SPARC
Stratosphere-troposphere Processes And their Role in Climate

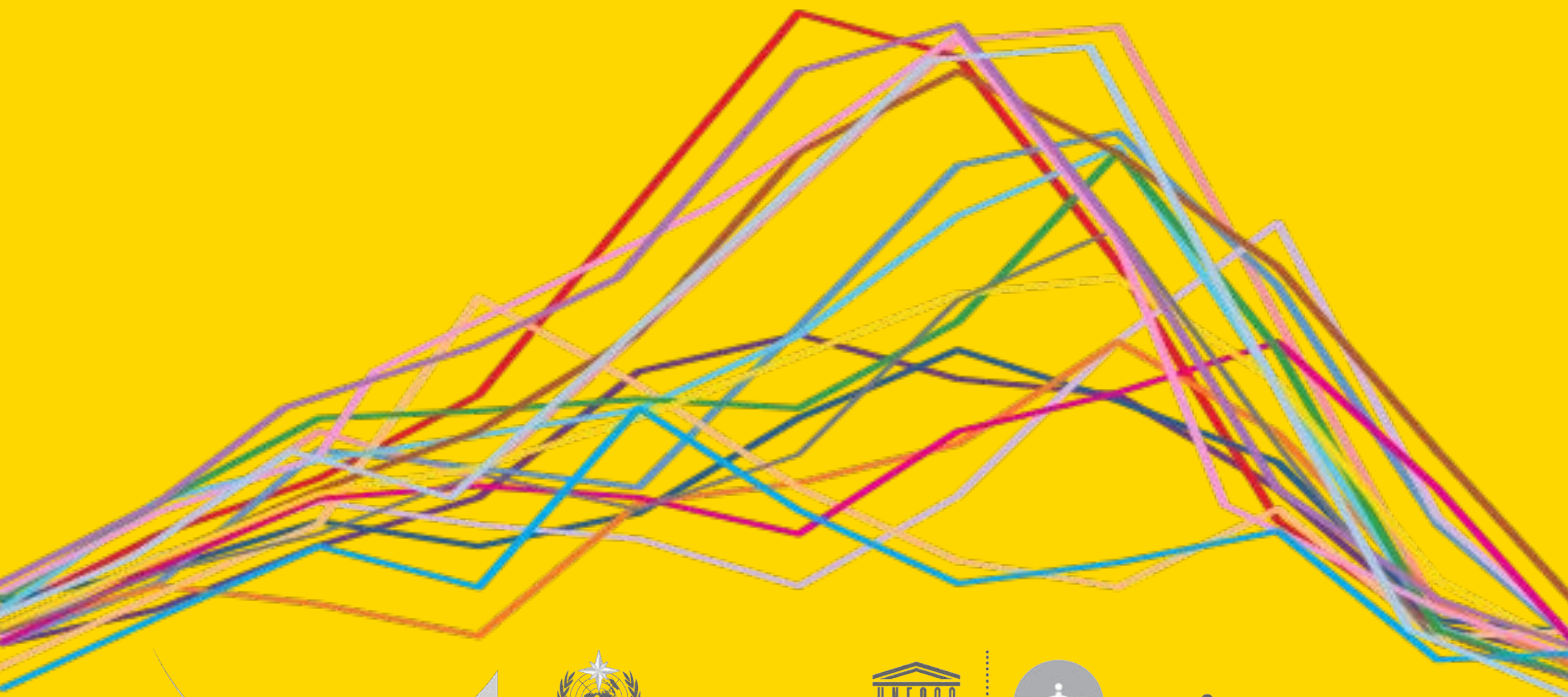
Core Project of the WMO/ISC/IOC
World Climate Research Programme

SPARC Reanalysis Intercomparison Project (S-RIP) Final Report

M. Fujiwara, G. L. Manney, L. J. Gray and J. S. Wright

S-RIP
SPARC Reanalysis Intercomparison Project

SPARC Report No. 10, WCRP Report 6/2021



WCRP World Climate Research Programme
WORLD METEOROLOGICAL ORGANIZATION
UNESCO United Nations Educational, Scientific and Cultural Organization
Intergovernmental Oceanographic Commission
International Science Council

SPARC Reanalysis Intercomparison Project (S-RIP) Final Report

January 2022

**Prepared by the SPARC Reanalysis Intercomparioson Project,
edited by M. Fujiwara, G.L. Manney, L.J. Gray, and J. S. Wright**

SPARC Report No. 10, WCRP Report 6/2021

Citation
SPARC, 2022: SPARC Reanalysis Intercomparison Project (S-RIP) Final Report. M. Fujiwara, G.L. Manney, L.J. Gray, and J.S. Wright (Eds.), SPARC Report No. 10, WCRP-17/2020, doi: 10.17874/800dee57d13, available at www.sparc-climate.org/publications/sparc-reports.

Electronic version available at **https://www.sparc-climate.org/sparc-report-no-10/**

ISSN 2296-5785 (Print)

Editing, Design, and Layout: Mareike Heckl, Sabrina Zechlau and Brigitte Ziegele

Authors & Reviewers[1]

SPARC Reanalysis Intercomparison Project Activity Leaders

Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Lesley J. Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK

Working group leaders of the SPARC S-RIP Activity

Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany
Simon Chabrillat	Royal Belgian Institute of Space Aeronomy	Belgium
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Michaela Hegglin	Department of Meteorology, University of Reading	UK
Kristin Krüger	Department of Geosciences, University of Oslo	Norway
Craig Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada

Report editors

Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Lesley J. Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China

Review editors

Judith Perlwitz	NOAA Physical Sciences Laboratory	USA
Vincent-Henri Peuch	European Centre for Medium-Range Weather Forecasts	UK, Italy
Seok-Woo Son	Seoul National University	Korea
Karen Rosenlof	National Oceanic and Atmospheric Administration	USA
Gufran Beig	Indian Institute of Tropical Meteorology	India

Authors and Contributors

Chapter 1: Introduction

Chapter lead authors

Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Lesley J. Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China

[1] Anonymous reviewers are not listed here.

Co-authors

James Anstey	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada	Canada
Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Edwin P. Gerber	Courant Institute of Mathematical Sciences, New York University	USA
V. Lynn Harvey	Laboratory for Atmospheric and Space Physics, University of Colorado	USA
Michaela I. Hegglin	Department of Meteorology, University of Reading	UK
Cameron R. Homeyer	School of Meteorology, University of Oklahoma	USA
John A. Knox	Department of Geography, University of Georgia	USA
Kirstin Krüger	Department of Geosciences, University of Oslo	Norway
Alyn Lambert	Jet Propulsion Laboratory, California Institute of Technology	USA
Craig S. Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan
Beatriz M. Monge-Sanz	Atmospheric, Oceanic and Planetary Physics, University of Oxford	UK
Michelle L. Santee	Jet Propulsion Laboratory, California Institute of Technology	USA
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
David G.H. Tan	European Centre for Medium-Range Weather Forecasts *(retired)*	UK
David R. Jackson	Met Office	UK
Saroja Polavarapu	Climate Research Division, Environment and Climate Change Canada	Canada
Gilbert P. Compo	(1) Cooperative Institute for Research in Environmental Sciences, University of Colorado (2) Physical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Rossana Dragani	European Centre for Medium-Range Weather Forecasts	UK
Wesley Ebisuzaki	National Oceanic and Atmospheric Administration	USA
Yayoi Harada	Japan Meteorological Agency	Japan
Chiaki Kobayashi	Japan Meteorological Agency	Japan
Krzysztof Wargan	(1) National Aeronautics and Space Administration (2) Science Systems and Applications, Inc.	USA
Jeffrey S. Whitaker	National Oceanic and Atmospheric Administration	USA

Reviewers

Joan Alexander	NorthWest Research Associates	USA
Greg Bodeker	Bodeker Scientzific	New Zealand

Chapter 2: Description of the Reanalysis Systems

Chapter lead authors

Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Craig Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA

Co-authors

James Anstey	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada	Canada
Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Gilbert P. Compo	(1) Cooperative Institute for Research in Environmental Sciences, University of Colorado (2) Physical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA

Rossana Dragani	European Centre for Medium-Range Weather Forecasts	UK
Wesley Ebisuzaki	National Oceanic and Atmospheric Administration	USA
Yayoi Harada	Japan Meteorological Agency	Japan
Chiaki Kobayashi	Japan Meteorological Agency	Japan
Will McCarty	National Aeronautics and Space Administration	USA
Andrea Molod	National Aeronautics and Space Administration	USA
Kazutoshi Onogi	Research Center for Advanced Science and Technology, the University of Tokyo *previously at Japan Meteorological Agency*	Japan
Steven Pawson	National Aeronautics and Space Administration	USA
Adrian Simmons	European Centre for Medium-Range Weather Forecasts	UK
David G.H. Tan	European Centre for Medium-Range Weather Forecasts *(retired)*	UK
Susann Tegtmeier	University of Saskatchewan	Canada
Krzysztof Wargan	(1) National Aeronautics and Space Administration (2) Science Systems and Applications, Inc.	USA
Jeffrey S. Whitaker	National Oceanic and Atmospheric Administration	USA
Cheng-Zhi Zou	National Oceanic and Atmospheric Administration	USA

Reviewers

Mark Buehner	Meteorological Research Division, Environment and Climate Change Canada	Canada
Alison Ming	Faculty of Mathematics, University of Cambridge	UK

Chapter 3: Overview of Temperature and Winds

Chapter lead authors

Craig Long	National Oceanic and Atmospheric Administration *(retired)*	USA
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan

Co-authors

Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Daniel Mitchell	University of Bristol	UK
Corwin Wright	University of Bath	UK

Reviewers

Alison Ming	Faculty of Mathematics, University of Cambridge	UK
William Randel	Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research	USA

Chapter 4: Overview of Ozone and Water Vapour

Chapter lead authors

Sean M. Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Michaela I. Hegglin	Department of Meteorology, University of Reading	UK

Co-authors

Rossana Dragani	European Centre for Medium-Range Weather Forecasts	UK
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Yayoi Harada	Japan Meteorological Agency	Japan
Chiaki Kobayashi	Japan Meteorological Agency	Japan
Craig Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA

Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Eric R. Nash	Science Systems and Applications, Inc	USA
Gerald L. Potter	National Aeronautics and Space Administration, Goddard Space Flight Center	USA
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Tao Wang	National Aeronautics and Space Administration, Jet Propulsion Laboratory	USA
Krzysztof Wargan	(1) National Aeronautics and Space Administration Global Modeling and Assimilation Office (2) Science Systems and Applications, Inc	USA
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China

Reviewers

Lucien Froidevaux	Jet Propulsion Laboratory, California Institute of Technology	USA
Gabriele Stiller	Institute of Meteorology and Climate Research - Atmospheric Trace Gases and Remote Sensing, Karlsruhe Institute of Technology, Germany	Germany
Mark Weber	Institute of Environmental Physics, University of Bremen	Germany

Chapter 5: Brewer-Dobson Circulation

Chapter lead authors

Beatriz M. Monge-Sanz	Atmospheric, Oceanic and Planetary Physics, University of Oxford	UK
Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany

Co-authors

Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Mohamadou Diallo	Institute of Energy and Climate Research, Stratosphere, Forschungzentrum Jülich	Germany
Florian Haenel	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Paul Konopka	Institute of Energy and Climate Research, Stratosphere,Forschungzentrum Jülich	Germany
Bernard Legras	Laboratoire de Météorologie Dynamique, IPSL, UPMC/ENS/CNRS/Ecole Polytech-nique	France
Felix Ploeger	Institute of Energy and Climate Research, Stratosphere, Forschungzentrum Jülich	Germany
Thomas Reddmann	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Gabriele Stiller	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China
Marta Abalos	Department of Earth Physics and Astrophysics, Universidad Complutense de Madrid	Spain
Harald Boenisch	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration (NOAA)	USA
Hella Garny	(1) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) (2) Meteorological Institute, Ludwig-Maximilians University Munich	Germany
Peter Hitchcock	Department of Earth and Atmospheric Sciences, Cornell University	USA
Kazuyuki Miyazaki	Tropospheric Composition Group, Jet Propulsion Laboratory, California Institute of Technology *previously at: Japan Agency for Marine-Earth Science and Technology, Japan*	USA
Howard K. Roscoe	British Antarctic Survey	UK
Kaoru Sato	Department of Earth and Planetary Science, The University of Tokyo	Japan
Mengchu Tao	Institute of Energy and Climate Research, Stratosphere, Forschungzentrum Jülich	Germany
Darryn Waugh	Department of Earth and Planetary Sciences, Johns Hopkins University	USA

Reviewers

Neal Butchart	Met Office Hadley Centre	UK
Amy Butler	NOAA Chemical Sciences Laboratory	USA
Chaim Garfinkel	The Fredy & Nadine Herrmann Institute of Earth Sciences, Hebrew University	Israel
Adrian Simmons	Copernicus Department, European Centre for Medium-Range Weather Forecasts	Germany

Chapter 6: Extratropical Stratosphere-Troposphere Coupling

Chapter lead authors

Edwin P. Gerber	Courant Institute of Mathematical Sciences, New York University	USA
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan

Co-authors

Blanca Ayarzagüena	(1)Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid (2)Instituto de Geociencias, CSIC-UCM	Spain
David Barriopedro	Instituto de Geociencias, CSIC-UCM	Spain
Thomas J. Bracegirdle	British Antarctic Survey	UK
Amy H. Butler	National Oceanic and Atmospheric Administration/Chemical Sciences Laboratory	USA
Natalia Calvo	Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid	Spain
Steven C. Hardiman	Met Office Hadley Centre	UK
Peter Hitchcock	Department of Earth and Atmospheric Sciences, Cornell University	USA
Maddalen Iza	Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid	Spain
Ulrike Langematz	Institut für Meteorologie, Freie Universität Berlin	Germany
Hua Lu	British Antarctic Survey	UK
Gareth Marshall	British Antarctic Survey	UK
Andrew Orr	British Antarctic Survey	UK
Froila M. Palmeiro	(1)Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid (2)Barcelona Supercomputing Center	Spain
Seok-Woo Son	School of Earth and Environmental Sciences, Seoul National University	South Korea
Masakazu Taguchi	Department of Earth Science, Aichi University of Education	Japan

Reviewers

Daniela Domeisen	(1) ETH Zurich (2) University of Lausanne	Switzerland
Chaim Garfinkel	The Fredy & Nadine Herrmann Institute of Earth Sciences, Hebrew University	Israel
Alexey Karpechko	Finnish Meteorological Instiute	Finland

Chapter 7: Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS)

Chapter lead authors

Cameron R. Homeyer	School of Meteorology, University of Oklahoma	USA
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA

Co-authors

Luis F. Millán	Jet Propulsion Laboratory, California Institute of Technology	USA
Alexander C. Boothe	(1) School of Meteorology, University of Oklahoma (2) Meteorologist, National Weather Service	USA
Tao Xian	(1) School of Meteorology, University of Oklahoma (2) School of Earth and Space Sciences, University of Science and Technology of China	USA China

Mark A. Olsen	NOAA/Oceanic and Atmospheric Research/Weather Program Office *previously at: Goddard Earth Sciences Technology and Research, Morgan State University*	USA
Michael J. Schwartz	Jet Propulsion Laboratory, California Institute of Technology	USA
Michaela I. Hegglin	Department of Meteorology, University of Reading	UK
Zachary D. Lawrence	(1) Cooperative Institute for Research in Environmental Sciences (CIRES), Univ. of Colorado (2) NOAA Physical Sciences Laboratory (PSL) (3) NorthWest Research Associates	USA
Krzysztof Wargan	(1) Science Systems and Applications Inc. (2) Global Modeling and Assimilation Office, NASA Goddard Space Flight Center	USA

Reviewers

Clara Orbe	(1) NASA Goddard Institute for Space Studies (2) Department of Applied Physics and Applied Mathematics, Columbia University	USA
Eric Ray	CIRES, University of Colorado Boulder and NOAA Chemical Sciences Laboratory	USA
Darryn Waugh	Earth and Planetary Science, Johns Hopkins University	USA

Chapter 8: Tropical Tropopause Layer

Chapter lead authors

Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Kirstin Krüger	Section for Meteorology and Oceanography, Department of Geosciences, University of Oslo	Norway

Co-authors

Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany
Nicholas A. Davis	Atmospheric Chemistry Observations & Modeling Lab, National Center for Atmospheric Research	USA
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Cameron R. Homeyer	School of Meteorology, University of Oklahoma	USA
Ioana Ivanciu	GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel	Germany
Young-Ha Kim	Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt *previously at: Severe Storm Research Center, Ewha Womans University, South Korea*	Germany
Bernard Legras	Laboratoire de Météorologie Dynamique, CNRS/ENS-PSL	France
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Eriko Nishimoto	NTT DATA INTELLILINK Corporation *previously at: Japan Agency for Marine Earth Science and Technology*	Japan
Matthias Nützel	Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen)	Germany
Robin Pilch Kedzierski	GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel	Germany
James S. Wang	Institute for Advanced Sustainability Studies	Germany
Tao Wang	Jet Propulsion Laboratory, California Institute of Technology	USA
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China

Reviewers

Andrew Gettelmann	National Center for Atmospheric Research	USA
Peter Haynes	Department of Applied Mathematics and Theoretical Physics, University of Cambridge	UK
William Randel	Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research	USA

Chapter 9: Quasi-Biennial Oscillation

Chapter lead authors

James A. Anstey	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada	Canada
Lesley J. Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK

Co-authors

Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Ioana Ivanciu	GEOMAR Helmholtz-Zentrum für Oceanforschung Kiel	Germany
Yoshio Kawatani	Japan Agency for Marine-Earth Science and Technology	Japan
George Kiladis	Physical Sciences Division, NOAA/Earth System Research Laboratory	USA
Young-Ha Kim	Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt *previously at: Severe Storm Research Center, Ewha Womans University, South Korea*	Germany
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan
Verena Schenzinger	Institute for Meteorology and Geophysics, University of Vienna	Austria
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Corwin Wright	Centre for Space, Atmospheric and Oceanic Science, University of Bath	UK

Reviewers

Lawrence Coy	(1) Science Systems and Applications, Inc. (2) and NASA Goddard Space Flight Center	USA
Kevin Hamilton	International Pacific Research Center, University of Hawaii	USA
Matt Hitchman	Atmospheric and Oceanic Science, University of Wisconsin - Madison	USA
Hiroaki Naoe	Department of Climate and Geochemistry Research, Japan Meteorological Agency	Japan
Scott Osprey	Atmospheric, Oceanic and Planetary Physics, University of Oxford	UK

Chapter 10: Polar Processes

Chapter lead authors

Michelle L. Santee	NASA Jet Propulsion Laboratory, California Institute of Technology	USA
Alyn Lambert	NASA Jet Propulsion Laboratory, California Institute of Technology	USA
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA

Co-authors

Zachary D. Lawrence	(1) Cooperative Institute for Research in Environmental Sciences (CIRES), Univ. of Colorado (2) NOAA Physical Sciences Laboratory (3) NorthWest Research Associates	USA
Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Lars Hoffmann	Forschungszentrum Jülich GmbH	Germany
Sean P. Palmer	Department of Physics, New Mexico Institute of Mining and Technology	USA
Ken Minschwaner	Department of Physics, New Mexico Institute of Mining and Technology	USA

Reviewers

Jens-Uwe Grooß	Institute of Energy and Climate Research - Stratosphere, Forschungszentrum Juelich	Germany
Michael Höpfner	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Farahanz Khosrawi	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany

Chapter 11: Upper Stratosphere and Lower Mesosphere

Chapter lead authors

V. Lynn Harvey	(1) Laboratory for Atmospheric and Space Physics (2) Atmospheric and Oceanic Sciences Department, University of Colorado	USA
John Knox	Department of Geography, University of Georgia	USA

Co-authors

Jeff France	White Ridge Solutions	USA
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Lesley Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK
Toshihiko Hirooka	Department of Earth and Planetary Sciences, Kyushu University	Japan
Peter Hitchcock	Department of Earth and Atmospheric Sciences, Cornell University	USA
Matthew Hitchman	Department of Atmospheric and Oceanic Sciences, University of Wisconsin – Madison	USA
Yoshio Kawatani	Japan Agency for Marine-Earth Science and Technology	Japan
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
John McCormack	Heliophysics Division, Science Mission Directorate, NASA Headquarters *previously at Space Science Division, Naval Research Laboratory*	USA
Yvan Orsolini	Norwegian Institute for Air Research	Norway
Takatoshi Sakazaki	Graduate School of Science, Kyoto University	Japan
Yoshihiro Tomikawa	(1) National Institute of Polar Research (2) The Graduate University for Advanced Studies (SOKENDAI)	Japan

Reviewers

Ruth Liebermann	NASA Goddard Institute for Space Studie	USA
Theodore Shepherd	Department of Meteorology, University of Reading	UK
David Siskind	Space Science Division, Naval Research Laboratory	USA
Anne Smith	National Center for Atmospheric Research	USA

Chapter 12: Synthesis Summary

Chapter lead authors

Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Lesley J. Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China
James Anstey	Canadian Centre for Climate Modelling and Analysis Environment and Climate Change Canada	Canada
Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Edwin P. Gerber	Courant Institute of Mathematical Sciences, New York University	USA
V. Lynn Harvey	Laboratory for Atmospheric and Space Physics, University of Colorado	USA
Michaela I. Hegglin	Department of Meteorology, University of Reading	UK
Cameron R. Homeyer	School of Meteorology, University of Oklahoma	USA
John A. Knox	Department of Geography, University of Georgia	USA

Kirstin Krüger	Department of Geosciences, University of Oslo	Norway
Alyn Lambert	Jet Propulsion Laboratory, California Institute of Technology	USA
Craig S. Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan
Beatriz M. Monge-Sanz	Atmospheric, Oceanic and Planetary Physics, University of Oxford	UK
Michelle L. Santee	Jet Propulsion Laboratory, California Institute of Technology	USA
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Zachary D. Lawrence	(1) Cooperative Institute for Research in Environmental Sciences, Univ. of Colorado (2) NOAA Physical Sciences Laboratory (PSL) (3) NorthWest Research Associates	USA
David G. H. Tan	European Centre for Medium-Range Weather Forecasts *(retired)*	UK
David R. Jackson	Met Office	UK
Saroja Polavarapu	Climate Research Division, Environment and Climate Change Canada	Canada
Gilbert P. Compo	(1) Cooperative Institute for Research in Environmental Sciences, University of Colorado (2) Physical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Rossana Dragani	European Centre for Medium-Range Weather Forecasts	UK
Wesley Ebisuzaki	National Oceanic and Atmospheric Administration	USA
Yayoi Harada	Japan Meteorological Agency	Japan
Chiaki Kobayashi	Japan Meteorological Agency	Japan
Krzysztof Wargan	(1) National Aeronautics and Space Administration (2) Science Systems and Applications, Inc.	USA
Jeffrey S. Whitaker	National Oceanic and Atmospheric Administration	USA

Acknowledgements

We acknowledge the scientific guidance and sponsorship of the World Climate Research Programme, coordinated in the framework of SPARC. We are grateful to Theodore Shepherd and Greg Bodeker (past SPARC co-chairs) for strong encouragement and valuable advice during the proposal phase of the project during 2011 - 2012, and Joan Alexander (serving SPARC co-chair during 2013 - 2015), Neil Harris (serving SPARC co-chair since 2014), Judith Perlwitz (serving SPARC co-chair during 2016 - 2020), and Seok-Woo Son (serving SPARC co-chair since 2019) for continued support and encouragement. We are also very grateful to the review handling editors of this report, Judith Perlwitz, Seok-Woo Son, Vincent-Henri Peuch, Karen Rosenlof, and Gufran Beig, and to the reviewers. The Information Initiative Center of Hokkaido University, Japan, has hosted the S-RIP web server since 2014. We thank the reanalysis centres for providing their support and data products. The British Atmospheric Data Centre (BADC) of the UK Centre for Environmental Data Analysis (CEDA) has provided a virtual machine for data processing, a group workspace for data storage, and public data archive sites. We thank Yulia Zyulyaeva for contributions to S-RIP as an original member of the working group, as a chapter co-lead through October 2014, and as the designer of the S-RIP logo. We thank Diane Pendlebury for contributions to S-RIP as a chapter co-lead through August 2015. We thank Quentin Errera, the lead of the SPARC Data Assimilation working group, for co-organizing the 2014, 2015, 2016, and 2017 S-RIP workshops. We also thank John McCormack for co-organizing the 2017 workshop. Travel support for some participants of the 2013 planning meeting, the 2014, 2015, 2016, and 2017 workshops, and the 2018 chapter-lead meeting was provided by SPARC. We thank Peter Haynes and Gabriele Stiller for their work as guest editors for the special issue "The SPARC Reanalysis Intercomparison Project (S-RIP)" in the journals Atmospheric Chemistry and Physics (ACP) and Earth System Science Data (ESSD); we offer special thanks to William Lahoz (1960 - 2019) for his role as the other guest editor of this special issue, which he performed excellently even as he battled his final illness. We thank the hard works at the SPARC Office, in particular Mareike Heckl, Brigitte Ziegele, and Sabrina Zechlau for their typesetting of the report. Work at the Jet Propulsion Laboratory, California Institute of Technology, was carried out under a contract with the National Aeronautics and Space Administration. **https://creativecommons.org/licenses/by/3.0/**.

Table of contents

Chapter 1: Introduction **1**

Chapter 2: Description of the Reanalysis Systems **15**

2.1 Introduction 17
2.2 Forecast models 20
2.2.1 Summary of basic information 20
2.2.2 Major physical parameterizations 22
2.2.3 Boundary and other specified conditions 27
2.2.4 Surface air and land surface treatments 31
2.3 Assimilation Schemes 35
2.3.1 Basics of data assimilation 35
2.3.2 Data assimilation in reanalysis systems 38
2.4 Observational Data 40
2.4.1 Summary of basic information 40
2.4.2 Quality control procedures 47
2.4.3 Summary of key upper air observations and known issues 49
2.4.3.1 Radiosonde data 49
2.4.3.2 Satellite data 51
2.4.3.3 Aircraft data 54
2.4.4 Water vapour 54
2.5 Execution streams 55
2.5.1 What is an ‘execution stream’? 55
2.5.2 Summary of stream execution 55
2.6 Archived data 57
References 59
Appendix A: Vertical levels of the models 69
Major abbreviations and terms 75

Chapter 3: Overview of Temperature and Winds **81**

3.1 Introduction 83
3.2 Improvements from older reanalyses to newer versions 84
3.3 Reanalysis global mean temperature anomaly variability 84
3.4 Reanalysis ensemble mean (REM) 86

3.4.1 Methodology 86
3.4.2 Climatology of the REM....87
3.4.3 Agreement among the REM members 91
3.5 Intercomparisons of the reanalyses 94
3.5.1 Temperature 94
3.5.2 Zonal wind 99
3.6 Amplitude of polar annual temperature cycle....103
3.7 Comparisons with satellite temperature observations 104
3.7.1 HIRDLS and MLS temperatures....104
3.7.2 Comparisons with COSMIC temperature observations....106
3.7.3 Atmospheric layer temperature anomalies....107
3.7.4 Ringing of SH polar latitude temperature differences....111
3.8 Comparisons against other observations....111
3.8.1 Ozonesonde temperatures....111
3.8.2 Long-duration balloon observations 112
3.8.3 Rocketsonde observations 113
3.9 Effects of volcanic eruptions and other natural variabilities....113
3.10 Summary and conclusions 114
3.11 Key findings and recommendations 116
References....117
Major abbreviations and terms....120

Chapter 4: Overview of Ozone and Water Vapour 123

4.1 Introduction....125
4.2 Description of ozone and water vapour in reanalyses....125
4.2.1 NCEP-NCAR (R1) and NCEP-DOE (R2) 127
4.2.2 CFSR 128
4.2.3 ERA-40 129
4.2.4 ERA-Interim 130
4.2.5 ERA5....131
4.2.6 JRA-25 and JRA-55 131
4.2.7 MERRA 131
4.2.8 MERRA-2....132
4.3 Data....133
4.3.1 Reanalysis data processing 133
4.3.2 SBUV and TOMS/OMI total column ozone....133
4.3.3 SPARC Data Initiative limb satellite observations 133
4.3.4 Aura MLS satellite data....134

4.3.5 SWOOSH merged limb satellite data record ... 134
4.4 Evaluation of reanalysis ozone products ... 134
4.4.1 Total column ozone seasonal cycle ... 134
4.4.2 Zonal mean ozone cross-sections ... 136
4.4.3 Ozone monthly mean vertical profiles and seasonal cycles ... 137
4.4.4 Ozone interannual variability ... 139
4.4.5 Ozone time series in equivalent latitude coordinates ... 139
4.4.6 Ozone quasi-biennial oscillation ... 141
4.4.7 Ozone hole area ... 144
4.4.8 Long-term evolution of ozone ... 145
4.5 Evaluation of reanalysis water vapour products ... 150
4.5.1 Zonal mean water vapour cross-sections ... 150
4.5.2 Water vapour monthly mean vertical profiles and seasonal cycles ... 151
4.5.3 Water vapour interannual variability ... 151
4.5.4 Tropical tape recorder in water vapour ... 153
4.6 Summary ... 155
4.7 Key findings and recommendations ... 157
References ... 159
Major abbreviations and terms ... 162

Chapter 5: Brewer-Dobson Circulation ... 165

5.1 Introduction ... 167
5.1.1 General description of the BDC and approach ... 167
5.1.2 Chapter objectives ... 167
5.2 Diagnostics description ... 168
5.2.1 Dynamical variables ... 168
5.2.2 Transport tracers from offline simulations ... 169
5.3 Offline models description ... 171
5.3.1 Description of the BASCOE model ... 171
5.3.2 Description of the CLaMS model ... 171
5.3.3 Description of the KASIMA model ... 172
5.3.4 Description of the TOMCAT/SLIMCAT model ... 172
5.3.5 Description of the TRACZILLA model ... 173
5.4 Description of tracers observations ... 173
5.4.1 “Standard” mean AoA observations for model intercomparisons ... 173
5.4.2 Long timeseries of mean AoA in the Northern Hemisphere ... 173
5.4.3 MIPAS AoA dataset based on tracer observations ... 174
5.4.4 BAS Polar tracer observations ... 174

5.4.5 Stratospheric water vapour tape-recorder observations 174
5.5 Comparison results 175
5.5.1 Results from dynamical variables 175
5.5.2 Results from transport tracers simulations 186
5.6 Discussion 206
5.7 Conclusions and recommendations 210
5.7.1 Conclusions from dynamics diagnostics 210
5.7.2 Conclusions from transport tracers simulations 210
5.7.3 Recommendations to reanalyses users 211
5.7.4 Recommendations to reanalyses centres 212
References 214
Major abbreviations and terms 218

Chapter 6: Extratropical Stratosphere–Troposphere Coupling 221

6.1 Introduction 223
6.2 Context and background 223
6.3 Reanalysis datasets 225
6.4 Sudden stratospheric warming events 225
6.4.1 Identifying SSW events 226
6.4.2 Characterizing SSW events 226
6.4.3 Sampling uncertainty vs. reanalysis uncertainty 231
6.4.4 Assessing the internal consistency of SSW events in Reanalyses 233
6.5Annular modes 235
6.5.1 Consistency of the annular mode index across reanalyses in the post and pre-satellite periods 235
6.5.2 Sampling uncertainty vs. reanalysis uncertainty 238
6.6 Stratospheric final warming events 240
6.7 Modulation of stratosphere-troposphere coupling by ENSO and QBO 241
6.7.1 Troposphere-stratosphere coupling through ENSO 242
6.7.2 Blocking patterns associated to SSWs and the modulation of ENSO 243
6.7.3 Nonlinear modulation of the extratropical stratosphere by ENSO and QBO 244
6.8 Stratosphere-troposphere coupling through the antarctic hzone hole 246
6.9 Outlook, key findings, and recommendations 249
References 255
Appendix A: Detection and classification of major SSW events 260
Major abbreviations and terms 264

Chapter 7: Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS)............... 265

7.1 Introduction 267
7.2 Reanalyses and general approach 267
7.3 The extratropical tropopause 268
7.3.1 Lapse rate tropopause altitudes 269
7.3.2 Dynamical tropopause altitudes 271
7.3.3 Multiple tropopauses 273
7.4 Jet streams 277
7.4.1 Jet characterization for UTLS studies 277
7.4.2 Climatology of UTLS jets in reanalyses 279
7.4.3 Trends in UTLS jets in reanalyses 281
7.5 Transport and mixing 282
7.5.1 Stratosphere-troposphere exchange 282
7.5.2 Mixing and transport barriers 284
7.5.3 Mass flux across 380 K isentropic surface 288
7.6 UTLS ozone 290
7.6.1 Ozone in tropopause-relative coordinates 291
7.6.2 Ozone in equivalent latitude coordinates 293
7.6.3 Ozone in jet-relative coordinates 294
7.6.4 Ozone mini-holes 298
7.7 Summary and recommendations 299
References 303
Major abbreviations and terms 307

Chapter 8: Tropical Tropopause Layer.. 309

8.1 Introduction 311
8.2 Temperature and tropopause characteristics 312
8.2.1 Observational data sets 312
8.2.2 Climatology 312
8.2.3 Interannual variability and long-term changes 315
8.2.4 Key findings and recommendations 317
8.3 Clouds and convection 318
8.3.1 Observational data sets 318
8.3.2 Spatial distribution of high clouds 320
8.3.3 Vertical profiles 321
8.3.4 Cloud radiative effects 323
8.3.5 Relationships with other variables 326
8.3.6 Temporal variability 328

8.3.7 Key findings and recommendations 330
8.4 Diabatic heating rates 331
8.4.1 Total diabatic heating 331
8.4.2 Radiative heating....332
8.4.3 Non-radiative heating 337
8.4.4 Key findings and recommendations 338
8.5 Transport 339
8.5.1 Dehydration point distribution 339
8.5.2 TTL residence time 341
8.5.3 TTL tropical upwelling....343
8.5.4 Key findings and recommendations 343
8.6 Wave activity 345
8.6.1 Horseshoe-shaped structure at the 100 hPa temperature....345
8.6.2 Equatorial waves.... 346
8.6.3 Key findings 348
8.7 Width of the TTL....348
8.7.1 Zonally-resolved subtropical jet diagnostic 349
8.7.2 Zonally-resolved tropopause break diagnostic 351
8.7.3 Zonal mean subtropical jet and tropopause break diagnostics....352
8.7.4 Key findings and recommendations....353
8.8 South Asian summer monsoon 353
8.8.1 Anticyclone: climatology and variability 354
8.8.2 Vertical velocity....359
8.8.3 Diabatic heating....361
8.8.4 Transport....363
8.8.5 Ozone....364
8.8.6 Regional analysis of clouds and radiative effects....365
8.8.7 Key findings and recommendations 371
8.9 Summary, key findings, and recommendations....372
References....376
Appendix A: Supplementary material 382
Major abbreviations and terms....388

Chapter 9: Quasi-Biennial Oscillation 391

9.1 Introduction....393
9.1.1 Issues in representing the QBO....393
9.1.2 Data and methods....395
9.2 Monthly-mean equatorial variability 396
9.2.1 Time evolution 396

9.2.2 QBO amplitude and phase transitions....403
9.2.3 Mean meridional circulation....411
9.2.4 Near-tropopause temperature....413
9.3 Tropical waves and QBO forcing....416
9.3.1 Characteristics of the equatorial waves....418
9.3.2 Momentum budget of the QBO....421
9.3.3 Direct comparison with satellite data....426
9.4 QBO teleconnections....428
9.4.1 Stratospheric teleconnections....429
9.4.2 Tropospheric teleconnections....438
9.4.3 QBO teleconnections in context....439
9.4.4 Surface teleconnections....441
9.5 Summary, key findings, and recommendations....444
9.5.1 Summary for monthly-mean equatorial variability....444
9.5.2 Summary for tropical waves and QBO forcing....445
9.5.3 Summary for QBO teleconnections....446
9.5.4 Recommendations....446
References....449
Appendix A: Additional Figures....453
Major abbreviations and terms....488

Chapter 10: Polar Processes....491

10.1 Introduction....493
10.2 Description of atmospheric measurements....495
10.2.1 Aura Microwave Limb Sounder....495
10.2.2 Envisat MIPAS....496
10.2.3 CALIPSO CALIOP....496
10.2.4 COSMIC GNSS-RO....496
10.2.5 Concordiasi superpressure balloon measurements....497
10.3 Overview of reanalysis polar temperature differences....498
10.4 Polar temperature and vortex diagnostics....498
10.5 Polar diabatic heating rates....502
10.6 Concordiasi superpressure balloon comparisons....505
10.6.1 Temperatures and winds....505
10.6.2 Trajectories....506
10.7 PSC thermodynamic-consistency diagnostics....507
10.7.1 Mountain waves....507
10.7.2 Thermodynamic equilibrium....508
10.7.3 Temperature difference profiles....509

10.7.4 Summary of the temperature differences ... 510
10.8 Chemical modeling diagnostics ... 510
10.8.1 Details of the BASCOE system and experimental setup ... 511
10.8.2 Analysis approach ... 512
10.8.3 Case study: 2009 Antarctic winter ... 512
10.8.4 Case study: 2009/2010 Arctic winter ... 519
10.8.5 Comparisons of chemical ozone loss ... 520
10.8.6 Discussion and implications ... 521
10.9 Summary, key findings, and recommendations ... 522
References ... 526
Major abbreviations and terms ... 531

Chapter 11: Upper Stratosphere and Lower Mesosphere ... 533

11.1 Introduction ... 535
11.1.1 Reanalysis products used in this chapter ... 535
11.1.2 Satellite observational products ... 536
11.1.3 Upper boundary conditions in the reanalyses ... 537
11.1.4 Physical parameterizations in the reanalyses specific to USLM ... 537
11.1.5 Long-term effects of data assimilation discontinuities ... 537
11.1.6 Variability among reanalyses ... 539
11.2 Climatology of the USLM ... 541
11.2.1 Seasonal zonal means ... 541
11.2.2 Annual cycles ... 546
11.2.3 Long-term variability ... 550
11.3 Tropical dynamics ... 554
11.3.1 Semi-annual oscillation ... 554
11.3.2 Middle-atmosphere Hadley circulation ... 556
11.3.3 Inertial instability ... 557
11.4 Polar dynamics ... 559
11.4.1 Polar vortices ... 559
11.4.2 Planetary waves ... 561
11.4.3 Elevated stratopause events ... 561
11.5 Tides and normal modes ... 564
11.5.1 Tides ... 564
11.5.2 Quasi-2-day wave ... 566
11.5.3 Quasi-5-day wave ... 569
11.6 Summary, key findings, and recommendations ... 571
References ... 574
Major abbreviations and terms ... 581

Chapter 12: Synthesis Summary ... 583

12.1 Key findings and recommendations by chapter ... 585
12.1.1 Chapter 3: Overview of Temperature and Winds ... 585
12.1.2 Chapter 4: Overview of Ozone and Water Vapour ... 587
12.1.3 Chapter 5: Brewer–Dobson Circulation ... 588
12.1.4 Chapter 6: Extratropical Stratosphere–Troposphere Coupling ... 591
12.1.5 Chapter 7: Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS) ... 592
12.1.6 Chapter 8: Tropical Tropopause Layer (TTL) ... 594
12.1.7 Chapter 9: Quasi-Biennial Oscillation (QBO) ... 597
12.1.8 Chapter 10: Polar Processes ... 599
12.1.9 Chapter 11: Upper Stratosphere and Lower Mesosphere ... 601
12.2 Overall findings and reanalysis user recommendations ... 604
12.3 Recommendations for improving reanalyses and their evaluation ... 604
12.3.1 S-RIP survey results and related product needs from S-RIP studies ... 605
12.3.2 Data access issues ... 607
12.3.3 Documentation issues ... 607
12.4 Prospects for the future ... 608
Major abbreviations and terms ... 609

Chapter 1: Introduction

Chapter lead authors

Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Lesley J. Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China

Co-authors

James Anstey	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada	Canada
Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Edwin P. Gerber	Courant Institute of Mathematical Sciences, New York University	USA
V. Lynn Harvey	Laboratory for Atmospheric and Space Physics, University of Colorado	USA
Michaela I. Hegglin	Department of Meteorology, University of Reading	UK
Cameron R. Homeyer	School of Meteorology, University of Oklahoma	USA
John A. Knox	Department of Geography, University of Georgia	USA
Kirstin Krüger	Department of Geosciences, University of Oslo	Norway
Alyn Lambert	Jet Propulsion Laboratory, California Institute of Technology	USA
Craig S. Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan
Beatriz M. Monge-Sanz	Atmospheric, Oceanic and Planetary Physics, University of Oxford	UK
Michelle L. Santee	Jet Propulsion Laboratory, California Institute of Technology	USA
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
David G.H. Tan	European Centre for Medium-Range Weather Forecasts *(retired)*	UK
David R. Jackson	Met Office	UK
Saroja Polavarapu	Climate Research Division, Environment and Climate Change Canada	Canada
Gilbert P. Compo	(1) Cooperative Institute for Research in Environmental Sciences, University of Colorado (2) Physical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Rossana Dragani	European Centre for Medium-Range Weather Forecasts	UK
Wesley Ebisuzaki	National Oceanic and Atmospheric Administration	USA
Yayoi Harada	Japan Meteorological Agency	Japan
Chiaki Kobayashi	Japan Meteorological Agency	Japan
Krzysztof Wargan	(1) National Aeronautics and Space Administration (2) Science Systems and Applications, Inc.	USA
Jeffrey S. Whitaker	National Oceanic and Atmospheric Administration	USA

Fujiwara *et al.* (2017a) have published a shortened version of this chapter.

1.1 Motivation and goals

An atmospheric reanalysis system consists of a global forecast model, input observations, and an assimilation scheme, which are used in combination to produce best estimates (analyses) of past atmospheric states. Whereas operational analysis systems are continuously updated with the intention of improving numerical weather predictions, reanalysis systems are (typically) fixed throughout their lifetime. Using a fixed assimilation–forecast model system to produce analyses of observational data previously analysed in the context of operational forecasting (the "re" in "reanalysis") helps to prevent the introduction of inhomogeneities in the analysed fields due to changing assimilation-forecast model systems (*Trenberth and Olson,* 1988; *Bengtsson and Shukla,* 1988; see also *Fujiwara et al.*, 2017a), although artificial changes still arise from other sources (especially from changes in the quality and/or quantity of the input observational data).

The Stratosphere-troposphere Processes And their Role in Climate (SPARC) project is one of the four core projects of the World Climate Research Programme (WCRP). Researchers interested in SPARC use global atmospheric reanalysis products (**Table 1.1**) (1) to understand a wide range of processes and variability in the atmosphere, (2) to validate chemistry climate models, and (3) to investigate and identify climate change (e.g., *SPARC,* 2010; *Randel et al.,* 2004; *SPARC*, 2002, and references therein). Even for more recent reanalyses, however, different results may be obtained for the same diagnostic due to the different methodologies used to construct the reanalysis data sets (see, *e.g.*, *Fujiwara et al.*, 2017a for examples). There is a need, therefore, for a coordinated intercomparison of reanalysis data sets with respect to key diagnostics that can help to clarify the causes of these differences. The results can then be used to provide guidance on the appropriate use of reanalysis products in scientific studies, particularly those of relevance to SPARC. Forecasting and research centres that produce reanalyses also benefit from coordinated user feedback, which helps to drive improvements in the next generation of reanalysis products.

Table 1.1: *Global atmospheric reanalysis data sets available as of July 2018. See end of chapter 1 for abbreviations.*

Reanalysis Centre	Name of the Reanalysis Product
ECMWF	ERA-40, ERA-Interim, ERA-20C, CERA-20C, ERA5 [1]
JMA	JRA-25/JCDAS, JRA-55
NASA	MERRA, MERRA-2
NOAA/NCEP	NCEP-NCAR R1, NCEP-DOE R2, CFSR/CFSv2 [2]
NOAA and Univ. Colorado	20CR

[1] Some ERA5 data have been available since July 2018, ERA5 data from 1979 onward have been available since January 2019, and a preliminary version of ERA5 1950-1978 data have been available since November 2020. Because most of the studies in this report were finalized before ERA5 was readily available, full evaluation of ERA5 has not been made. However, Chapter 2 includes information on the ERA5 system, and some chapters show some ERA5 results.

[2] CFSR is for the period from January 1979 to December 2010, and CFSv2 is for the period from January 2011 to present. We strongly recommend explicitly referring to the combination "CFSR/CFSv2" in documenting any study that uses these products across the 2010-2011 transition.

The SPARC Reanalysis Intercomparison Project (S-RIP) was initiated in 2011 and officially started in 2013 to conduct a coordinated intercomparison of all major global atmospheric reanalysis data sets. While the focus is on the stratosphere, the intercomparison also encompasses the troposphere and lower mesosphere where appropriate.

The goals of S-RIP are as follows:

i. To create a communication platform between SPARC-related researchers and the reanalysis centres;

ii. To better understand the differences among current reanalysis products and their underlying causes and to contribute to future reanalysis improvements; and

iii. To provide guidance to reanalysis data users by documenting the results of this reanalysis intercomparison in peer reviewed papers and this S-RIP report.

This chapter discusses the scope and plans of S-RIP based on the S-RIP Implementation Plan (February 2014) and updated information.

1.2 Scope

The S-RIP activity focuses predominantly on reanalyses, although some chapters include diagnostics from operational analyses when appropriate. Available reanalysis data sets (as of July 2018) are listed in **Table 1.1**. The guidelines for the choice of reanalysis data sets are detailed below. Many of the chapters focus primarily on newer reanalysis systems that assimilate upper-air measurements and produce data at relatively high resolution (*i.e.*, ERA-Interim, JRA-55, MERRA, MERRA-2, and CFSR). The ERA5 reanalysis, which was released during the latter stages of the activity, is not fully evaluated but is included in some intercomparisons. Selected long-term reanalyses that assimilate only surface meteorological observations (*e.g.*, NOAA-CIRES 20CR, ERA-20C, and CERA-20C) are also evaluated where appropriate. Some chapters include comparisons with older reanalyses (NCEP-NCAR R1, NCEP-DOE R2, ERA-40, and JRA-25/JCDAS), because these products have been extensively used in the past and are still being used for some studies, and because such comparisons can provide insight into the potential shortcomings of past research results.

Other chapters only include a subset of these reanalysis data sets, since some reanalyses have already been shown to perform poorly for certain diagnostics or do not extend high enough (*e.g.*, pressures less than 10 hPa) in the atmosphere. At the beginning of each chapter an explanation is given as to why specific reanalysis data sets were included or excluded.

The minimum intercomparison period is 1980 - 2010. This period starts with the availability of MERRA-2 shortly after the advent of high-frequency remotely sensed data in late 1978 and ends with the transition between CFSR and CFSv2. Some chapters also consider the pre-satellite era before 1979 and/or include results for more recent years. Some chapters use shorter intercomparison periods for some diagnostics due to limitations in the observational record available for comparison and/or computational resources.

1.3 Outline of this report

Summarised below are the components of this S-RIP report (see also **Figure 1.1**). On initiation of the project in 2013, it was planned to publish an interim report containing preliminary versions of Chapters 1 - 4 prior to the completion of the full report. However, following the publication of three papers covering the material in the planned interim report (*Fujiwara et al.*, 2017a; *Long et al.*, 2017; *Davis et al.*, 2017), it was decided in early 2018 that preparing a separate report was unnecessary, and the interim report was cancelled in favor of focusing on the full report. *Chapters 1 and 2* are introductory chapters. *Chapters 3 and 4* are overview chapters, for major dynamical variables in the former and for ozone and water vapour in the latter. *Chapters 5 - 11* are more process-oriented chapters, and are arranged according to, and focus on, different regions or processes within the atmosphere. It is noted that stratosphere-troposphere exchange processes are primarily evaluated in *Chapter 7* (Extratropical upper troposphere and lower stratosphere), while *Chapter 5* (Brewer-Dobson circulation) primarily evaluates the mass transport within the stratosphere. Also, *Chapter 6* (Extratropical stratosphere-troposphere coupling) deals with dynamical coupling, not transport processes. Furthermore, the processes in the upper troposphere and lower stratosphere (UTLS) are discussed separately in *Chapters 7* (Extratropical UTLS) and *8* (Tropical Tropopause Layer, TTL); *Chapter 7* begins with an introduction to these two UTLS chapters, and explains the distinction between the two UTLS regions while identifying key processes and common diagnostics used to study these processes in each region. Some important topics, such as gravity wave drag and transport processes, are sufficiently pervasive that related aspects are distributed amongst several chapters. *Chapter 12* synthesizes the findings and recommendations.

In the summary sections of *Chapters 3 - 11* and in *Chapter 12*, we use the following terms to provide context to our recommendations for each diagnostic:

- **Demonstrated suitable**: the reanalysis product could be directly validated using observational or physical constraints and was found to be in close agreement with expectations

- **Suitable with limitations**: the reanalysis product could be directly validated using observational or physical constraints and exhibited limited agreement; or, appropriate constraints were unavailable but reanalysis products were consistent beyond specific limitations as described in the text

- **Use with caution**: the reanalysis system contains all elements necessary to provide a useful representation of this variable or process, but that representation has evident red flags (*e.g.*, disagreement with available observations; meaningful disagreements among reanalyses that cannot be resolved at this point)

- **Demonstrated unsuitable**: the reanalysis product has been flagged as unable to represent processes that are key for this diagnostic as assessed in this report or by previous studies. This category is reserved for situations where the reanalysis is missing something fundamental in its structure (*e.g.*, a model top at 3 hPa means NCEP-NCAR R1 is 'demonstrated unsuitable' for studying processes in the Upper Stratosphere and Lower Mesosphere (USLM))

- **Unevaluated**: the performance of the reanalysis product with respect to this diagnostic or variable has not been examined in this report or by previous studies

It is noted that many figures in this report use the S-RIP colour definitions for reanalysis datasets (see *Appendix A*). Note, however, that some figures use different colours; thus, the readers should always refer to the legends of the figures to distinguish reanalyses by colours.

Chapter 1 - Introduction: The S-RIP motivation, goals, rationale, and report structure are described. See also *Fujiwara et al.* (2017a).

Chapter 2 - Description of the reanalysis systems: This chapter includes detailed descriptions of the forecast model, assimilation scheme, and observational data assimilated for each reanalysis. It also provides information on execution streams and archived data products. This chapter covers much of the same material as *Fujiwara et al.* (2017a), but in more detail and with some additions and corrections included. An extended electronic-only version of *Chapter 2* (denoted Chapter 2E) is also available as an online supplement to this report (through **https://s-rip.ees.hokudai.ac.jp/** and **https://s-rip.github.io**).

Chapter 3 - Overview of temperature and winds: This chapter evaluates major dynamical variables (*e.g.*, zonal mean temperature, zonal mean wind) of all the recent and past reanalyses on standard pressure levels. This evaluation uses monthly mean and 2.5° zonal mean data sets and spans the satellite era from 1979 - 2014.

The first key plot shows the homogeneity (or in many cases inhomogeneity) of each reanalysis with respect to pressure over the time period of the reanalysis. Then, key plots of the ensemble climatological means and individual reanalysis anomalies from these means are presented. Inter-reanalysis variations are quantified. The validation of this climatology is based on independent observations (*i.e.*, those not used in the reanalyses) such as non-assimilated radiosondes, rocketsondes, and non-assimilated satellite data. Additionally, the chapter presents how the more recent reanalyses have progressed over time to greater agreement among themselves, especially with the assimilation of GNSS-RO data. This chapter is an extended version of *Long et al.* (2017).

Chapter 4 - Overview of ozone and water vapour: This chapter includes a detailed evaluation of ozone and water vapour in the reanalyses, using a range of observational data sets obtained from both nadir and limb satellite instruments. The diagnostics considered include climatological evaluations such as monthly zonal mean cross-sections and altitude profiles, seasonal cycles, and interannual variability. Some more advanced diagnostics, such as the Quasi-Biennial Oscillation (QBO) and equivalent latitude timeseries, are used to better understand the differences in the climatological evaluations, while a detailed investigation of the transport processes resulting in these distributions is covered in the later, more process-oriented chapters. In addition, this chapter includes some summary information on the assimilated observations and on the modelling of ozone and water vapour in each reanalysis system. This chapter is an extended version of *Davis et al.* (2017).

Chapter 5 - Brewer-Dobson circulation: This chapter focuses on evaluation and comparison of the stratospheric circulation, using diagnostics based on the residual mean meridional mass streamfunction (*e.g.*, tropical upwelling), and stratospheric transport tracers such as the age-of-air (AoA). Off-line chemistry transport models in Eulerian and Lagrangian frameworks are used to compute tracer and trajectory diagnostics for more recent reanalyses. Results are compared to those from observation-based datasets derived from satellite, ground-based, balloon, and aircraft observations of long-lived tracers such as SF_6, CO_2, and N_2O. Particular attention is given to comparing past trends in AoA from the different reanalyses with several model simulations.

Chapter 6 - Extratropical stratosphere-troposphere coupling: This chapter covers the representation of dynamical coupling between the troposphere and stratosphere in the reanalyses. It focuses on the coupling between the stratospheric polar vortex and the troposphere on daily to intraseasonal time scales, and how this short-term variability is modulated on interannual time scales, *e.g.*, by El Niño Southern Oscillation (ENSO), the QBO, and stratospheric ozone loss. In particular, dynamical metrics associated with Sudden Stratospheric Warming (SSW) events are considered, including changes in heat and momentum fluxes, blocking events, the meridional circulation, and vertical coupling of the zonal mean circulation as characterized by the annular modes. In addition, alternative strategies for characterizing SSWs are considered. Reanalysis uncertainty, *i.e.*, the spread between reanalyses, is contrasted with sampling uncertainty associated with natural variability; in most cases, uncertainty in stratospheric-tropospheric coupling metrics is dominated by the latter. The utility of surface and conventional-input reanalyses is also explored.

Chapter 7 - Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS): This chapter evaluates the processes in the UTLS specifically in the extratropics. Only the most recent reanalyses have resolution adequate to represent many ExUTLS processes, so older reanalyses are not analyzed. Diagnostics include characterization of the tropopause based on different definitions (including multiple tropopauses, vertical structure, comparison of temperature-gradient based tropopause characteristics with radiosonde observations, *etc.*); UTLS jet characteristics and long-term changes; atmospheric transport from trajectory model calculations; and diagnostics of mixing and stratosphere-troposphere exchange (STE). In addition, assimilated UTLS ozone from the more recent reanalyses is evaluated, including diagnostics of dynamically-driven column ozone variations, evidence of STE and mixing, and the relationships of ozone diagnostics to the dynamical variability. This chapter also includes comparisons of assimilated UTLS ozone with satellite observations.

Chapter 8 - Tropical Tropopause Layer (TTL): This chapter evaluates the tropical transition region between the well-mixed, convective troposphere and the highly stratified stratosphere in the reanalyses. The general TTL structure, as given by the vertical temperature profile, tropopause levels, and the level of zero radiative heating, is analysed. Diagnostics related to clouds and convection in the TTL include cloud fraction, cloud water content, and outgoing longwave radiation. The chapter takes into account the diabatic heat budget as well as dynamical characteristics of the TTL such as Lagrangian cold points, residence times, and wave activity. Finally, the width of the tropical belt based on tropical and extra-tropical diagnostics and the representation of the South Asian Summer Monsoon in the reanalyses are evaluated.

Chapter 9 - Quasi-Biennial Oscillation (QBO): The diagnostics in this chapter include analysis of the tropical QBO in zonal wind and temperature, tropical waves and the QBO zonal momentum budget, and extra-tropical teleconnections of the QBO. Observations used for validation include operational and campaign radiosondes, and satellite observations from GNSS-RO, HIRDLS, SABER, COSMIC and AIRS.

Chapter 10 - Polar processes: This chapter focuses on microphysical and chemical processes in the winter polar lower stratosphere, such as polar stratospheric cloud (PSC) formation; denitrification and dehydration; heterogeneous chlorine activation and deactivation; and chemical ozone loss. These are "threshold" phenomena that depend critically on meteorological conditions.

A range of diagnostics is examined to quantify differences between reanalyses and their impact on polar process studies, including minimum lower stratospheric temperatures, area and volume of stratospheric air cold enough to support PSC formation, maximum latitudinal gradients in potential vorticity (a measure of the strength of the winter polar vortex), area of the vortex exposed to sunlight each day, vortex break-up dates, and polar cap average diabatic heating rates. For such diagnostics, the degree of agreement between reanalyses is an important direct indicator of the systems' inherent uncertainties, and comparisons to independent measurements are frequently not feasible. For other diagnostics, however, comparisons with atmospheric observations are very valuable. The representation of small-scale temperature and horizontal wind fluctuations and the fidelity of Lagrangian trajectory calculations are evaluated using observations obtained during long-duration superpressure balloon flights launched from Antarctica. Comparisons with satellite measurements of various trace gases and PSCs are made to assess the thermodynamic consistency between reanalysis temperatures and theoretical PSC equilibrium curves. Finally, to explore how the spatially and temporally varying differences between reanalyses interact to affect the conclusions of typical polar processing studies, simulated fields of nitric acid, water vapour, several chlorine species, nitrous oxide, and ozone from a chemistry-transport model driven by the different reanalyses for specific Arctic and Antarctic winters are compared to satellite measurements.

Chapter 11 - Upper Stratosphere and Lower Mesosphere (USLM): This chapter focuses on the uppermost levels in the reanalyses, where assimilated data sources are most sparse. The first part of the chapter includes a brief discussion of the effects of the model top and physical parameterizations relevant to the USLM. Long-term signatures of discontinuities in data assimilation and variability among reanalyses are then presented. A climatology of the basic state variables of temperature, horizontal winds, and residual circulation velocities is given. The climatology includes estimates of variability among the reanalyses. Annual cycles highlight the dependence of reanalysis difference on time of year. We then document dominant modes of variability in the reanalyses in the tropical regions and at high latitudes, and longer-term variability including solar cycle, volcanic, ENSO, and QBO signals. The tropical Semi-Annual Oscillation (SAO), the middle-atmosphere Hadley circulation, and the occurrence of inertial instability are compared among the reanalyses. High latitude processes considered include polar vortex variability and extreme disruptions therein observed during "elevated stratopause" events. Planetary wave amplitudes are quantified and compared to observations. The chapter ends with a comparison of solar atmospheric tides, 2-day wave amplitudes, and 5-day wave amplitudes in the USLM.

Chapter 12 - Synthesis summary: This chapter summarizes the key findings and the common patterns across the report, and provides suggestions as to the appropriateness of individual reanalyses for studies of particular atmospheric processes. It provides recommendations for future research and reanalysis development.

1.4 Development of the S-RIP team

The need for a coordinated reanalysis intercomparison project was proposed and discussed at the 8th SPARC Data Assimilation Workshop held at Brussels, Belgium in June 2011 (*Jackson and Polavarapu*, 2012; **Figure 1.2**), leading to the proposal of the SPARC Reanalysis Intercomparison Project (S-RIP) in January 2012 (*Fujiwara et al.*, 2012). In February 2012, S-RIP was officially endorsed by the SPARC Scientific Steering Group (SSG) as an emerging activity of SPARC. A first S-RIP session was held at the subsequent 9th SPARC Data Assimilation workshop in Socorro, New Mexico, USA, in June 2012 (*Jackson et al.*, 2013; **Figure 1.2**), followed by the formation of the scientific Working Group (11 members) and the confirmation of the reanalysis centre contacts (8 members) by August 2012. The Working Group then proceeded to discuss chapter titles, co-leads, and initial contributors to the final SPARC report, and organised an S-RIP Planning Meeting for the following year.

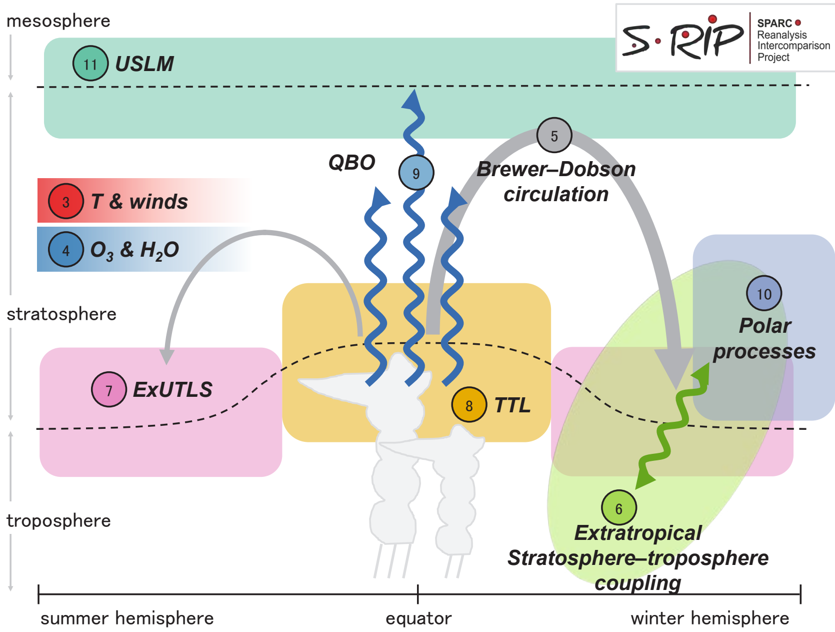


***Figure 1.1:** Schematic illustration of the atmosphere showing the processes and regions that are covered in this report. The numbers are the chapter numbers. Domains approximate the main focus areas of each chapter and should not be interpreted as strict boundaries. Chapters 3 and 4 cover the entire domain. (Updated from Fujiwara et al., 2017a.)*

Figure 1.2: *Photographs from (left) the 8th SPARC Data Assimilation Workshop at Brussels in 2011 and (right) the 9th SPARC Data Assimilation workshop at Socorro in 2012.*

The S-RIP Planning Meeting, with 39 participants, was hosted by David Jackson at the UK Met Office in Exeter, UK from 29 April to 1 May 2013 (*Fujiwara and Jackson*, 2013; **Figure 1.3**). The purpose of that meeting was to finalise the report outline, to determine the diagnostics list and observational data required for validation for each chapter, to agree on general guidelines and protocols, and to define the project timetable. The S-RIP Implementation Plan was submitted to the SSG in January 2014, at which point S-RIP was officially endorsed by the SSG as a full activity of SPARC. Since this official launch of S-RIP, two side meetings were held during the SPARC General Assemblies (Queenstown, New Zealand, in January 2014; and Kyoto, Japan, in October 2018), and four annual workshops were held (**Figure 1.4**). The 2014 workshop was hosted by Craig Long at the NOAA Center for Weather and Climate Prediction, College Park, Maryland, USA in September 2014 (*Errera et al.*, 2015). The 2015 workshop was hosted by Bernard Legras and held at Pierre and Marie Curie University, Paris, France in October 2015 (*Errera et al.*, 2016). The 2016 workshop was hosted by James Anstey and held at Victoria, Canada in October 2016 (*Fujiwara et al.*, 2017b). The 2017 workshop was hosted by Beatriz Monge-Sanz and Rossana Dragani at the ECMWF, Reading, UK in October 2017 (*McCormack et al.*, 2018). The 2014, 2015, 2016, and 2017 workshops were co-organized with Quentin Errera (and John McCormack for the 2017 one) and were held at the same place in the same week as the SPARC Data Assimilation workshops, with a one-day joint session. In June 2018, an S-RIP chapter-lead meeting was hosted by Gloria Manney and held at the NorthWest Research Associates (NWRA), Boulder, USA (**Figure 1.5**).

1.5 Management and communication

S-RIP was initially co-led by Masatomo Fujiwara (Japan), and David Jackson (UK), until April 2014 when Jackson stepped down. David Tan (UK), served as a co-lead between September 2014 and July 2015, working together with Masatomo Fujiwara. In October 2015, Jonathon Wright was assigned as a co-editor of the S-RIP Report with Masatomo Fujiwara. Since November 2015, S-RIP has been co-led by Masatomo Fujiwara, Gloria Manney (USA), and Lesley Gray (UK). The co-leads are members of a wider Working Group, who help steer the direction of the project and coordinate the specifics of the work. The Working Group members are David Tan (UK; until July 2015), Thomas Birner (USA, now in Germany), Simon Chabrillat (Belgium), Sean Davis (USA), Yulia Zyulyaeva (Russia; until October 2014), Michaela Hegglin (UK), Kirstin Krüger (Germany, now in Norway), Craig Long (USA), Susann Tegtmeier (Germany, now in Canada), Gloria Manney (USA), Lesley Gray (UK; since November 2015), and Masatomo Fujiwara (Japan).

Each reanalysis centre also has designated a contact who is involved in S-RIP and whose presence is vital to ensure the two-way flow of knowledge between researchers participating in S-RIP and the reanalysis centres. The reanalysis centre contacts are David Tan (ECMWF; until July 2015), Rossana Dragani (ECMWF; since July 2015), Craig Long (NOAA/NCEP), Wesley Ebisuzaki (NOAA/NCEP), Kazutoshi Onogi (JMA), Yayoi Harada (JMA), Steven Pawson (NASA; until April 2016), Krzysztof Wargan (NASA; since April 2016), Gilbert Compo (NOAA and University of Colorado), and Jeffrey Whitaker (NOAA).

Figure 1.3: *Figure 1.3. Photograph from the S-RIP Planning Meeting at Exeter in 2013.*

Figure 1.4: *Photographs from (top left) the 2014 S-RIP workshop at NOAA at College Park in the USA, (top right) the 2015 workshop at Paris, France, (bottom left) the 2016 workshop at Victoria, Canada, and (bottom right) the 2017 workshop at ECMWF at Reading, UK. These four workshops were held jointly with the SPARC Data Assimilation workshops, and these photographs were taken on the joint-session day.*

Each chapter of the report selected co-leads who organised the production of relevant diagnostics and the chapter writing, along with several contributors. The chapter co-leads are listed in **Table 1.2**.

The project has been monitored by the S-RIP co-leads via email communications with the chapter co-leads. Full S-RIP workshops have been held annually (see *Section 1.4*) to discuss emerging scientific results, the current status of each chapter, planning of evaluations, writing of papers, and completion of chapters. Individual chapter workshops were also held, usually jointly with other relevant workshops and conferences. The latest project information has been disseminated through the S-RIP website (at **https://s-rip.ees.hokudai.ac.jp**; being migrated to **https://s-rip.github.io**), constructed by Jonathon Wright and Masatomo Fujiwara, which includes a public section and an internal Wiki to facilitate the preparation of the report (see **Figure 1.6**). A virtual machine for data processing and a group workspace for storing data have been provided by the British Atmospheric Data Centre (BADC) of the UK Centre for Environmental Data Analysis (CEDA), as negotiated by James Anstey and Lesley Gray. S-RIP common grid files have been archived at a zenodo site (**https://zenodo.org/record/3754753**; **https://doi.org/10.5281/zenodo.3754753**). A CFSR/CFSv2 model level data set has been converted to netCDF format using the High Resolution Initial Conditions (HIC) binary files and forecast files that are archived by NOAA's NCEI (**https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/climate-forecast-system-version2-cfsv2**; the converted data files are available upon request from Sean M. Davis and Karen H. Rosenlof at NOAA). Zonal-mean data sets prepared for S-RIP are archived at the CEDA (Martineau *et al.*, 2018). These include dynamical variables and derived quantities prepared by Patrick Martineau (**https://doi.org/10.5285/b241a7f536a244749662360bd7839312**) as well as heating rates prepared by Jonathon Wright (**https://doi.org/10.5285/70146c789eda4296a3c3ab6706931d56**).

Quasi-monthly S-RIP news emails have been sent to the participants and other interested researchers to share the latest information relevant to the project and to keep the volunteer participants motivated. Following the discussion at the 2015 S-RIP workshop, in February 2016 a special issue on "The SPARC Reanalysis Intercomparison Project (S-RIP)" was launched in Atmospheric Chemistry and Physics (ACP), a journal of the European Geosciences Union (EGU).

Figure 1.5: *Photograph from the 2018 S-RIP chapter-lead meeting at NWRA at Boulder, USA.*

Table 1.2: *Chapter titles and co-leads.*

	Title	Co-leads
1	Introduction	Masatomo Fujiwara, Gloria Manney, Lesley Gray, Jonathon Wright
2	Description of the Reanalysis System	Jonathon Wright, Masatomo Fujiwara, Craig Long
3	Overview of Temperature and Winds	Craig Long, Masatomo Fujiwara
4	Overview of Ozone and Water Vapour	Michaela Hegglin, Sean Davis
5	Brewer-Dobson Circulation	Beatriz Monge-Sanz, Thomas Birner
6	Extratropical Stratosphere-Troposphere Coupling	Edwin Gerber, Patrick Martineau
7	Extra-tropical Upper Troposphere and Lower Stratosphere (ExUTLS)	Cameron Homeyer, Gloria Manney
8	Tropical Tropopause layer (TTL)	Susann Tegtmeier, Kirstin Krüger
9	Quasi-Biennial Oscillation (QBO)	James Anstey, Lesley Gray
10	Polar Processes	Michelle Santee, Alyn Lambert, Gloria Manney
11	Upper Stratosphere and Lower Mesosphere (USLM)	V. Lynn Harvey, John Knox
12	Synthesis Summary	Masatomo Fujiwara, Gloria Manney, Lesley Gray, Jonathon Wright

The editors of this special issue are Peter Haynes, Gabriele Stiller, and William Lahoz. Later (in January 2017), this special issue was extended to an inter-journal special issue in ACP and Earth System Science Data (ESSD), another journal produced by the EGU. Gabriele Stiller is the special-issue editor for ESSD. This special issue is one of the ways to encourage researchers to publish S-RIP related works. As of 9 November 2021, there are 48 published papers in this special issue.

1.6 Links to other projects

S-RIP has close links to several other SPARC activities, including the SPARC Data Assimilation Working Group (SPARC-DA), the SPARC Network on Assessment of Predictability (SNAP), SPARC Dynamical Variability (DynVar), the SPARC QBO initiative (QBOi), and the Observed Composition Trends and Variability in the UTLS (OCTAV-UTLS) activities. These activities share a common focus on stratospheric analyses and, in the case of SNAP, on the impacts of these analyses on weather forecasting. The reanalyses evaluated and compared by S-RIP are widely used to validate climate models, establishing a direct connection between the activities of S-RIP and those of the Chemistry-Climate Model Initiative (CCMI). S-RIP activities also overlap with several other SPARC activities, such as the Temperature Changes activity, the SPARC Data Initiative, and the Gravity Waves activity. The leaders of several of these activities are also involved in the S-RIP Working Group and/or serving as chapter co-leads or contributors in the preparation of the S-RIP report, thus enhancing opportunities for coordination and collaboration.

Figure 1.6: *A snapshot of the front page of the S-RIP website (7 May 2021).*

S-RIP has been publicised at meetings of the WMO Working Group on Numerical Experimentation (WGNE; **http://www.wmo.int/pages/prog/arep/wwrp/rescrosscut/resdept_wgne.html**; now at **http://wgne.meteoinfo.ru/**), where the project was well received and prompted discussion about a parallel WGNE activity focused on tropospheric reanalyses. In 2016, the WCRP Task Team for Intercomparison of ReAnalyses (TIRA; **https://reanalyses.org/atmosphere/wcrp-task-team-intercomparison-reanalyses-tira/**) was established under the WCRP Data Advisory Council (WDAC), and one of the S-RIP co-leads (Masatomo Fujiwara) became a member (and later, a co-lead) of TIRA as the SPARC liaison. Finally, activities associated with S-RIP have the potential to be important components of the WMO Global Framework for Climate Services (GFCS; **http://www.wmo.int/gfcs/**; now at **https://gfcs.wmo.int**).

Acknowledgements

We acknowledge the scientific guidance and sponsorship of the World Climate Research Programme, coordinated in the framework of SPARC. We are grateful to Theodore Shepherd and Greg Bodeker (past SPARC co-chairs) for strong encouragement and valuable advice during the proposal phase of the project during 2011 - 2012, and Joan Alexander (serving SPARC co-chair during 2013 - 2015), Neil Harris (serving SPARC co-chair since 2014), Judith Perlwitz (serving SPARC co-chair during 2016 - 2020), and Seok-Woo Son (serving SPARC co-chair since 2019) for continued support and encouragement. We are also very grateful to the review handling editors of this report, Judith Perlwitz, Seok-Woo Son, Vincent-Henri Peuch, Karen Rosenlof, and Gufran Beig, and to the reviewers. The Information Initiative Center of Hokkaido University, Japan, has hosted the S-RIP web server since 2014. We thank the reanalysis centres for providing their support and data products. The British Atmospheric Data Centre (BADC) of the UK Centre for Environmental Data Analysis (CEDA) has provided a virtual machine for data processing, a group workspace for data storage, and public data archive sites. We thank Yulia Zyulyaeva for contributions to S-RIP as an original member of the working group, as a chapter co-lead through October 2014, and as the designer of the S-RIP logo. We thank Diane Pendlebury for contributions to S-RIP as a chapter co-lead through August 2015. We thank Quentin Errera, the lead of the SPARC Data Assimilation working group, for co-organizing the 2014, 2015, 2016, and 2017 S-RIP workshops. We also thank John McCormack for co-organizing the 2017 workshop. Travel support for some participants of the 2013 planning meeting, the 2014, 2015, 2016, and 2017 workshops, and the 2018 chapter-lead meeting was provided by SPARC. We thank Peter Haynes and Gabriele Stiller for their work as guest editors for the special issue "The SPARC Reanalysis Intercomparison Project (S-RIP)" in the journals Atmospheric Chemistry and Physics (ACP) and Earth System Science Data (ESSD); we offer special thanks to William Lahoz (1960 - 2019) for his role as the other guest editor of this special issue, which he performed excellently even as he battled his final illness. We thank the hard works at the SPARC Office, in particular Mareike Heckl, Brigitte Ziegele, and Sabrina Zechlau for their typesetting of the report. Work at the Jet Propulsion Laboratory, California Institute of Technology, was carried out under a contract with the National Aeronautics and Space Administration.

References

Bengtsson, L. and J. Shukla, 1988: Integration of space and in situ observations to study global climate change. *Bull. Amer. Meteorol. Soc.*, **69**, 1130 - 1143, doi: 10.1175/1520-0477(1988)069<1130:IOSAIS>2.0.CO;2.

Davis, S.M., *et al.*, 2017: Assessment of upper tropospheric and stratospheric water vapour and ozone in reanalyses as part of S-RIP. *Atmos. Chem. Phys.*, **17**, 12743 - 12778, doi: 10.5194/acp-17-12743-2017.

Errera, Q., M. Fujiwara, C. Long, and D. Jackson, 2015: Report from the 10th SPARC data assimilation workshop and the 2014 SPARC Reanalysis Intercomparison Project (S-RIP) workshop in Washington DC, USA. *SPARC Newsletter*, **44**, 31 - 38, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Errera, Q., M. Fujiwara, and B. Legras, 2016: The 2015 S-RIP workshop and 11th SPARC data assimilation workshop. *SPARC Newsletter,* **47**, 12 - 19, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Fujiwara, M., S. Polavarapu, and D. Jackson, 2012: A proposal of the SPARC Reanalysis/Analysis Intercomparison Project. *SPARC Newsletter*, **38**, 14 - 17, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Fujiwara, M., and D. Jackson, 2013: SPARC Reanalysis Intercomparison Project (S-RIP) Planning Meeting, 29 April-1 May, 2013, Exeter, UK. *SPARC Newsletter*, **41**, 52-55, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Fujiwara, M., *et al.*, 2017a: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417-1452, doi:10.5194/acp-17-1417-2017, 2017a.

Fujiwara, M., *et al.*, 2017b, The 12th SPARC Data Assimilation Workshop and 2016 S-RIP Workshop. *SPARC Newsletter*, **48**, 41 - 44, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Jackson, D., and S. Polavarapu, 2012: Report on the 8th SPARC Data Assimilation Workshop, 20 - 22 June 2011, Brussels, Belgium. *SPARC Newsletter*, **38**, 9 - 13, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Jackson, D., G. Manney, and K. Minschwaner, 2013: Report on the 9th SPARC Data Assimilation Workshop in Socorro, NW, USA. *SPARC Newsletter*, **40**, 21 - 29, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Long, C.S., *et al.*, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14593 - 14629, doi: 10.5194/acp-17-14593-2017.

Martineau, P., J.S. Wright, N. Zhu, and M. Fujiwara, 2018: Zonal-mean data set of global atmospheric reanalyses on pressure levels. *Earth Syst. Sci. Data*, **10**, 1925 - 1941, doi: 10.5194/essd-10-1925-2018.

McCormack, J., *et al.*, 2018: The 2017 S-RIP workshop and the 13th SPARC data assimilation workshop. *SPARC Newsletter*, **50**, 26 - 29, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Randel, W., et al., 2004: The SPARC intercomparison of middle-atmosphere climatologies. *J. Climate*, **17**, 986 - 1003, doi: 10.1175/1520-0442(2004)017<0986:TSIOMC>2.0.CO;2.

SPARC, 2002: Intercomparison of middle atmosphere climatologies edited by W. Randel, M.-L. Chanin, and C. Michaut, *WCRP*-**116**, *WMO/TD*-No. **1142**, *SPARC Report* No. **3**, available at: **https://www.sparc-climate.org/publications/sparc-reports/.**

SPARC, 2010: Chemistry-climate model validation, edited by V. Erying, T. Shepherd, and D. Waugh, *WCRP*-**30**, *WMO/TD*-No. **40**, *SPARC Report* No. **5**, available at: **https://www.sparc-climate.org/publications/sparc-reports/.**

Trenberth, K.E., and J.G. Olson, 1988: An evaluation and intercomparison of global analyses from the National Meteorological Center and the European Centre for Medium-Range Weather Forecasts. *Bull. Amer. Meteorol. Soc.*, **69**, 1047 - 1057, doi: 10.1175/1520-0477(1988)069<1047:AEAIOG>2.0.CO;2.

Appendix A: S-RIP colour definitions

Many figures in this report use the S-RIP colour definitions for reanalysis datasets shown Table A.1. Note that some figures use different colours; thus, the readers should always refer to the legends of the figures to distinguish reanalyses by colours.

Table A.1: *The S-RIP colour definitions for reanalysis and other datasets*

Reanalyses	Red, Green, Blue	Hexadecimal	Notes
MERRA-2	226, 31, 38	#E21F26	
MERRA	246, 153, 153	#F69999	
ERA-Interim	41, 95, 138	#295F8A	
ERA5	95, 152, 198	#5F98C6	
ERA-40	175, 203, 227	#AFCBE3	
JRA-55	114, 59, 122	#723B7A	
JRA-55C, JRA-55AMIP	173, 113, 181	#AD71B5	See *Chapter 2*
JRA-25/JCDAS	214, 184, 218	#D6B8DA	
NCEP-NCAR R1	245, 126, 32	#F57E20	
NCEP-DOE R2	253, 191, 110	#FDBF6E	
20CR v2c	236, 0, 140	#EC008C	See *Chapter 2*
20CR v2	247, 153, 209	#F799D1	See *Chapter 2*
CERA-20C	0, 174, 239	#00AEEF	
ERA-20C	96, 200, 232	#60C8E8	
CFSR/CFSv2	52, 160, 72	#34A048	include CFSv2 if post-2010 data are included
REM	179, 91, 40	#B35B28	reanalysis ensemble mean
Other	255, 215, 0	#FFD700	
Observations	0, 0, 0	#000000	observations - black
Other observations	119, 119, 119	#777777	observations - grey

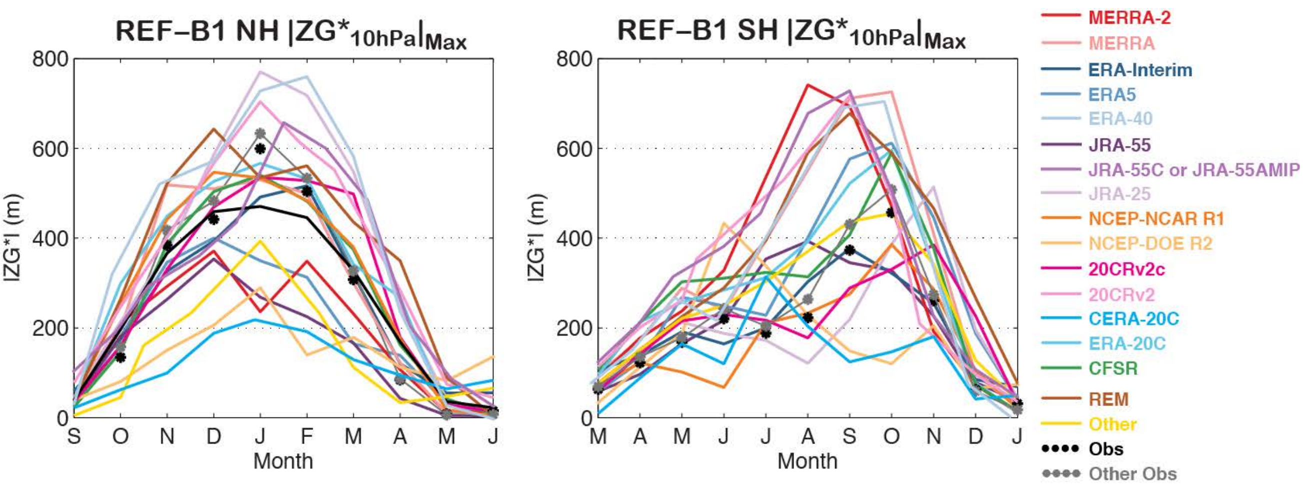


Figure A.1: *An example showing all the colours in Table A.1.*

Major abbreviations and terms

20CR	20th Century Reanalysis
AIRS	Atmospheric Infrared Sounder
AoA	Age of Air
BADC	British Atmospheric Data Centre
CCMI	Chemistry-Climate Model Initiative
CEDA	Centre for Environmental Data Analysis
CERA-20C	ECMWF 10-member ensemble of coupled climate reanalyses of the 20th century
CFSR	Climate Forecast System Reanalysis of the NCEP
CFSv2	Climate Forecast System version 2
CIRES	Cooperative Institute for Research in Environmental Sciences (NOAA and University of Colorado Boulder)
COSMIC	Constellation Observing System for Meteorology Ionosphere and Climate
DOE	Department of Energy
DynVAR	Dynamical Variability
ECMWF	European Centre for Medium-Range Weather Forecasts
EGU	European Geosciences Union
ENSO	El Niño Southern Oscillation
ERA-20C	ECMWF 20th century reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA-Interim	ECMWF interim reanalysis
ERA5	the fifth major global reanalysis produced by ECMWF
ExUTLS	Extra-tropical Upper Troposphere and Lower Stratosphere
GFCS	Global Framework for Climate Services
GNSS-RO	Global Navigation Satellite System Radio Occultation
HIC	High Resolution Initial Conditions
HIRDLS	HIgh Resolution Dynamics Limb Sounder
JCDAS	JMA Climate Data Assimilation System
JMA	Japan Meteorological Agency
JRA-25	Japanese 25-year Reanalysis
JRA-55	Japanese 55-year Reanalysis
JRA-55AMIP	Japanese 55-year Reanalysis based on AMIP-type simulations
JRA-55C	Japanese 55-year Reanalysis assimilating Conventional observations only
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NCEP-DOE R2	Reanalysis 2 of the NCEP and DOE
NCEP-NCAR R1	Reanalysis 1 of the NCEP and NCAR
netCDF	Network Common Data Form
NOAA	National Oceanic and Atmospheric Administration
R1	Reanalysis 1 of the NCEP and NCAR
R2	Reanalysis 2 of the NCEP and DOE
OCTAV-UTLS	NorthWest Research Associates
NWRA	NorthWest Research Associates
PSC	Polar Stratospheric Cloud

QBO	Quasi-Biennieal Oscillation
QBOi	QBO initiative
REM	Reanalysis Ensemble Mean
SABER	Sounding of the Atmosphere using Broadband Emission Radiometry
SAO	Semi-Annual Oscillation
SNAP	SPARC Network on Assessment of Predictability
SPARC	Stratosphere-troposphere Processes And their Role in Climate
SPARC-DA	SPARC Data Assimilation working group
S-RIP	SPARC Reanalysis Intercomparison Project
SSG	Scientific Steering Group
SSWs	Sudden Stratospheric Warmings
STE	Stratosphere-Troposphere Exchange
TIRA	Task Team for Intercomparison of ReAnalyses of the WCRP
TTL	Tropical Tropopause Layer
USLM	Upper Stratosphere and Lower Mesosphere
UTLS	Upper Troposphere and Lower Stratosphere
WCRP	World Climate Research Programme
WDAC	WCRP Data Advisory Council
WGNE	WMO Working Group on Numerical Experimentation
WMO	World Meteorological Organization

Chapter 2: Description of the Reanalysis Systems

Chapter lead authors

Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Craig Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA

Co-authors

James Anstey	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada	Canada
Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Gilbert P. Compo	(1) Cooperative Institute for Research in Environmental Sciences, University of Colorado (2) Physical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Rossana Dragani	European Centre for Medium-Range Weather Forecasts	UK
Wesley Ebisuzaki	National Oceanic and Atmospheric Administration	USA
Yayoi Harada	Japan Meteorological Agency	Japan
Chiaki Kobayashi	Japan Meteorological Agency	Japan
Will McCarty	National Aeronautics and Space Administration	USA
Andrea Molod	National Aeronautics and Space Administration	USA
Kazutoshi Onogi	Research Center for Advanced Science and Technology, the University of Tokyo *previously at Japan Meteorological Agency*	Japan
Steven Pawson	National Aeronautics and Space Administration	USA
Adrian Simmons	European Centre for Medium-Range Weather Forecasts	UK
David G.H. Tan	European Centre for Medium-Range Weather Forecasts *(retired)*	UK
Susann Tegtmeier	University of Saskatchewan	Canada
Krzysztof Wargan	(1) National Aeronautics and Space Administration (2) Science Systems and Applications, Inc.	USA
Jeffrey S. Whitaker	National Oceanic and Atmospheric Administration	USA
Cheng-Zhi Zou	National Oceanic and Atmospheric Administration	USA

Abstract. Information on key components of twelve global atmospheric reanalysis systems with output data available in 2018 is summarized, including brief descriptions of the forecast models, assimilation schemes, and observational data used in these systems. Details of the execution streams and archived data products are also provided. Tables are used extensively to facilitate comparison of different reanalysis systems, and are arranged so that readers interested in one or more systems can easily find and compare relevant information. The information in this chapter will be referred to in the interpretation of results presented in the other chapters of this S-RIP report. This chapter is not intended to provide a complete description of the reanalysis systems; readers requiring further details are encouraged to refer to the cited literature and the online documentation provided for each system. A condensed version of the material in this chapter has been provided by Fujiwara *et al.* (2017). A longer and more detailed version (denoted Chapter 2E) is provided as an electronic file on the S-RIP website at **https://s-rip.ees.hokudai.ac.jp** (being migrated to **https://s-rip.github.io**).

Fujiwara *et al.* (2017) have published a shortened version of this chapter.

Content

2.1 Introduction 17
2.2 Forecast models 20
2.2.1 Summary of basic information 20
2.2.2 Major physical parameterizations 22
2.2.3 Boundary and other specified conditions 27
2.2.3.1 Sea surface temperature and sea ice 30
2.2.3.2 Ozone 30
2.2.3.3 Aerosols 30
2.2.3.4 Carbon dioxide and other radiatively active gases 30
2.2.3.5 Solar cycle 31
2.2.4 Surface air and land surface treatments 31
2.3 Assimilation Schemes 35
2.3.1 Basics of data assimilation 35
2.3.2 Data assimilation in reanalysis systems 38
2.4 Observational Data 40
2.4.1 Summary of basic information 40
2.4.2 Quality control procedures 47
2.4.3 Summary of key upper air observations and known issues 49
2.4.3.1 Radiosonde data 49
2.4.3.2 Satellite data 51
2.4.3.3 Aircraft data 54
2.4.4 Water vapour 54
2.5 Execution streams 55
2.5.1 What is an 'execution stream'? 55
2.5.2 Summary of stream execution 55
2.6 Archived data 57
References 59
Appendix A: Vertical levels of the models 69
A2.1 ERA-40 and ERA-Interim 69
A2.2 ERA-20C 70
A2.3 ERA5 70
A2.4 JRA-25/JCDAS 72
A2.5 JRA-55 72
A2.6 MERRA and MERRA-2 73
A2.7 NCEP-NCAR R1 and NCEP-DOE R2 73
A2.8 CFSR 74
A2.9 20CR 74
Major abbreviations and terms 75

2.1 Introduction

An atmospheric reanalysis system consists of a global forecast model, input observations, and an assimilation scheme that blends input observations with short-range forecasts. These systems produce global atmospheric data that represents best estimates (analyses) of past atmospheric states. The information collected in these analyses is then propagated forward in time and space by subsequent forecasts. In this chapter, we provide summary descriptions of the key components of the twelve global atmospheric reanalysis systems listed in **Table 2.1**. Our descriptions of these systems are by necessity incomplete. Further details may be found in the cited literature, particularly the publications listed in **Table 2.1**, or in the technical documentation compiled and provided by the reanalysis centres. A list of the acronyms used in this chapter is provided in the *Appendix* at the end of this chapter.

We classify reanalysis systems according to their observational inputs and temporal coverage. The three classes of reanalysis systems include “full input” systems (which assimilate surface and upper-air conventional and satellite data), “conventional input” systems (which assimilate surface and upper air conventional data but do not assimilate satellite data), and “surface input” systems (which assimilate surface data only). Some reanalysis centres also provide companion “AMIP-type” simulations, which do not assimilate any observational data and are constrained by applying observed sea surface temperatures, sea ice, and other boundary or forcing conditions on the atmospheric forecast model. We also broadly distinguish reanalyses of the “satellite era” (1979 - present) and reanalyses that provide data for dates before January 1979, with the latter referred to as “extended” reanalyses. All reanalyses are affected by changes in assimilated observations, as discussed below, but such temporal inconsistencies are especially important to keep in mind for extended reanalyses that assimilate satellite data during the later part of the record.

Four reanalyses produced by ECMWF are considered: ERA-40, ERA-Interim, ERA-20C, and ERA5. ERA-40 (*Uppala et al.*, 2005) is an extended full input reanalysis covering 45 years from September 1957 through August 2002. No satellite data were assimilated for dates prior to January 1973; ERA-40 is therefore a conventional input reanalysis from September 1957 through December 1972. ERA-40 represented an important improvement relative to the first generation of modern reanalysis systems and continues to be used in many studies that require long-term atmospheric data. ERA-Interim (*Dee et al.*, 2011) is a full input reanalysis of the satellite era (1979 -present) that applies several corrections and modifications to the system used for ERA-40. Major focus areas during the production of ERA-Interim included improving the representations of the hydrologic cycle and the stratospheric circulation relative to ERA-40, as well as improving the consistency of the reanalysis products in time. ERA5 (*Hersbach et al.*, 2020) is intended as the full input replacement for ERA-Interim, with finer resolution in time and space (see also *Section 2.2* and *Appendix A*) and the ability to assimilate several new types of observational data (see also *Section 2.4*). ERA5 is an extended reanalysis covering 1950 to present, and the first full input reanalysis to be conducted together with an ensemble of data assimilations, which allows for a more robust characterization of uncertainty in the analysis state. Some ERA5 data have been available since July 2018, ERA5 data from 1979 onward have been available since January 2019, and a preliminary version of ERA5 1950 - 1978 data have been available since November 2020. Products from ERA5 are evaluated in some chapters of this report. While ERA5 could not be included in the interim version of this chapter (*Fujiwara et al.*, 2017), we document its structure here in tandem with the other reanalysis systems considered by S-RIP. ERA-20C (*Poli et al.*, 2016) is a surface input reanalysis of the twentieth century (1900 - 2010). ERA-20C directly assimilates only surface pressure and surface wind observations, and can therefore generate reanalyses of the atmospheric state that extend further backward in time. Data from ERA-20C extend up to 0.01 hPa, but the lack of upper-air observational constraints means that these data should be used with caution in the upper troposphere and above. We omit the earlier ECMWF reanalysis products FGGE (*Bengtsson et al.*, 1982) and ERA-15 (*Gibson et al.*, 1997), as well as recent coupled atmosphere–ocean reanalysis efforts at ECMWF using the CERA data assimilation system (*Laloyaux et al.*, 2016).

Two reanalyses produced by JMA and cooperating institutions are considered: JRA-25/JCDAS and JRA-55. JRA-25 (*Onogi et al.,* 2007), a joint effort by JMA and CRIEPI, was the first reanalysis produced using the JMA forecast model and data assimilation system. This reanalysis originally covered 25 years from 1979 through 2004, and was extended an additional 10 years (through the end of January 2014) as JCDAS using an identical system. JRA-55 (*Kobayashi et al.*, 2015) is an extended full input reanalysis with coverage from 1958 through the present. JRA-55 is the first reanalysis system to apply a 4D-Var data assimilation scheme (see *Section 2.3*) to upper-air data during the pre-satellite era (note however that ERA-20C has also used 4D-Var to assimilate surface observations during the pre-satellite era, while extension of ERA5 backward in time to 1950 has recently been completed). Along with the JRA-55 reanalysis, JMA has provided two companion products: JRA-55C (*Kobayashi et al.*, 2014), a conventional input reanalysis that excludes satellite observations from the assimilation, and JRA-55AMIP, an ensemble of AMIP-type forecast model simulations without data assimilation.

Two full input reanalyses produced by NASA GMAO are considered: MERRA and MERRA-2. MERRA (*Rienecker et al.*, 2011) was conceived by NASA GMAO as a reanalysis of the satellite era (starting in January 1979), with particular focus on leveraging the large amounts of data produced by NASA's Earth Observing System (EOS) satellite constellation and improving the representations of the water and energy cycles relative to earlier reanalyses. MERRA production was discontinued after February 2016. Motivated by the inability of the MERRA system to ingest some recent data types, GMAO has developed the follow-on reanalysis MERRA-2 (*Gelaro et al.*, 2017). MERRA-2, which covers 1980 - present, includes substantial upgrades to the model (*Molod et al.*, 2015) and changes to the data assimilation system and input data (*McCarty et al.*, 2016). Several new data sources are used that were not assimilated by MERRA, including hyperspectral radiances from IASI and CrIS, microwave radiances from ATMS, MLS temperature and ozone profiles, and GNSS-RO bending angles. One significant and unique feature of MERRA-2 is the assimilation of aerosol optical depth observations (*Randles et al.*, 2017; *Buchard et al.*, 2017), with analysed aerosols fed back to the forecast model radiation scheme. An earlier NASA reanalysis (*Schubert et al.*, 1993; *Schubert et al.*, 1995) covering 1980–1995 was produced by NASA's DAO (now GMAO) using the GEOS-1 data assimilation system; this reanalysis is no longer publicly available and is not included in the S-RIP intercomparison.

Four reanalyses produced by NOAA and cooperating organizations are considered: NCEP–NCAR R1, NCEP–DOE R2, CFSR/CFSv2, and NOAA–CIRES 20CR. NCEP–NCAR R1 (*Kalnay et al.*, 1996; *Kistler et al.*, 2001) was the first modern reanalysis system with extended temporal coverage (1948 - present). This system, which uses a modified 1995 version of the NCEP forecast model, remains in widespread use. NCEP–DOE R2 covers the satellite era (1979 - present) using essentially the same model, but corrects some important errors and limitations (*Kanamitsu et al.*, 2002). More recently, NCEP has produced CFSR using a 2007 version of the NCEP forecast model (*Saha et al.*, 2010). CFSR contains a number of improvements relative to *R1* and *R2* in both the forecast model and data assimilation system, including higher horizontal and vertical resolutions, more sophisticated model physics, and the ability to assimilate satellite radiances directly (rather than temperature retrievals). CFSR was also the first coupled global reanalysis of the atmosphere–ocean–sea ice system. Production of CFSR was transitioned to a newer version of the NCEP data assimilation system (CFSv2; *Saha et al.*, 2014) on 1 January 2011. This transition from CFSR to CFSv2 should not be confused with the transfer of CFSv2 production from NCEP EMC to NCEP operations, which occurred at the start of April 2011. The model used for CFSv2 has a different horizontal resolution and includes minor changes to physical parameterizations. Because CFSv2 has been touted as a continuation of CFSR, we treat CFSR and CFSv2 as a paired system in this chapter, including brief descriptions of differences between the original and updated systems where relevant. However, we note that subsequent chapters of this report document many significant differences between CFSR and CFSv2, and suggest that users of these products should be cautious when conducting studies that span the 1 January 2011 transition date (see also *Section 2.5*). NOAA–CIRES 20CR (*Compo et al.*, 2011) is the first reanalysis to span more than 100 years. Like ERA-20C, 20CR is a surface input reanalysis. Unlike ERA-20C, which uses a 4D-Var approach to assimilate both surface pressure and surface winds, 20CR uses an EnKF approach (see *Section 2.3*) and assimilates only surface pressure data. The forecast model used in 20CR is similar in many ways to that used in CFSR, but with much coarser vertical and horizontal grids. Because of its relatively coarse vertical resolution (see *Appendix A*) and the lack of direct observational inputs in the upper atmosphere, output from 20CR should be used with care, particularly in the upper troposphere and above. Although two updated versions of 20CR (20CRv2c and 20CRv3; see *Slivinski et al.*, 2019) have been released since the beginning of the S-RIP activity, this report focuses on the earlier 20CRv2 (*Compo et al.*, 2011) unless otherwise indicated.

The influence of observational data on reanalysis products differs not only by the type of reanalysis (*e.g.*, "full input" versus "surface input"), but also by variable (see, *e.g.*, the variable classification proposed by *Kistler et al.*, 2001). Atmospheric temperatures, horizontal winds, and geopotential heights are strongly influenced by the assimilation of observational data even in earlier reanalysis systems, although these variables may be determined mainly by the forecast model in regions or periods where observations are sparse or uncertain. Observational constraints on tropospheric water vapour are weaker but still influential, and some recent reanalysis systems assimilate data that establish constraints on ozone, total water, precipitation, and/or aerosol optical depth. Variables that are largely determined by the forecast model or surface boundary conditions (such as surface fluxes and tendency terms for heat, moisture, and momentum) are considered less reliable and should be used with caution and/or validated against independent estimates.

The SPARC community has particular interest in upper tropospheric and stratospheric ozone and water vapour. This chapter touches briefly on the treatment of these variables, with detailed intercomparisons deferred to *Chapter 4*. Many reanalysis systems simulate ozone using photochemistry schemes of varying complexity and assimilate satellite ozone retrievals during the period after 1979. Some reanalysis systems provide an ozone analysis but use a climatological ozone distribution for radiation calculations in the forecast model. Additional details regarding the treatment of ozone are provided in **Table 2.11**. Reanalysis estimates of stratospheric water vapour are rudimentary and often unreliable. Adjustments due to data assimilation are typically suppressed above a specified upper boundary that varies by reanalysis system, and are in several cases replaced by relaxation to a constant value or zonal mean climatology. Stratospheric air is dehydrated mainly at the tropical tropopause and transported and diffused from there, with only a few systems attempting to represent the source of water vapour due to methane oxidation (see **Table 2.24** for further details).

Table 2.1: *List of global atmospheric reanalysis systems considered in this report.*

Reanalysis system	Reference	Description
ERA-40	*Uppala et al.* (2005)	Class: full input; extended Centre: ECMWF Coverage: September 1957 to August 2002
ERA-Interim	*Dee et al.* (2011)	Class: full input; satellite era Centre: ECMWF Coverage: January 1979 to August 2019.
ERA-20C	*Poli et al.* (2016)	Class: surface input; extended Centre: ECMWF Coverage: January 1900 to December 2010 *Note: A companion ensemble of AMIP-style simulations (ERA-20CM; Hersbach et al., 2015) is also available.*
ERA5	*Hersbach et al.* (2020)	Class: full input; extended Centre: ECMWF Coverage: currently January 1979 to present; a preliminary version of extension backward in time to January 1950 has also been released. *Note: ERA5.1, a rerun covering 2000–2006, has been conducted to address a cold bias in the lower stratosphere during this period.*
JRA-25 / JCDAS	*Onogi et al.* (2007)	Class: full input; satellite era Centre: JMA and CRIEPI Coverage: January 1979 to January 2014 *Note: January 2005 through January 2014 are from JCDAS, a real-time extension of JRA-25.*
JRA-55	*Kobayashi et al.* (2015); *Harada et al.* (2016)	Class: full input; extended Centre: JMA Coverage: January 1958 to present *Note: Two ancillary products are also available: JRA-55C (a conventional input reanalysis covering November 1972 to December 2012; see Kobayashi et al., 2014) and JRA-55AMIP (which assimilates no observational data but uses the same boundary conditions as JRA-55).*
MERRA	*Rienecker et al.* (2011)	Class: full input; satellite era Centre: NASA GMAO Coverage: January 1979 to February 2016
MERRA-2	*Gelaro et al.* (2017)	Class: full input; satellite era Centre: NASA GMAO Coverage: January 1980 to present
NCEP-NCAR R1	*Kalnay et al.* (1996); *Kistler et al.* (2001)	Class: full input; extended Centre: NOAA/NCEP and NCAR Coverage: January 1948 to present
NCEP-DOE R2	*Kanamitsu et al.* (2002)	Class: full input; satellite era Centre: NOAA/NCEP and the DOE AMIP-II project Coverage: January 1979 to present
CFSR / CFSv2	*Saha et al.* (2010); *Saha et al.* (2014)	Class: full input; satellite era Centre: NOAA/NCEP Coverage: January 1979 to present *Note: Official data coverage by CFSR (CDAS-T382) extends through December 2010; production was migrated to the CFSv2 (CDAS-T574) analysis system starting from 1 January 2011. Although it has a different horizontal resolution (Table 2.2) and includes minor changes to physical parameterizations, CFSv2 can be considered as a continuation of CFSR for most purposes.*
NOAA-CIRES 20CR v2	*Compo et al.* (2011)	Class: surface input; extended Centre: NOAA and the University of Colorado CIRES Coverage: November 1869 to December 2012 *Note: Updated versions of 20CR covering 1851–2011 (20CR version 2c, released in 2015) and 1836–2015 (20CR version 3, released in 2019) have been completed and made available, but are not documented in this chapter. See Slivinski et al. (2019) for details.*

2.2 Forecast models

2.2.1 Summary of basic information

Table 2.2 provides a summary of key information regarding the forecast models used in each reanalysis, including the analysis system, the horizontal grid, and the number of levels in the vertical coordinate. The forecast models and data assimilation systems used in reanalyses are typically frozen versions of operational systems for numerical weather prediction. The atmospheric model used in a reanalysis thus often has much in common with the model used for operational numerical weather forecasting at the same forecasting centre around the time that reanalysis was started. Model names and generations are listed in the second column of **Table 2.2**.

The information on horizontal grids provides a rough idea of the finest horizontal scales represented by the models. We describe the horizontal grid structures of models that use spectral dynamical cores (*e.g.*, *Machenhauer*, 1979) using two separate notations. All of the models considered here use spectral dynamical cores except for MERRA and MERRA-2. Regular Gaussian grids are denoted by Fn and Tk. Fn refers to a regular Gaussian grid with 2n latitude bands and (in most cases) 4n longitude bands, while Tk indicates horizontal truncation at wave number k in the spectral dynamical core. The longitude grid spacing in a standard Fn regular Gaussian grid is 90°/n, so that the geographical distance between neighbouring grid cells in the east–west direction shrinks toward the poles. *R1*, *R2*, and 20CR use modified regular Gaussian grids with 4(n+1) longitude bands and longitude spacings of 90°/(n+1). Linear reduced Gaussian grids (*Hortal and Simmons*, 1991; *Courtier and Naughton*, 1994) are denoted by Nn and TLk, where the latter again indicates truncation at horizontal wave number k. The number of latitude bands in the Nn reduced Gaussian grid is also 2n, but the number of longitudes per latitude circle decreases from the equator (where it is 4n) toward the poles. Longitude grid spacing in reduced Gaussian grids is therefore quasi-regular in distance rather than degrees (**Table 2.2**). More details on Gaussian grids are available at **https://confluence.ecmwf.int/display/FCST/Gaussian+grids** (accessed 5 June 2020). Unlike the other reanalysis systems discussed in this chapter, the MERRA and MERRA-2 atmospheric models use finite volume dynamical cores. MERRA applied this dynamical core on a regular latitude–longitude grid (*Lin*, 2004), while MERRA-2 uses a cubed-sphere grid (*Putman and Lin*, 2007). The latter type of grid is denoted by Cn, following a similar convention as Fn and Nn (*i.e.*, approximately 4n longitude bands along the equator).

Table 2.3 lists the vertical locations of the model tops and describes special treatments applied in the uppermost layers of each model. Common special treatments include the use of a diffusive 'sponge layer' near the model top. Sponge layers mitigate the effects of the finite 'lid height' that must be assumed in numerical models of the atmosphere. The application of enhanced diffusion in a sponge layer damps upward propagating waves as they near the model top, thereby preventing unphysical reflection of wave energy at the model top that would in turn introduce unrealistic resonance in the model atmosphere (*Lindzen et al.*, 1968). It is worth noting, however, that diabatic heating and momentum transfer associated with the absorption of wave energy by sponge layers and other simplified representations of momentum damping (such as Rayleigh friction; see, *e.g.*, *Holton and Wehrbein*, 1980) may still introduce spurious behaviour in model representations of middle atmospheric dynamics (*Shepherd and Shaw*, 2004; *Shepherd et al.*, 1996). Most of the forecast models used by reanalysis systems include a sponge layer, but the formulation of this layer varies. The models that do not, such as that used to produce NCEP-NCAR R1, are known to include spurious wave reflection from the model top that affects their performance in the upper atmosphere.

Table 2.2: *Basic details of the forecast models used in the reanalyses. Horizontal grid spacing is expressed in degrees for regular grids and in kilometres for reduced grids.*

Reanalysis system	Model	Horizontal grid	Vertical grid
ERA-40	IFS Cycle 23r4 (2001)	N80: ~125 km (TL159)	60 (hybrid σ–p)
ERA-Interim	IFS Cycle 31r2 (2007)	N128: ~79 km (TL255)	60 (hybrid σ–p)
ERA-20C	IFS Cycle 38r1 (2012)	N80: ~125 km (TL159)	91 (hybrid σ–p)
ERA5	IFS Cycle 41r2 (2016)	N320: ~31 km (TL639)	137 (hybrid σ–p)
JRA-25 / JCDAS	JMA GSM (2004)	F80: 1.125°(T106)	40 (hybrid σ–p)
JRA-55	JMA GSM (2009)	N160: ~55 km (TL319)	60 (hybrid σ–p)
MERRA	GEOS 5.0.2 (2008)	1/2° latitude, 2/3° longitude	72 (hybrid σ–p)
MERRA-2	GEOS 5.12.4 (2015)	C180: ~50 km (cubed sphere)	72 (hybrid σ–p)
NCEP-NCAR R1	NCEP MRF (1995)	F47: 1.875° (T62)	28 (σ)
NCEP-DOE R2	Modified MRF (1998)	F47: 1.875° (T62)	28 (σ)
CFSR **CFSv2**	NCEP CFS (2007) NCEP CFS (2011)	F288: 0.3125° (T382) F440: 0.2045° (T574)	64 (hybrid σ–p) 64 (hybrid σ–p)
NOAA-CIRES 20CR v2	NCEP GFS (2008)	F47: 1.875° (T62)	28 (hybrid σ–p)

All of the reanalysis systems discussed in this chapter use hybrid σ–p vertical coordinates (*Simmons and Burridge*, 1981), with the exception of NCEP-NCAR R1 and NCEP-DOE R2, which use σ vertical coordinates. The number of vertical levels ranges from 28 (*R1*, *R2*, and 20CR) to 137 (ERA5), and top levels range from 3 hPa (*R1* and *R2*) to 0.01 hPa (MERRA, MERRA-2, ERA5, and ERA-20C).

Table 2.3: *Model top levels and special dynamical treatments applied in the uppermost model levels.*

Reanalysis system	Top level	Special treatment of uppermost levels
ERA-40	0.1 hPa	A sponge layer is applied at pressures less than 10 hPa by adding an additional function to the horizontal diffusion terms. This function, which varies with wavenumber and model level, acts as an effective absorber of vertically-propagating gravity waves. Rayleigh friction is also implemented at pressures less than 10 hPa.
ERA-Interim	0.1 hPa	Same as ERA-40.
ERA-20C	0.01 hPa	Similar to ERA-Interim, but an additional first order 'mesospheric' sponge layer is implemented at pressures less than 1 hPa. As in ERA-40 and ERA-Interim, Rayleigh friction is still applied at pressures less than 10 hPa, but the coefficient is reduced to account for the inclusion of parameterized non-orographic gravity wave drag (**Table 2.6**).
ERA5	0.01 hPa	Similar to ERA-20C, but Rayleigh friction is no longer applied.
JRA-25 / JCDAS	0.4 hPa	A sponge layer is applied by gradually enhancing horizontal diffusion coefficients with increasing height at pressures less than 100 hPa. Rayleigh damping is applied to temperature deviations from the global average on each of the uppermost three levels.
JRA-55	0.1 hPa	Sponge layer treatment is similar to JRA-25, but with Rayleigh friction implemented at pressures less than 50 hPa.
MERRA	0.01 hPa	A sponge layer consisting of the nine uppermost model levels (pressures less than ~0.24 hPa) is implemented by increasing the horizontal divergence damping coefficient (see also **Table 2.7**). Advection at the top model level is reduced to first order.
MERRA-2	0.01 hPa	Same as MERRA.
NCEP-NCAR R1	3 hPa	No sponge layer or other special treatment.
NCEP-DOE R2	3 hPa	No sponge layer or other special treatment.
CFSR / CFSv2	~0.266 hPa	Linear Rayleigh damping with a time scale of 5 days is applied at pressures less than ~2 hPa. The horizontal diffusion coefficient also increases with scale height throughout the atmosphere.
NOAA-CIRES 20CR v2	~2.511 hPa	No sponge layer or other special treatment.

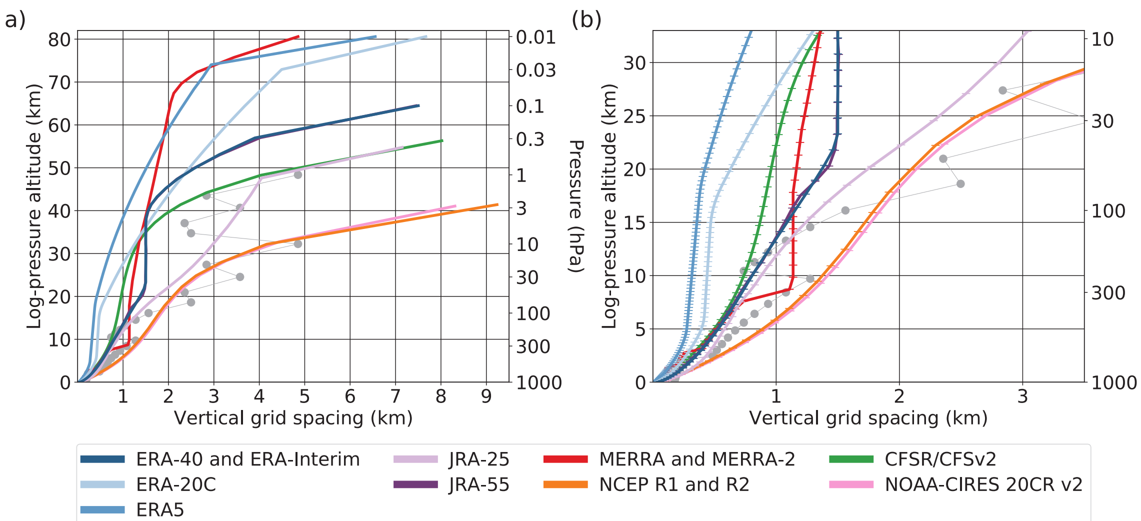


Figure 2.1: *Approximate vertical resolutions of the reanalysis forecast models for (a) the full vertical range of the reanalyses and (b) the surface to 33 km (~10 hPa). Altitude and vertical grid spacing are estimated using log-pressure altitudes (z* = H ln[p0/p]), where the surface pressure p0 is set to 1000 hPa and the scale height H is set to 7 km. The grid spacing indicating the separation of two levels is plotted at the altitude of the upper of the two levels, so that the highest altitude shown in (a) indicates the height of the top level. Some reanalyses use identical vertical resolutions; these systems are listed together in the legend. Other reanalyses have very similar vertical resolutions when compared with other systems, including JRA-55 (similar but not identical to ERA-40 and ERA-Interim) and 20CR (similar but not identical to R1 and R2). Approximate vertical spacing associated with the isobaric levels on which ERA-40 and ERA-Interim reanalysis products are provided (grey discs) is shown in both panels for context. Reproduced from Fujiwara et al. (2017).*

Figure 2.1 shows approximate vertical resolutions for the reanalysis systems in log-pressure altitude, assuming a scale height of 7 km and a surface pressure of 1000 hPa. A number of key differences are evident, including large discrepancies in the height of the top level (**Figure 2.1a**) and variations in vertical resolution through the upper troposphere and lower stratosphere (**Figure 2.1b**). These model grids differ from the isobaric levels on which many reanalysis products are provided. Vertical spacing associated with an example set of these isobaric levels (corresponding to ERA-40 and ERA-Interim) is included in **Figure 2.1** for context. See *Appendix A* for lists of model levels and further details of the vertical grid.

2.2.2 Major physical parameterizations

In this section we briefly describe some influential physical parameterizations used in the reanalysis forecast models, including those for longwave and shortwave radiation (**Table 2.4**), stratiform clouds (**Table 2.5**), moist convection (**Table 2.6**), gravity wave drag (**Table 2.7**), and horizontal and vertical diffusion (**Table 2.8**). Further details and additional references for many of these parameterizations are provided in the extended digital version of this chapter (*Chapter 2E*).

Table 2.4: *Radiative transfer schemes used in the forecast models of the reanalysis systems. A more complete discussion is provided in Chapter 2E.*

Reanalysis system	Radiative transfer scheme
ERA-40	Shortwave: *Fouquart and Bonnel* (1980) with four spectral intervals. Longwave: RRTM (*Mlawer et al.*, 1997). Radiation calculations are performed every 3 hours on a T63 horizontal grid.
ERA-Interim	Shortwave: Updated version of *Fouquart and Bonnel* (1980). Longwave: RRTM (*Mlawer et al.*, 1997). The scheme is a revised version of that used in ERA-40 with hourly radiation calculations on a T95 horizontal grid (*Dee et al.*, 2011).
ERA-20C	Shortwave: RRTM-G (*Mlawer et al.*, 1997; *Iacono et al.*, 2008). Longwave: RRTM-G (*Mlawer et al.*, 1997; *Iacono et al.*, 2008). Radiation calculations are performed every 3 hours on a T63 horizontal grid. A McICA approach with generalized overlap is used to represent the radiative effects of clouds (*Morcrette et al.*, 2008).
ERA5	Similar to ERA-20C, but with radiation calculations performed hourly on a T319 horizontal grid.
JRA-25 / JCDAS	Shortwave: *Briegleb* (1992) Longwave: line absorption based on the random band model of *Goody* (1952). Radiation calculations are performed on the full model grid, with calculations every hour for shortwave radiation and every three hours for longwave radiation.
JRA-55	Shortwave: *Briegleb* (1992), updated to use the formulation of *Freidenreich and Ramaswamy* (1999) for shortwave absorption by O_2, O_3, and CO_2. Longwave: *Murai et al.* (2005). Radiation calculations are performed on the full model grid, with calculations every hour for shortwave radiation and every three hours for longwave radiation.
MERRA	Shortwave: *Chou and Suarez* (1999). Longwave: *Chou et al.* (2001). Radiation calculations are performed hourly on the full model grid.
MERRA-2	Same as MERRA.
NCEP-NCAR R1	Shortwave: GFDL (*Lacis and Hansen*, 1974). Longwave: GFDL (*Schwarzkopf and Fels*, 1991; *Fels and Schwarzkopf*, 1975). Radiation calculations are performed every 3 hours on a 128×64 linear grid.
NCEP-DOE R2	Shortwave: *Chou and Lee* (1996). Longwave: GFDL (*Schwarzkopf and Fels*, 1991; *Fels and Schwarzkopf*, 1975; same as *R1*). Radiation calculations are performed hourly on the full model grid.
CFSR / CFSv2	Shortwave: Modified RRTM-G (*Clough et al.*, 2005). Longwave: Modified RRTM-G (*Clough et al.*, 2005). Radiation calculations are performed hourly on the full model grid. A McICA approach with maximum–random overlap is used for representing the radiative effects of clouds in CFSv2, but not in CFSR.
NOAA-CIRES 20CR v2	Shortwave: Modified RRTM-G (*Clough et al.*, 2005). Longwave: Modified RRTM-G (*Clough et al.*, 2005). Radiation calculations are performed hourly on the full model grid.

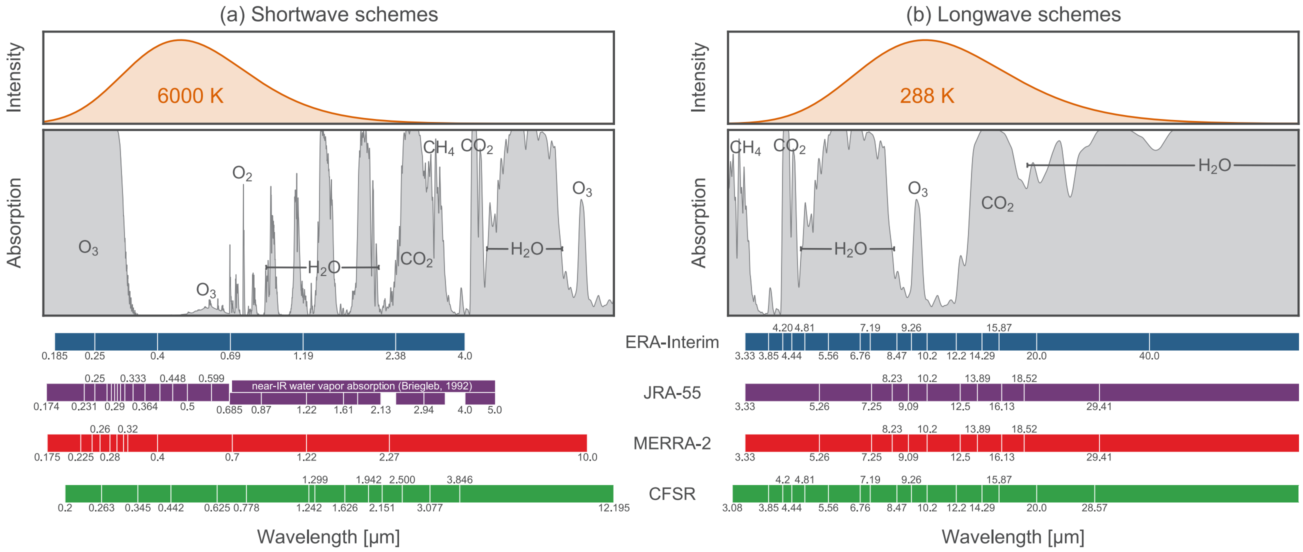


Figure 2.2: *Spectral bands in the radiation schemes used in four recent reanalyses.*

Other pertinent items include the treatment of incoming solar radiation, surface boundary conditions, and radiatively active gases and aerosols, which are summarized in *Section 2.2.3* (see also references in **Table 2.1**), as well as representations of land surface properties, which are described very briefly in **Section 2.2.4**.

The radiative parameterisations used in the forecast model components of reanalysis systems are broadband schemes, in which the radiative spectrum is discretized into a small set of spectral intervals or bands. The form of this discretization is dictated primarily by the presence of radiatively active constituents in the atmosphere and the wavelengths at which these constituents are active. Radiative fluxes and heating rates are computed by integrating across all spectral bands. Note that the radiative transfer schemes used in the atmospheric forecast models (**Table 2.4**) differ from the radiative transfer schemes used to process satellite radiances for data assimilation (**Table 2.19**).

Assumptions on cloud overlapping during radiation calculations are described in *Chapter 2E*.

Parameterizations of stratiform or “large-scale” clouds

Table 2.5: *Non-convective (stratiform) cloud parameterizations used in the forecast models of the reanalysis systems. A more complete discussion is provided in Chapter 2E.*

Reanalysis System	Cloud Parameterization
ERA-40	A prognostic cloud scheme (*Tiedtke*, 1993), in which cloud fraction and cloud water content both evolve according to physical sources and sinks.
ERA-Interim	Similar to ERA-40, but updated to include a treatment for ice supersaturation at temperatures less than 250 K (*Tompkins et al.*, 2007).
ERA-20C	Similar to ERA-Interim, but updated to permit separate estimates of liquid and ice water in non-convective clouds.
ERA5	Same as ERA-20C.
JRA-25 / JCDAS	A modified version of the parameterization proposed by *Smith* (1990), but with stratocumulus cloud fractions following *Kawai and Inoue* (2006).
JRA-55	Same as JRA-25.
MERRA	A prognostic scheme developed by *Bacmeister et al.*, (2006). Convectively-detrained “anvil” condensate is tracked separately from condensate formed in situ, with the former converted to the latter over a specified e-folding timescale.
MERRA-2	As in MERRA, but with new constraints on distributions of total water following *Molod* (2012) and a modified function governing the partitioning of cloud water into liquid and ice during cloud formation.
NCEP-NCAR R1	Diagnosed as a function of grid-scale relative humidity; known to produce discontinuities around 0°E and 180°E longitude (*Kanamitsu et al.*, 2002).
NCEP-DOE R2	Diagnosed as a function of grid-scale RH; modified from that used by R1 to eliminate the discontinuities around 0°E and 180°E.
CFSR / CFSv2	A simple cloud physics parameterization with prognostic cloud condensate (*Zhao and Carr*, 1997). Cloud fraction is diagnosed as a function of cloud water content and relative humidity (*Xu and Randall*, 1996).
NOAA-CIRES 20CR v2	Same as CFSR.

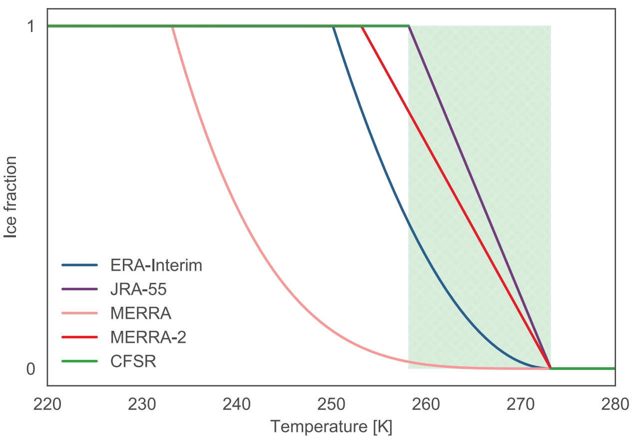


Figure 2.3: *Partitioning of prognostic cloud condensate between the ice and liquid phases as a function of temperature in five recent reanalysis systems. See Chapter 2E for details.*

in the reanalysis systems (listed in **Table 2.5**) influence surface fluxes and the atmospheric state via couplings with radiative transfer, precipitation, and convection. The simplest parameterisations diagnose stratiform cloud cover at each time step as a function of the difference between the grid-scale relative humidity and a critical relative humidity. The existence of clouds in the model atmosphere thus depends on the relative humidity exceeding this critical threshold. NCEP-NCAR R1 and NCEP-DOE R2 use this type of "diagnostic" parameterization. Although computationally inexpensive, diagnostic cloud parameterizations have a number of intrinsic flaws (see, *e.g.*, *Xu and Krueger*, 1991), and have been replaced in more recent reanalyses by variations on the "prognostic" approach pioneered by *Sundqvist* (1978). Prognostic parameterizations simulate the evolution of key cloud variables, such as cloud fraction, cloud water content, and precipitation, and allow for the persistence and advection of convectively-detrained anvil clouds across multiple time steps, as well as the inclusion of more sophisticated approaches to simulating the autoconversion of cloud condensate to rain and snow. The prognostic cloud parameterizations used in reanalyses consider two primary sources of stratiform clouds. The first of these, detrainment of cloud condensate from moist convection, depends on the formulation of the convection schemes documented in **Table 2.6**. The second source, in situ condensation resulting from large-scale cooling, may be represented either via empirically-based PDFs (*e.g.*, *Molod*, 2012; *Smith*, 1990) or by prognostic equations that track the physical sources and sinks of stratiform cloud (*e.g.*, *Tiedtke*, 1993).

Another potentially influential difference among the prognostic cloud schemes used in reanalysis systems is the approach to partitioning cloud condensate into ice and liquid phases (**Figure 2.3**), which affects both the optical properties (and hence radiative transfer) and microphysical properties (and hence autoconversion and precipitation) of the

Table 2.6: *Convective parameterizations used in the forecast models of the reanalysis systems. A more complete discussion is provided in Chapter 2E.*

Reanalysis System	Convective Parameterization
ERA-40	Deep, shallow, and mid-level cumulus convection are parameterized using a bulk mass flux scheme based on that proposed by *Tiedtke* (1989). Each simulated convective cloud consists of a single pair of entraining/detraining plumes that represent updraught and downdraught processes.
ERA-Interim	Similar to ERA-40, but modified in several respects to improve the diurnal cycle of convection, increase convective precipitation efficiency, and make more explicit distinctions among shallow, mid-level and deep convective clouds (*Dee et al.*, 2011).
ERA-20C	Similar to ERA-Interim but with modified representations of entrainment and detrainment rates and a revised convective adjustment time scale.
ERA5	Similar to ERA-20C but with a new closure that better accounts for coupling between the boundary layer and free troposphere, improving the diurnal cycle of convection (*Bechtold et al.*, 2014).
JRA-25 / JCDAS	An 'economical prognostic' mass-flux type Arakawa–Schubert cumulus scheme (*JMA*, 2007; *Arakawa and Schubert*, 1974).
JRA-55	Similar to JRA-25 but with a new triggering mechanism (*Xie and Zhang*, 2000).
MERRA	A version of the relaxed Arakawa–Schubert cumulus scheme (*Moorthi and Suarez*, 1992).
MERRA-2	Same as MERRA, but with a new stochastic Tokioka-type entrainment condition that limits the occurrence of plumes with very small entrainment rates (*Molod et al.*, 2015).
NCEP-NCAR R1	Deep convective clouds are simulated using a simplified Arakawa–Schubert convection scheme (*Pan and Wu*, 1995; *Arakawa and Schubert*, 1974); shallow convective clouds are simulated using a Tiedtke-type scheme (*Tiedtke*, 1989).
NCEP-DOE R2	Similar to NCEP-NCAR R1, but with minor tuning applied.
CFSR / CFSv2	Same underlying schemes as *R1* and *R2*, but with substantial updates as described by *Moorthi et al.* (2001, 2010) and *Saha et al.* (2010).
NOAA-CIRES 20CR v2	Same as CFSR.

simulated clouds. As with the cloud schemes themselves, this partitioning may be either diagnostic or prognostic. See *Chapter 2E* for further details.

Moist convection is another critical subgrid-scale process that must be parameterized in atmospheric models (*Arakawa*, 2004). All of the reanalyses described in this chapter represent moist convection using versions of bulk mass-flux parameterizations (*Tiedtke*, 1989; *Arakawa and Schubert*, 1974), which have as their conceptual basis the "hot tower" hypothesis of *Riehl and Malkus* (1958). These parameterizations represent the statistical effects of convection in a given grid cell via one or more updraft and downdraft plumes, which are in turn coupled to the background environment via entrainment and detrainment, diabatic heating, and the vertical transport of tracers and momentum. Key differences in the convective parameterizations used by the reanalysis systems include the trigger function, the principal closure, whether and to what extent momentum and tracer transport are included, restrictions on the properties of the individual plumes (*e.g.*, entrainment, detrainment, cloud base, and cloud top), and assumptions governing the production and partitioning of rainfall and cloud condensate. We summarize the convection schemes used in each reanlysis in **Table 2.6**. In *Chapter 2E*, we briefly describe the two aspects, trigger functions and closure assumptions.

Gravity wave drag (GWD) parameterisations are used in reanalysis forecast models to represent the systematic effects of momentum deposition on the resolved flow by small-scale (*i.e.*, unresolved) gravity waves. As a relative fraction of the momentum budget the importance of GWD forcing generally increases with altitude, becoming a dominant contribution in the mesosphere (*Polavarapu et al.*, 2005), but effects can also be significant at lower altitudes, such as on the upper poleward flank of the tropospheric subtropical jet (*McFarlane*, 1987; *Palmer et al.*, 1986). GWD parameterisations are typically implemented in atmospheric models via separate schemes for orographic and non-orographic gravity waves. All reanalysis systems considered here include orographic GWD parameterisation, but only ERA-20C, ERA5, MERRA, MERRA-2, and CFSv2 include non-orographic GWD parameterizations (**Table 2.7**). *Chapter 2E* has some further discussions on orographic and non-orographic gravity waves.

Table 2.7: *Gravity wave drag parameterizations used in the forecast models of the reanalysis.*

Reanalysis System	Gravity Wave Drag Parameterization
ERA-40	Subgrid-scale orographic drag is parameterized using the scheme developed by *Lott and Miller* (1997). The representation of the orographic gravity wave source follows *Miller* (1989) and *Baines and Palmer* (1990), and accounts for three-dimensional variability in the amplitude and orientation of wave stress. Non-orographic gravity wave drag is represented as Rayleigh friction above the stratopause.
ERA-Interim	Same as ERA-40.
ERA-20C	Subgrid-scale orographic drag is parameterized similarly to ERA-40 and ERA-Interim, but with slight modifications that increase gravity wave activity. Non-orographic gravity wave drag is included using the parameterization proposed by *Scinocca* (2003); see also *Orr et al.* (2010).
ERA5	Same as ERA-20C, except with a latitudinal dependence of non-orographic launch flux.
JRA-25 / JCDAS	The orographic gravity wave drag parameterization consists of a long wave (wavelengths over 100 km) component and a short wave (wavelengths of ~10 km) component (*Iwasaki et al.*, 1989a, 1989b). Long waves are assumed to propagate upward and break mainly in the stratosphere, where they exert drag (*Palmer et al.*, 1986). Short waves are regarded as trapped and dissipating within the troposphere. Non-orographic gravity wave drag is not included.
JRA-55	Same as JRA-25.
MERRA	MERRA includes parameterizations that compute drag due to the breaking of orographic (McFarlane, 1987) and non-orographic (after *Garcia and Boville*, 1994) gravity waves.
MERRA-2	Similar to MERRA, but with an increased latitudinal profile of the gravity wave drag background source at tropical latitudes and increased intermittency (*Molod et al.*, 2015).
NCEP-NCAR R1	An orographic gravity wave drag scheme based on *Palmer et al.* (1986), *Pierrehumbert* (1987), and *Helfand et al.* (1987) is included in the forecast model. Non-orographic gravity wave drag is not included.
NCEP-DOE R2	Same as NCEP-NCAR R1.
CFSR / CFSv2	The orographic gravity wave drag parameterization is based on the scheme proposed by *Kim and Arakawa* (1995). Sub-grid scale mountain blocking is represented using the scheme developed by *Lott and Miller* (1997). Although non-orographic gravity wave drag is not considered in CFSR, a simple representation of non-orographic gravity wave drag is included in CFSv2 via the parameterization proposed by *Chun and Baik* (1998).
NOAA-CIRES 20CR v2	The orographic gravity wave drag parameterization is the same as in CFSR. Non-orographic gravity wave drag is not considered.

Table 2.8 briefly describes the implementations of horizontal and vertical diffusion in the atmospheric forecast models used by the reanalysis systems. All of the systems that use spectral dynamical cores on Gaussian or reduced Gaussian grids (see above) use implicit linear diffusion in spectral space, although the implementations vary from second-order (NCEP-NCAR R1, NCEP-DOE R2, and 20CR) to eighth-order (CFSR). MERRA and MERRA-2, which are built on finite volume dynamical cores, use slightly different implementations of explicit second-order diffusion. Representations of vertical diffusion in the free troposphere and above are based on first order *K*-type closures. One of the most notable differences among these parameterizations as implemented in the reanalysis systems is the presence or absence of a critical Richardson number, above which turbulent mixing no longer occurs (*Flannaghan and Fueglistaler*, 2014). See the extended *Chapter 2E* (and Figure 2.4) for additional information. Consideration of turbulence in the surface layer and ABL introduces a wider array of parameterizations for turbulent mixing, which are listed in **Table 2.8** but not introduced in detail. Differences in these parameterisations may influence surface exchanges of enthalpy and

Table 2.8: *Representations of vertical and horizontal diffusion in the forecast models used by reanalysis systems.*

Reanalysis System	Representations of Vertical and Horizontal Diffusion
ERA-40	Horizontal diffusion: Implicit linear fourth-order diffusion in spectral space. Vertical diffusion: Vertical diffusion in the free atmosphere and in the ABL under stable conditions is based on the revised Louis scheme (*Beljaars*, 1995; *Louis*, 1979;) for positive Richardson numbers and on Monin–Obukhov similarity for negative Richardson numbers. Vertical diffusion in the ABL under unstable conditions is based on the non-local scheme proposed by *Troen and Mahrt* (1986). Turbulent fluxes in the surface layer are calculated using bulk formulae based on Monin–Obukhov similarity.
ERA-Interim	Horizontal diffusion: Same as ERA-40. Vertical diffusion: Vertical diffusion in the free atmosphere and in the ABL under stable conditions is as in ERA-40. Vertical diffusion in the ABL under unstable conditions is based on an eddy-diffusivity mass-flux (EDMF) scheme (*Köhler et al.*, 2011). Turbulent fluxes in the surface layer are calculated using bulk formulae based on Monin–Obukhov similarity.
ERA-20C	Horizontal diffusion: Same as ERA-40. Vertical diffusion: Similar to ERA-Interim, but with vertical diffusion above the lower troposphere based on Monin–Obukhov similarity under all conditions (rather than the revised Louis scheme) and the inclusion of a simple empirical parameterization to represent unresolved vertical wind shear.
ERA5	Similar to ERA-20C, but with the empirical parameterization of unresolved vertical wind shear removed.
JRA-25 / JCDAS	Horizontal diffusion: Implicit linear fourth-order diffusion in spectral space. Vertical diffusion: Vertical diffusion of momentum, heat, and moisture are represented using the level 2 turbulence closure scheme developed by *Mellor and Yamada* (1974). Surface turbulent fluxes are calculated using bulk formulae based on Monin–Obukhov similarity.
JRA-55	Same as JRA-25.
MERRA	Horizontal diffusion: Explicit second-order horizontal divergence damping is included in the dynamical core. Vertical diffusion: Vertical diffusion in the free atmosphere and in the boundary layer under stable conditions is based on a local gradient Richardson number closure (*Louis et al.,* 1982), but a tuning parameter severely suppresses turbulent mixing at pressures less than ~900 hPa. Vertical diffusion in the boundary layer under unstable conditions is based on the non-local scheme proposed by *Lock et al.* (2000).
MERRA-2	Horizontal diffusion: Similar to MERRA, but with an additional second-order Smagorinsky divergence damping. Vertical diffusion: Similar to MERRA in most respects, with the addition of a Monin–Obukhov-type parameterization to represent turbulent fluxes across the surface layer (*Helfand and Schubert*, 1995). The tuning parameter that suppressed turbulent mixing at pressures less than ~900 hPa in MERRA has been removed, but diffusion coefficients are still usually very small in the free atmosphere.
NCEP-NCAR R1	Horizontal diffusion: Implicit linear second-order diffusion in spectral space. Horizontal diffusion along model σ layers led to the occurrence of spurious "spectral precipitation", particularly in mountainous areas at high latitudes. A special precipitation product was produced to correct this issue. Vertical diffusion: Local K diffusion (*Louis et al.*, 1982) is applied in both the ABL and the free atmosphere with a uniform background diffusion coefficient.
NCEP-DOE R2	Horizontal diffusion: Implicit linear second-order diffusion in spectral space. Issues with spectral precipitation caused by horizontal diffusion are greatly reduced relative to *R1*. Vertical diffusion: Local K diffusion (*Louis et al.*, 1982) is applied in the free atmosphere with a uniform background diffusion coefficient. Non-local diffusion is applied in the ABL (*Hong and Pan*, 1996).
CFSR / CFSv2	Horizontal diffusion: Implicit linear eighth-order diffusion in spectral space. Vertical diffusion: Local K diffusion (*Louis et al.*, 1982) is applied in the free atmosphere with a background diffusion coefficient that decreases exponentially with pressure. Non-local vertical diffusion is applied in the ABL (*Hong and Pan*, 1996).
NOAA-CIRES 20CR v2	Horizontal diffusion: Implicit linear second-order diffusion in spectral space. Vertical diffusion: Same as CFSR.

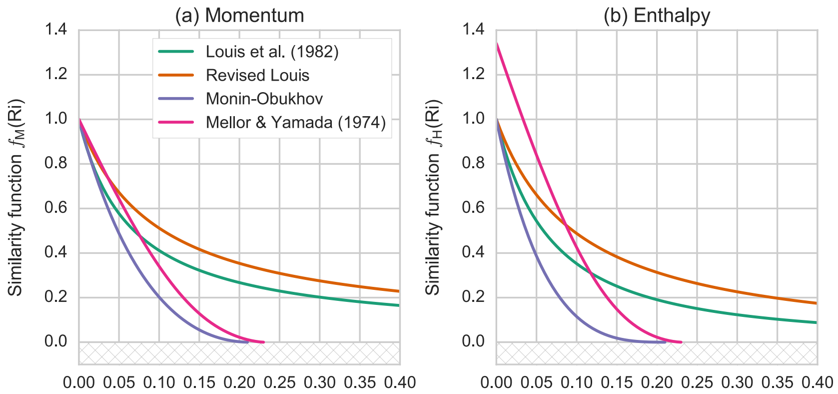


***Figure 2.4:** Similarity functions for parameterized turbulent transfer of (a) momentum and (b) enthalpy (heat and moisture) as a function of the gradient Richardson number (Ri) based on four turbulence schemes used in the free troposphere by reanalysis systems. See Chapter 2E for details.*

momentum. Different treatments of surface roughness lengths over land and ocean can also influence energy and momentum fluxes into the atmosphere; these aspects are documented in Table 2.9 of *Chapter 2E* but are omitted here.

2.2.3 Boundary and other specified conditions

This section describes the boundary and other specified conditions that can be regarded as "externally supplied forcings" for each reanalysis system. These conditions comprise the elements of the reanalysis that are not taken from the forecast model or data assimilation but are used to produce the outputs. **Figure 2.5** shows three examples of how externally-specified boundary conditions may vary amongst reanalysis systems.

The factors that may be considered "external" vary somewhat among reanalyses because the forecast and assimilation components have provided a progressively larger fraction of the inputs (initial conditions) for the forecast model as reanalysis systems have developed. Ozone is a prime example. As discussed below, all of the

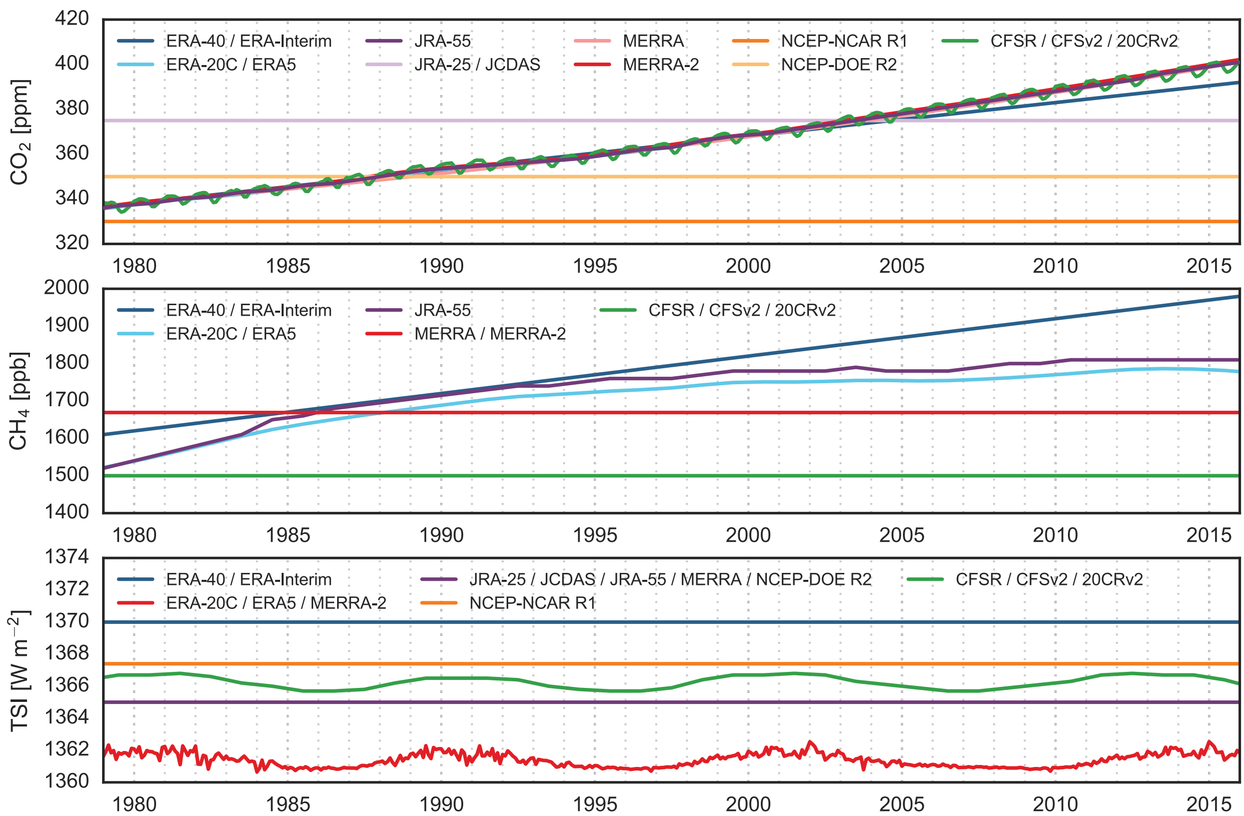


***Figure 2.5:** Time series of boundary and specified conditions for CO_2 (top), CH_4 (center), and TSI (bottom) used by the reanalysis systems. The CH_4 climatology used in MERRA and MERRA-2 varies in both latitude and height; here a "tropospheric mean" value is calculated as a mass- and area-weighted integral between 1000 hPa and 288 hPa to facilitate comparison with the "well-mixed" values used by most other systems. ERA-20C and ERA5 also apply rescalings of annual mean values of both CO_2 and CH_4 that vary in latitude and height; here the base values are shown (note that the ERA-20C/ERA5 time series in panel a is obscured by those for JRA-55 and MERRA-2). Time series of TSI neglect seasonal variations due to the ellipiticity of the Earth's orbit, as these variations are applied similarly (but not identically) across reanalysis systems. Additional information on CO_2 and CH_4 is provided in **Table 2.13**, and additional information on TSI is provided in **Table 2.14**. Reproduced from Fujiwara et al. (2017).*

[1] **Table 2.9** (titled as "Sources and representations of surface roughness in the reanalysis systems") is only shown in *Chapter 2E.*

reanalysis systems except for NCEP-NCAR R1, NCEP-DOE R2, and NOAA-CIRES 20CR; JRA-55 and ERA-40 prior to 1978; and ERA5 prior to April 1970) assimilate satellite ozone measurements. Some of these reanalysis systems (notably ERA-40, ERA-Interim, ERA-20C, and ERA5) use ozone climatologies rather than internally generated ozone fields for radiation calculations in the forecast model. MERRA-2 assimilates aerosol optical depths and uses internally generated aerosol fields for the radiation calculations, while other systems use

Table 2.10: *Treatment of sea surface temperature and sea ice.* [1]

Reanalysis System	Sea Surface Temperature and Sea Ice
ERA-40	Monthly data from the Met Office HadISST1 product was used before November 1981, replaced by weekly data from the NOAA–NCEP 2D-Var product from December 1981 through June 2001 and NOAA OISSTv2 from July 2001 through August 2002 (*Reynolds et al.*, 2002). A special sea ice analysis and a method of specifying SST in grid boxes with partial ice-cover were used. Interpolation was used to produce daily values.
ERA-Interim	Similar to ERA-40 but NCEP RTG sea surface temperatures were used from January 2002 through January 2009 and OSTIA (*Donlon et al.*, 2012) was used from February 2009 through August 2019.
ERA-20C	Daily gridded SST and sea ice are from HadISST version 2.1.0.0 (*Titchner and Rayner*, 2014) at 0.25° horizontal resolution. Daily fields are obtained via cubic interpolation from monthly analyses, with the temporal average of daily fields constrained to match the analysed monthly mean.
ERA5	Daily gridded SSTs are from HadISST version 2.1.0.0 between January 1949 and August 2007, and from OSTIA for September 2007 onwards. Sea ice cover is from HadISST version 2.0.0.0 from January 1950 through December 1978, from reprocessed OSI SAF fields between January 1979 and August 2007, and from operational OSI SAF estimates for September 2007 onwards. Data through August 2007 are at 0.25° horizontal resolution, while data from September 2007 to present are at 0.05° horizontal resolution. When necessary, daily fields are obtained from monthly analyses using the same procedure as ERA-20C.
JRA-25 / JCDAS	Daily COBE SSTs (*Ishii et al.*, 2005) were used. COBE SSTs are based on the ICOADS and Kobe data collections, and do not include satellite data. Daily sea ice distributions prepared for COBE are based on reports by *Walsh and Chapman* (2001) for the Northern Hemisphere and *Matsumoto et al.* (2006) for the Southern Hemisphere.
JRA-55	Daily COBE SSTs and sea ice distributions are used, with minor updates from those used for JRA-25/JCDAS. Southern Hemisphere sea ice coverage is based on a climatology before October 1978, and based on *Matsumoto et al.* (2006) after October 1978.
MERRA	Weekly NOAA OISST data at 1° resolution (*Reynolds and Smith*, 1994) are linearly interpolated in time to the model time steps.
MERRA-2	Monthly 1° gridded data (*Taylor et al.*, 2000) are used prior to 1982, daily 0.25° gridded data (*Reynolds et al.*, 2007) through March 2006, and daily 0.05° gridded data from OSTIA (*Donlon et al.*, 2012) from April 2006.
NCEP-NCAR R1	SSTs are taken from the Met Office Global Ice and Sea Surface Temperature (GISST) data set for 1981 and earlier, and from the NOAA OISST data set from 1982 to the present. Sea ice cover is from Navy/NOAA Joint Ice Center analyses before 1978, from SMMR observations for 1978 through 1987, and from SSM/I observations for 1988 through the present. Snow cover is from the NESDIS weekly snow cover analysis (Northern Hemisphere only) for September 1998 and earlier, and from the US Air Force global snow cover analysis from October 1998 through the present.
NCEP-DOE R2	SSTs and sea ice cover for January 1979 through 15 August 1999 are taken from data prepared for AMIP-II and provided by the PCMDI at Lawrence Livermore National Laboratory. SSTs and sea ice cover for 16 August 1999 through December 1999 are from monthly NOAA OISST and monthly NCEP operational sea ice analyses, interpolated to daily resolution. SSTs and sea ice cover for January 2000 to present are from daily NOAA OISST and NCEP operational sea ice analyses.
CFSR / CFSv2	The atmospheric model is coupled to the GFDL MOM version 4 ocean model and a two-layer sea ice model. AVHRR and AMSR satellite infrared observations of SST are assimilated in the SST analysis, along with in situ data from ships and buoys. The sea (and lake) ice concentration analysis products assimilate different observational data depending on the period, including microwave satellite observations when available. Temperatures at the atmosphere–ocean boundary are relaxed every six hours to separate SST analyses, including the 1° gridded HadISST1.1 from January 1979 through October 1981 and versions 1 and 2 of the 0.25° gridded OI analyses described by *Reynolds et al.* (2007) from November 1981. Further details of the coupling procedure and SST/sea ice analysis have been provided by *Saha et al.* (2010).
NOAA-CIRES 20CR v2	HadISST1.1 monthly mean SST and sea ice data are interpolated to daily resolution. Sea ice concentrations were accidentally mis-specified in coastal regions. This error results in warmer lower tropospheric temperatures in polar regions relative to ERA-40 and NCEP-NCAR R1 (*Compo et al.*, 2011). The error has been corrected in Version 2c of the reanalysis.

[1] Table 2.9: Sources and representations of surface roughness in the reanalysis systems is provided in *Chapter 2E*.

climatologies or neglect the role of aerosols altogether. CFSR is a coupled atmosphere–ocean–sea ice system, in which the SST and sea ice lower boundary conditions for the atmospheric model are generated by an ocean model (although temperatures at the boundary are relaxed every six hours to SST analyses similar to those used by other reanalysis systems). This section summarizes the treatment of SST, sea ice, ozone, aerosols, trace greenhouse gases (other than water vapour), and the solar cycle, with special notes where necessary. Dynamical variables, water vapour, and internally generated ozone (*i.e.*, variables that are often directly constrained by the set of assimilated observations) are discussed and evaluated in *Chapters 3 and 4* of this report.

Table 2.11: *Treatment of ozone. See also Chapter 4 of this report.*

Reanalysis System	Treatment of Ozone
ERA-40	TOMS and SBUV ozone retrievals were assimilated from 1978 onward. Ozone in the model is described using a linearization of the ozone continuity equation, including photochemical sources and sinks (*Cariolle and Déqué*, 1986; *Dethof and Hólm*, 2004). The model does not account for heterogeneous chemistry, but does include an empirical ozone destruction term to account for chemical loss in polar stratospheric clouds. Model-generated ozone is not used in the radiation calculations, which instead assume the climatological ozone distribution reported by *Fortuin and Langematz* (1995).
ERA-Interim	Ozone retrievals are assimilated from TOMS (1979 - present), SBUV (1979 - present), GOME (1996 - 2002), MIPAS (2003 - 2004), SCIAMACHY (2003 - 2008), MLS (2008 - present), and OMI (2008 - present). The ozone scheme is an updated version of that used in ERA-40 (*Dragani*, 2011; *Cariolle and Teyssèdre*, 2007). As in ERA-40, climatological ozone distributions from *Fortuin and Langematz* (1995) are used for radiation calculations.
ERA-20C	No ozone data are assimilated. The forecast model ozone parameterization is identical to that used in ERA-Interim. Model-generated ozone is not used in the radiation calculations, which instead use monthly three-dimensional ozone fields that evolve in time (*Cionni et al.*, 2011).
ERA5	The ozone scheme is the same as that used in ERA-Interim, but with substantial updates to the assimilated data. Reprocessed retrievals are assimilated from BUV (1970 - 1977), TOMS (1979 - 2003), SBUV v8.6 (1979 - present), CCI MIPAS (2005 - 2012) and SCIAMACHY (2003 - 2012), Aura MLS v4.2 (2004 - present) and OMI-DOAS (2004 - present). ERA5 also assimilates IR ozone-sensitive radiance not used in ERA-Interim, and uses variational bias correction (see *Section 2.4.3.2*) during the ozone analysis. Analyzed ozone is not used in the radiation calculations, which instead use an in-house ozone climatology from CAMSiRA (*Flemming et al.*, 2017).
JRA-25 / JCDAS	Daily ozone distributions were prepared in advance using the MRI-CCM1 offline chemical transport model with output “nudged” to satellite retrievals of total ozone. These distributions were provided to the forecast model for use in radiation calculations.
JRA-55	For 1979 and later, the approach is similar to that used by JRA-25/JCDAS, but uses an updated chemical transfer model with 68 vertical levels rather than 45. For 1958 - 1978, a monthly mean climatology generated from the 1980 - 1984 ozone analyses was used. These distributions were provided to the forecast model for use in radiation calculations.
MERRA	Version 8 SBUV ozone retrievals have been assimilated from October 1978 onward. The ozone parameterization is based on an empirical relationship between ozone and prognostic odd-oxygen that varies with height and the diurnal cycle (*Rienecker et al.*, 2008). The parameterization uses zonally-symmetric monthly production and loss rates derived from a 2-dimensional model as described by *Stajner et al.* (2008), but without representation of heterogeneous chemistry in polar regions. The forecast model uses analyzed ozone data in radiation calculations.
MERRA-2	Version 8.6 SBUV retrievals have been assimilated in reanalyses between 1980 and 2004. Starting from October 2004, these data have been replaced by retrieved MLS profiles (version 2.2 through 31 May 2015; version 4.2 from 1 June 2015) and OMI observations of total ozone (*McCarty et al.*, 2016). Assimilation of MLS retrievals at 261 hPa was discontinued starting on 1 May 2016 (*Wargan et al.*, 2017). The ozone parameterization is the same as that used in MERRA. The forecast model uses analyzed ozone data in radiation calculations.
NCEP-NCAR R1	Seasonal ozone climatologies reported by *London* (1962) and *Hering et al.* (1965) are used in radiation calculations. No ozone analysis is produced.
NCEP-DOE R2	The zonal mean ozone climatology published by *Rosenfield et al.* (1987) is used in radiation calculations, but the latitudinal orientation was reversed north-to-south. Although this error may cause some problems in the stratosphere, *Kanamitsu et al.* (2002) report that the overall impact is minor. No ozone analysis is produced.
CFSR / CFSv2	Version 8 SBUV profiles and total ozone retrievals were assimilated without bias adjustment. Prognostic ozone is parameterized using concentration-dependent climatological production and destruction terms generated by a 2-dimensional chemistry model (*McCormack et al.*, 2006). The forecast model uses analyzed ozone data for radiation calculations. Late 20th century levels of CFCs are included implicitly in the gas phase chemistry and ozone climatology used in the prognostic ozone parameterization.
NOAA-CIRES 20CR v2	No ozone data are assimilated. The ozone model is the same as that used in CFSR.

Table 2.12: *Treatment of aerosols.*

Reanalysis System	Treatment of Aerosols
ERA-40	Aerosols have been included in the radiation calculations using prescribed climatological aerosol distributions (*Tanré et al.*, 1984). These distributions include annual mean geographical distributions for maritime, continental, urban and desert aerosol types, in addition to uniformly distributed tropospheric and stratospheric 'background' aerosol loading. No trends or temporal variations (such as variations due to volcanic eruptions) were included.
ERA-Interim	Aerosols are included in the radiation calculations using updated climatological distributions (*Tegen et al.*, 1997). The climatological annual cycles of tropospheric aerosols have been revised relative to those used by ERA-40, as have the optical thickness values for tropospheric and stratospheric background aerosols. There is no evolution of volcanic aerosols.
ERA-20C	The evolution of tropospheric aerosols is based on data prepared for CMIP5 (*an Vuuren et al.*, 2011; *Lamarque et al.*, 2010). Volcanic sulphates (*Sato et al.*, 1993) and ash (*Tanré et al.*, 1984) are also included in the stratosphere. A more detailed description of the aerosol fields used in ERA-20C and ERA-20CM has been provided by *Hersbach et al.* (2015).
ERA5	Same as ERA-20C.
JRA-25 / JCDAS	Aerosols are represented using two aerosol profiles, one over land and one over sea (*WMO*, 1986). Neither interannual nor seasonal variations are considered.
JRA-55	Similar to JRA-25, but with optical depths adjusted to a 2-dimensional monthly climatology (*JMA*, 2013). Interannual variations, such as those due to volcanic eruptions, are not considered.
MERRA	Aerosols are represented using a climatological aerosol distribution generated using the Goddard Chemistry, Aerosol, Radiation, and Transport (GOCART) model (*Colarco et al.*, 2010).
MERRA-2	Aerosol optical depths from AVHRR, MODIS, MISR and AERONET are assimilated into the GEOS-5 GAAS (*Buchard et al.*, 2015, 2017; *Randles et al.*, 2017). Volcanic aerosols are included. The forecast model uses analyzed aerosols in radiation calculations for the entire production period. Additional details have been provided by *Randles et al.* (2017).
NCEP-NCAR R1	No aerosols.
NCEP-DOE R2	No aerosols.
CFSR / CFSv2	Aerosols are represented using a seasonally varying climatological global distribution of aerosol vertical profiles on a 5° grid (*Koepke et al.*, 1997). Monthly zonal mean volcanic aerosols in four latitude bands (90 - 45°S, 45°S - equator, equator - 45°N, 45 - 90°N) are specified based on data reported by *Sato et al.* (1993).
NOAA-CIRES 20CR v2	Same as CFSR.

2.2.3.1 Sea surface temperature and sea ice

Table 2.10 summarizes the treatment of SST and sea ice distributions in the reanalysis systems, including the names of SST and sea-ice datasets, special calibration or preprocessing details (*e.g.*, bias corrections, interpolations), and details of how the datasets were produced.

2.2.3.2 Ozone

Table 2.11 briefly summarizes the treatment of ozone in the reanalysis systems (detailed intercomparisons are deferred to *Chapter 4*). Some reanalysis systems assimilate satellite ozone measurements (from 1978/1979, and in one case 1970, onward) to produce an ozone analysis product, while some systems do not. Moreover, some systems that produce an ozone analysis use a climatological ozone distribution (rather than the ozone analysis) for radiation calculations in the forecast model. These distinctions are made explicit in **Table 2.11**. None of the reanalysis systems considered here assimilate data from ozonesondes.

2.2.3.3 Aerosols

Table 2.12 summarizes the treatment of stratospheric and tropospheric aerosols in the reanalysis systems. Some reanalysis systems consider tropospheric aerosols over continents and over oceans separately in the radiation scheme. Some reanalysis systems (but not all) account for changes in stratospheric aerosols due to volcanic eruptions. One reanalysis (MERRA-2) assimilates aerosol optical depths and uses analyzed aerosols in radiation calculations.

2.2.3.4 Carbon dioxide and other radiatively active gases

Table 2.13 summarizes the treatment of carbon dioxide and other radiatively active gases (except for water

Table 2.13: *Treatment of carbon dioxide and other radiatively active gases.*

Reanalysis System	CO_2 and Reactive Trace Gases
ERA-40	CO_2, CH_4, N_2O, CFC-11, and CFC-12 are assumed to have globally uniform concentrations throughout the atmosphere. The concentrations of these gases were set to the observed 1990 values plus a linear trend as specified by *IPCC* (1996).
ERA-Interim	Same as ERA-40.
ERA-20C	CO_2, CH_4, N_2O, CFC-11, and CFC-12 are specified according to CMIP5-recommended values (*Meinshausen et al.*, 2011). The IPCC RCP3PD scenario is followed for 2006 - 2010. Greenhouse gases are not assumed to be globally uniform; rather, they are rescaled to match specified seasonal cycles and zonal mean vertical distributions (*Hersbach et al.*, 2015).
ERA5	Same as ERA-20C, with extension of RCP3PD after 2010.
JRA-25 / JCDAS	A constant, globally uniform CO_2 concentration of 375 ppmv was assumed. CH_4, N_2O, CFCs, and HCFCs were not considered.
JRA-55	Daily values of CO_2, CH_4, N_2O, CFC-11, CFC-12, and HCFC-22 are specified by interpolating from annual mean values. For CO_2, CH_4, and N_2O these annual mean values are valid on 1 July; for CFC-11, CFC-12, and HCFC-22 they are valid on 31 December. All species are assumed to be globally uniform, with sources that vary in time (*Kobayashi et al.*, 2015; their Table 7).
MERRA	CO_2 concentrations are assumed to be globally uniform and are specified according to historical observed values. CH_4, N_2O, CFCs, and HCFCs are specified according to steady state monthly climatologies from the Goddard two-dimensional chemistry transport model (*Rienecker et al.*, 2008). These monthly climatologies vary in both latitude and pressure, but do not contain interannual variability.
MERRA-2	Annual global mean CO_2 concentrations follow the IPCC RCP4.5 scenario and are assumed to be uniform throughout the atmosphere. CH_4, N_2O, CFCs, and HCFCs are specified as in MERRA.
NCEP-NCAR R1	A constant, globally uniform CO_2 concentration of 330 ppmv is assumed. CH_4, N_2O, CFCs, and HCFCs are not considered.
NCEP-DOE R2	Similar to *R1*, but with a constant, globally uniform CO_2 concentration of 350 ppmv.
CFSR / CFSv2	Monthly mean 15°×15° distributions of CO_2 concentrations derived from historical WMO Global Atmosphere Watch observations are used. Constant values of CH_4, N_2O, O_2, and four types of halocarbons are also included in the radiation calculations.
NOAA-CIRES 20CR v2	Similar to CFSR for 1956 and later. Estimates of semi-annual average global mean CO_2 concentrations based on ice core data are specified for the period before 1956. Values of CH_4, N_2O, O_2, and four types of halocarbons are constant throughout.

vapour) in the reanalysis systems (see also **Figure 2.5**). Notes on the treatment of water vapour are provided in *Section 2.4.4*.

2.2.3.5 Solar cycle

The solar cycle (*i.e.*, changes in TSI with a period of ~11 years) is an important driver of atmospheric variability. This variability is incorporated in reanalysis systems in a variety of ways, including specified solar radiation at the TOA (boundary condition) and/or observations of temperature or ozone (data assimilation). **Table 2.14** briefly briefly summarizes the extent to which interannual variations in TSI are represented in each reanalysis system (see also **Figure 2.5**).

2.2.4 Surface air and land surface treatments

Treatments of surface air and land surface properties present a number of challenges for reanalyses. For example, sharp gradients and other types of spatial heterogeneity in land cover are difficult to represent in global models, but have important influences on the magnitudes and variability of water and energy fluxes between the land surface and the atmosphere. More specific to reanalyses, the spatial region for which near-surface observations may be considered representative is reduced in coastal regions and regions of complex topography. Land surface properties, such as soil moisture and soil temperature, also evolve relatively slowly, especially at deeper layers. As a result, these variables are among the main targets of model spin-up. Discontinuities in the land surface state at stream transitions (*Section 2.5*) can propagate into the atmosphere.

Reanalyses use two main approaches for producing surface air analysis variables over land (**Table 2.15**). The first approach, which is taken by ERA-40, ERA-Interim, ERA5, JRA-25, and JRA-55, is to assimilate screen-level station observations (*i.e.*, temperatures and dewpoint temperatures at 2-m height) in separate two-dimensional OI analyses (*Section 2.3*) of surface air variables (*e.g.*, *Simmons et al.*, 2004). The main benefits include stronger constraints on surface meteorological conditions and their influences on the LSM (see below);

however, this approach can also generate inconsistencies between the upper air and surface fields in the analysis. None of the reanalysis systems use the results of OI surface air analyses to initialize subsequent forecasts, although these analyses can still indirectly affect subsequent forecasts via influences on the land surface state. The second approach, which is taken by all other reanalyses described in this document, omits screen-level station observations from the analysis. Surface air analysis variables over land are still affected by surface pressure and (in the case of full-input reanalyses) upper air measurements assimilated during the standard analysis cycle. This approach establishes weaker observational constraints on the evolution of surface air and land surface conditions in regions where the observational network is dense, but has the

Table 2.14: *Influence of solar cycle on the reanalysis systems.*

Reanalysis System	Influence of the solar cycle
ERA-40	The ~11-year solar cycle is not included in the TSI boundary condition, with the base irradiance assumed to be constant at 1370 W m^{-2}; however, variations in this value due to changes in the distance between the Earth and the Sun have been incorporated as prescribed by *Paltridge and Platt* (1976). A programming error artificially increased the effective TSI by about 2 W m^{-2} relative to the specified value. *Dee et al.* (2011) reported that the impact of this error is mainly expressed as a warming of approximately 1 K in the upper stratosphere; systematic errors in other regions are negligible. The effects of the solar cycle are included in the assimilated upper-air temperatures, but are not included in the ozone passed to the forecast model (see **Table 2.11**).
ERA-Interim	Same as ERA-40.
ERA-20C	ERA-20C uses TSI variations provided for CMIP5 historical simulations by the SPARC SOLARIS-HEPPA working group with the TIM scaling applied, which take values ranging from 1360.2 W m^{-2} to 1362.7 W m^{-2} between 1900 and 2008. These variations account for solar cycle changes through 2008 and repeat the final cycle (April 1996 - June 2008) thereafter. Seasonal variations due to the ellipticity of the Earth's orbit are also included.
ERA5	Same as ERA-20C.
JRA-25 / JCDAS	A constant base TSI of 1365 W m^{-2} was assumed, including seasonal effects due to the ellipticity of the Earth's orbit (*Spencer*, 1971). Interannual variations in incoming solar radiation were not included in the TSI boundary condition, but were included in assimilated temperature and ozone observations.
JRA-55	Same as JRA-25. Note that interannual variations in incoming solar radiation are included in assimilated temperature observations for the whole period, but only included in ozone observations for 1979 and later.
MERRA	MERRA assumes a constant base TSI of 1365 W m^{-2}. Seasonal variations due to the ellipticity of the Earth's orbit are included. Although interannual variations in incoming solar radiation were not included in the TSI boundary condition, these variations could influence the model state through assimilated temperature and ozone observations.
MERRA-2	MERRA-2 uses TIM-corrected TSI variations provided for CMIP5 historical simulations by the SPARC SOLARIS-HEPPA working group, which take values ranging from 1360.6 to 1362.5 W m^{-2} between 1980 and 2008. These variations account for solar cycle changes through 2008 and repeat the final cycle (April 1996 - June 2008) thereafter. Seasonal variations due to the ellipticity of the Earth's orbit are included.
NCEP-NCAR R1	R1 uses a constant TSI of 1367.4 W m^{-2}. The ~11-year solar cycle is not included in the TSI boundary condition, but variations due to changes in orbital geometry are accounted for. The effects of the solar cycle are included in the assimilated upper-air temperatures, but are not included in the ozone passed to the forecast model (see **Table 2.11**).
NCEP-DOE R2	Similar to R1, but with a constant TSI of 1365 W m^{-2}.
CFSR / CFSv2	Annual average variations in TSI were specified according to data prepared by *H. van den Dool (personal communication,* 2006), with values ranging from 1365.7 W m^{-2} to 1367.0 W m^{-2}. The solar cycle after 2006 is repeated forwards (*e.g.*, insolation for 2007 is the same as that for 1996, that for 2008 is the same as that for 1997, and so on). The effects of the solar cycle are included in assimilated temperature and ozone observations; however, the prognostic ozone parameterization does not otherwise account for variations in incoming solar radiation.
NOAA-CIRES 20CR v2	Annual average variations in TSI were specified according to data prepared by *H. van den Dool (personal communication*, 2006), with values ranging from 1365.7 W m^{-2} to 1367.0 W m^{-2}. The solar cycle before 1944 is repeated backwards (*e.g.*, insolation for 1943 equals that for 1954, that for 1942 equals that for 1953, and so on) and the solar cycle after 2006 is repeated forwards (as in CFSR). Upper-air observations were neither assimilated nor included. The prognostic ozone scheme does not account for variations in incoming solar radiation.

Table 2.15: *Information about land surface models and analyses of surface air variables (if applicable) in the reanalysis systems. Surface air station observations are assimilated in ERA-40, ERA-Interim, ERA5, JRA-25, and JRA-55 in analysis steps separate from the standard upper-air analysis cycles. Other reanalyses do not assimilate these data. Additional details are provided in Chapter 2E.*

Reanalysis System	Surface models and analyses of surface air variables
ERA-40	The surface air and land surface analyses are performed outside of the main atmospheric reanalysis. Six-hourly OI analyses of surface air temperature and dewpoint temperature at 2-m height are produced using station observations over land and the background state from the most recent atmospheric analysis. Empirical relationships between surface air fields and soil properties are then used to update soil temperature and soil moisture in a four-level land surface model (*van den Hurk et al.*, 2000).
ERA-Interim	Essentially the same as ERA-40. The additional global land surface reanalysis ERA-Interim/Land was conducted for 1979-2010 using a newer version of the land surface model (*Balsamo et al.*, 2015) with atmospheric forcing from ERA-Interim and precipitation from GPCP.
ERA-20C	Surface pressure and surface winds (over ocean) are the only variables directly constrained by the data assimilation; no land surface analysis is performed. The land surface scheme is based on a new version of the land surface model (*Balsamo et al.*, 2015) relative to that used in ERA-Interim.
ERA5	Similar to ERA-Interim, but with substantial updates to the land surface analysis (*de Rosnay et al.*, 2014) and a new formulation of the LSM that better represents subgrid-scale water bodies and coastlines. A separate global land surface reanalysis ERA5-Land is being conducted with atmospheric forcing from ERA5.
JRA-25 / JCDAS	Surface air temperature, winds, and relative humidity are based on univariate OI analyses that assimilate meteorological station observations. Observation departures are computed relative to the background state at the analysis time rather than at the observation time. Soil temperature and soil moisture on three levels are based on a modified version of the SiB model (*Sato et al.*, 1989; *Sellers et al.*, 1986) forced by atmospheric reanalysis fields applied every 6 h.
JRA-55	Surface air analyses differ from those in JRA-25 in two ways. First, comparisons between observations and the first-guess background state are evaluated at observation times rather than analysis times. Second, screen-level observations over islands are not used as they may not be appropriately representative of conditions at the scale of the model grid cell. Representation of the land surface state is similar to that in JRA-25, but atmospheric forcing is applied every 3 h instead of every 6 h.
MERRA	MERRA did not conduct separate surface air or land surface analyses. Screen-level temperature and humidity measurements over land are not assimilated, although surface air variables in both ANA and IAU products are affected by surface pressure and upper air measurements assimilated during the analysis cycle. Estimates of land surface properties represent the time-integrated effects of coupling between the LSM (*Koster et al.*, 2000) and surface conditions and fluxes generated by the atmospheric model during the IAU "corrector" segment (see *Section 2.3*). A separate land surface analysis (MERRA-Land) was conducted by replacing model-generated precipitation with pentad-resolution GPCP data and using an updated version of the LSM (*Reichle et al.*, 2011).
MERRA-2	Like MERRA, MERRA-2 does not conduct a land surface analysis; however, precipitation inputs to the LSM are primarily based on observations rather than model-generated values between 60°S and 60°N (*Reichle et al.*, 2017a). The reanalysis does not assimilate screen-level temperature or humidity measurements over land. Surface meteorological variables over land thus primarily reflect the net effects of assimilated surface pressures, model-generated surface fluxes (which are directly affected by precipitation corrections), and the upper-air assimilated state (which is not). The LSM features several adjustments relative to MERRA and MERRA-Land (*Reichle et al.*, 2017b).
NCEP-NCAR R1	The reanalysis does not assimilate screen-level temperature or humidity measurements over land, although surface air variables are affected by surface pressure and upper air measurements assimilated during the standard analysis cycle. The land surface analysis includes soil moisture and soil temperature on two layers. Rather than an assimilation, this analysis is constructed by driving the 2-layer OSU LSM (*Pan and Mahrt*, 1987; *Mahrt and Pan*, 1984) using analyses of snow cover (**Table 2.16**) and atmospheric reanalysis fields as forcings. Soil moisture and temperature are relaxed toward a specified climatology.
NCEP-DOE R2	Similar to *R1*, but with precipitation inputs to the LSM corrected for consistency with pentad-mean precipitation data from CMAP. Also, the relaxation of soil variables to climatological values used in *R1* was not used in *R2*.
CFSR / CFSv2	Similar to *R1* and *R2*, but using the 4-layer Noah LSM (*Ek et al.*, 2003). The precipitation forcing is a blended estimate combining pentad-mean CMAP data, the CPC daily-mean gauge-based analysis, and precipitation produced by the atmospheric model. The weights for the blending depend on location, especially latitude. Other forcing data are taken from the coupled atmosphere–ocean reanalysis. The LSM is fully coupled to the atmosphere throughout the diurnal cycle, but the land surface analysis is performed only once per day (at 00UTC) for better consistency with the temporal resolution of the precipitation forcing.
NOAA-CIRES 20CR v2	Surface pressure is the only variable assimilated by the system; no land surface analysis is performed. The model is coupled to the 4-layer Noah LSM (*Ek et al.*, 2003).

benefit of producing a more internally-consistent atmospheric state. Reanalyses using this second approach are mutually independent with respect to external analyses of surface air temperatures over land (e.g., *CRUTEM*; *Osborn and Jones, 2014*); reanalyses using the first approach are not.

Land surface state variables that are simulated by atmospheric reanalyses include soil moisture and soil temperature. Analyses of these variables are not directly affected by data assimilation, but are instead produced by LSMs forced entirely or primarily by the reanalysis atmospheric state. In addition to the different treatments of surface air variables discussed above, a key difference among reanalyses in this respect is the source of the precipitation forcing, which may be taken from the atmospheric model, from observations, or from a combination of the two. The complexity and implementation of the land surface models used by reanalyses also varies widely. These aspects are covered in more detail in *Chapter 2E*.

Table 2.16: *Treatment of snow in the reanalysis systems. Additional details are provided in Chapter 2E.*

Reanalysis System	Treatment of Snow
ERA-40	A snow analysis is performed outside of the main atmospheric reanalysis using Cressman interpolation with successive corrections. Assimilated observations include station observations of snow depth and gridded estimates of snow cover from satellites. Observations of snow depth are limited to Canada before 1966 and to Canada and the former Soviet Union between 1966 and 1976 (*Uppala et al.*, 2005). The snow depth analysis is relaxed toward a climatology when observations are unavailable.
ERA-Interim	Similar to ERA-40.
ERA-20C	Snow depth, albedo, temperature and density are simulated using the model described by *Dutra et al.* (2010).
ERA5	Similar to ERA-Interim, but using a two-dimensional OI analysis (*de Rosnay et al.*, 2015) as opposed to Cressman interpolation. The snow model has also been updated relative to that used by ERA-Interim (*Dutra et al.*, 2010), and the snow depth analysis is no longer relaxed toward a climatology when observations are unavailable.
JRA-25 / JCDAS	A separate OI snow depth analysis is performed once per day. The first-guess background state combines the land-surface analysis and gridded satellite observations. Weekly NOAA snow cover analyses are used in place of gridded satellite observations when the latter are unavailable. The analysis ingests in situ observations of snow depth from selected archives (*Onogi et al.*, 2007).
JRA-55	Some differences relative to JRA-25. The first-guess background state combines the land-surface analysis, gridded satellite observations, and climatological values over ice sheets. Climatological values are used in place of gridded satellite observations when the latter are unavailable. The analysis ingests in situ observations of snow depth from selected archives (*Kobayashi et al.*, 2015).
MERRA	The evolution of snow mass, depth, and heat content is simulated using a three-layer snow model (*Stieglitz et al.*, 2001). No snow analysis is produced.
MERRA-2	Similar to MERRA in most respects; however, a detailed representation of the surface properties of land ice sheets is introduced that includes the evolution of overlying snow layers (*Gelaro et al.*, 2017). No snow analysis is produced.
NCEP-NCAR R1	Snow is treated as a single layer of frozen water with a uniform density. Weekly snow cover analyses from the NSIDC are used for the NH between 1967 and September 1998, after which they are replaced with daily analyses. Snow cover analyses are not available in the SH or in the NH before 1967; climatologies are used instead. Weekly analyses are not interpolated in time, so snow variables change discontinuously every seven days. Model-simulated snow depths are ignored and replaced using an empirical function of model temperature. Several errors have been identified (*Kanamitsu et al.*, 2002; *Kistler et al.*, 2001). For example, the snow cover analysis mistakenly re-used 1973 data for the entire 1974 - 1994 period, and conversion of snow to water during melting was overestimated by three orders of magnitude.
NCEP-DOE R2	Snow is simulated as a single layer of frozen water with a uniform density via a budget equation that accounts for accumulation and melting. Weekly analyses of NH snow cover from the NSIDC are interpolated to daily resolution until September 1998, after which they are replaced with daily analyses. Snow cover analyses are not available in the SH, where model-generated values are used instead. The model-predicted evolution of snow depth is used when it is consistent with ingested snow cover. When this condition is not met, snow is either removed or added, with snow depth in the latter case determined via an empirical function of model temperature.
CFSR / CFSv2	Snow is simulated using a three-layer snow model (*Ek et al.*, 2003; *Koren et al.*, 1999). Simulated snow variables are evaluated and adjusted using external analyses of global snow depth and NH snow cover. These external analyses are not available for dates prior to February 1997, but are used to supplement and correct the snow depth analyses after this date. Model-estimated snow depths are only adjusted if they differ from the analysed depth by more than a factor of two, and are used as is when analysed values are unavailable. A prognostic snow layer is also included in the sea ice model.
NOAA-CIRES 20CR v2	Snow is simulated using a three-layer snow model (*Ek et al.*, 2003; *Koren et al.*, 1999).

Snow cover and its evolution have important impacts on climate (*e.g.*, *Cohen and Entekhabi*, 1999), including the stratospheric circulation and its coupling with the troposphere (*Cohen et al.*, 2014; *Allen and Zender*, 2010; *Orsolini and Kvamstø*, 2009). **Table 2.16** summarizes the models and analysis techniques used to represent snow in reanalyses. Several of the reanalyses produce analyses of snow cover and snow depth using station observations of snow depth. Gridded, observationally-based analyses of snow cover and/or depth may be assimilated as additional constraints, used to help constrain the background state prior to assimilating station observations, or applied (when available) as the primary determinant for the presence or absence of snow. Four of the reanalyses (ERA-20C, MERRA, MERRA-2, NOAA–CIRES 20CRv2) simulate the evolution of snow using snow models forced by the atmospheric reanalysis and the land surface state, with no adjustment based on observational data.

2.3 Assimilation Schemes

2.3.1 Basics of data assimilation

This section provides a brief overview of data assimilation concepts and methods as implemented in current reanalysis systems. More detailed summaries have been provided by *Krishnamurti and Bounoua* (1996), *Bouttier and Courtier* (1999), and *Kalnay* (2003), among others. In this context, an analysis is a best estimate of the true state of the atmosphere at a given time t. Reanalysis systems use objective analysis methods that employ mathematical optimization (data assimilation) techniques to combine model-generated forecasts and observed data, given constraints that are intended to preserve consistency. The results should be reproducible, internally consistent, and spatially continuous.

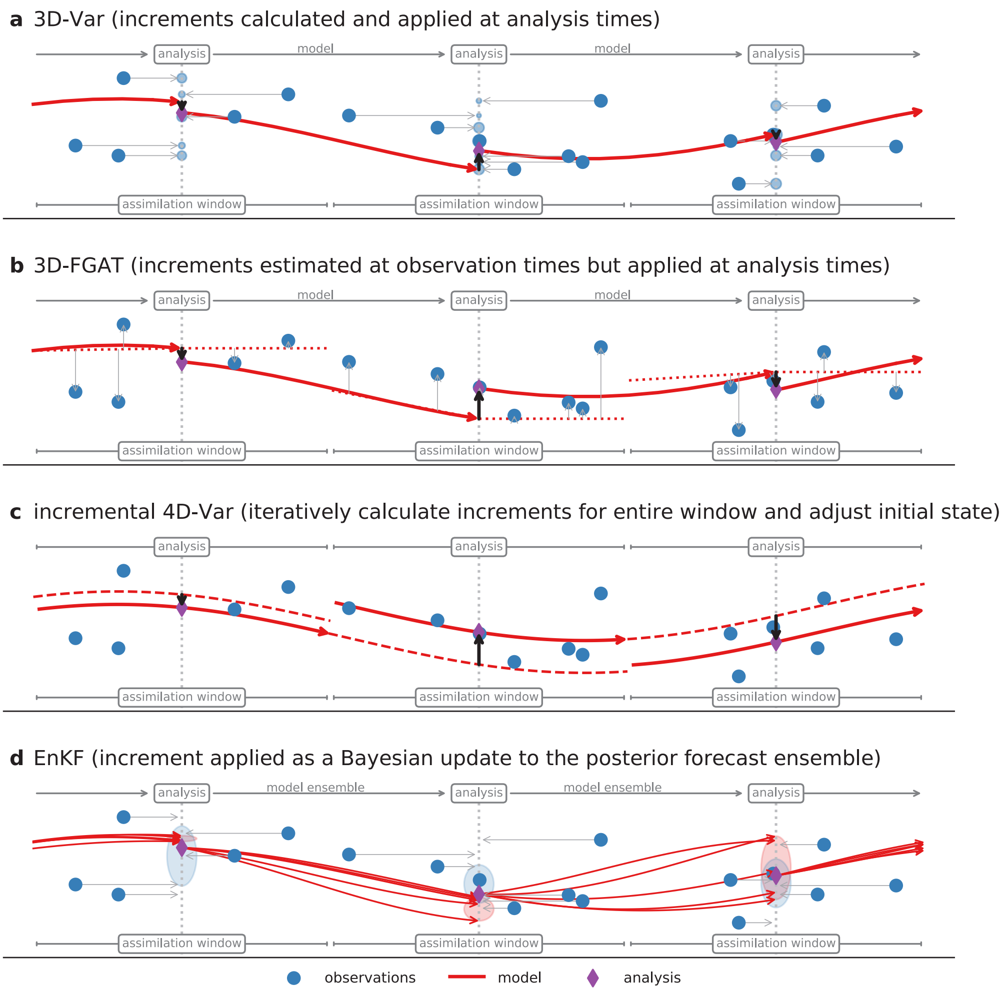


Figure 2.6: *Simplified schematic representations of four data assimilation strategies used by current reanalyses: (a) 3D-Var; (b) 3D-FGAT (here the 'semi-FGAT' approach used by NCEP–NCAR R1 and NCEP–DOE R2 is shown); (c) incremental 4D-Var; and (d) EnKF. Blue circles represent observations, red lines represent the model trajectory, and purple diamonds indicate the analysis. The dotted red lines in (b) represent linearly interpolated/extrapolated first guesses used to estimate increments at observation times. The dashed red lines in (c) represent the initial forecasts, prior to iterative adjustments. These illustrations are conceptual, and should not be taken as exact depictions of the much more complex strategies used by reanalysis systems. Updated from Fujiwara et al. (2017).*

Data ingested into an analysis system may include observations and variables from a first guess background state (such as a previous analysis or forecast). Analysis systems are constructed to be consistent with known or assumed physical properties (such as smoothness, hydrostatic balance, geostrophic or gradient-flow balance, or more complex non-linear balances). Both the observations and the background state include important information, and neither should be considered as 'truth': both the model and observations include errors and uncertainties. An analysis system must therefore adopt a consistent and objective strategy for minimizing the differences between the analysis and the (unknown) true state of the atmosphere. Such strategies are intended to reduce the extent to which errors and uncertainties in both observations and the first-guess background state influence the final analysis state. To this end, data assimilation algorithms often employ statistics to represent the range of potential uncertainties in the background state, observations, and any techniques used to convert between model and observational space (*i.e.*, observation operators), and ultimately aim to minimize these potential uncertainties.

An observation operator (also sometimes referred to as a "forward operator") is a function that converts information from the first guess background state space to the observation space, thus permitting direct comparisons between the model state and observed variables. Different types of observations require different types of observation operators. Key functions performed by observation operators include spatial interpolation from the model grid to observation locations and the transformation of model variables to observed quantities (*i.e.*, the estimation of satellite radiances via the application of a radiative transfer model to the first guess profile; see also **Table 2.19**). Errors in the observation operators constitute a portion of the observation errors considered by the data assimilation scheme.

The analysis methods used by current reanalysis systems include variational methods (3D-Var and 4D-Var) and the ensemble Kalman filter (EnKF). Variational methods (*e.g., Talagrand*, 2010) minimize an objective cost function that simultaneously penalizes differences between the analysis and observations and differences between the analysis and the model background state, with consideration of uncertainties in both the observations and the model. Implementations of variational data assimilation may be applied to derive optimal states at discrete times (3D-Var), or to identify optimal state trajectories within finite time windows (4D-Var). In EnKF (*e.g., Evensen*, 2009), an ensemble of forecasts is used to define a set of background states (the prior ensemble), which is then combined with observations and associated uncertainties to derive a set of analysis states that is consistent with the posterior distribution. The optimal analysis states are determined by applying a Kalman filter (*Kalman*, 1960) to this posterior ensemble (see also *Evensen and van Leeuwen*, 2000). If a single analysis state is required, this is typically derived by averaging the ensemble members, although this approach often leads to fields that are spatially smoother than any of the individual ensemble members, particularly in regions of sharp gradients. One of the key advantages of 3D-Var, 4D-Var, and EnKF methods relative to many earlier implementations of data assimilation is the ability to account for indirect and possibly nonlinear relationships between observed quantities and analysis variables. This ability permits the direct assimilation of satellite radiance data without an intermediate retrieval step (*Tsuyuki and Miyoshi*, 2007), and underpins many of the recent advances in reanalysis development.

Figure 2.6 shows simplified one-dimensional schematic representations of four data assimilation strategies used by current reanalysis systems (3D-Var, 3D-FGAT, 4D-Var, and EnKF). In the following discussion, we frequently refer to the analysis increment, which is defined as the adjustment applied to the first guess (forecast) background state following the assimilation of observational data (*i.e.*, the difference between the analysis state and the first guess background state). We also use the term observation increment, which refers to the difference between the observation and the background state after the observation operator is applied. This concept is also referred to in the literature as the observational 'innovation' (see detailed discussion by *Kalnay*, 2003). The analysis increment reflects the combined adjustment after evaluating and weighting (see also *Section 2.4.2*) all observation increments within an assimilation window, where the assimilation window is the time period containing observations that influence the analysis. The assimilation window used in reanalyses is typically between 6 and 12 hours long but can be as long as 24 hours. This window is often (but not always) centred at the analysis time. Core differences among the data assimilation strategies used in current reanalysis systems can be understood in terms of how the analysis increment is calculated and applied.

The 3D-Var method (**Figure 2.6a**) calculates and applies analysis increments only at discrete analysis times. Observation increments within the assimilation window may either be treated as though they were all at the analysis time (which approximates the average observation time) or weighted by when they occurred (so that observations collected closer to the analysis time have a stronger impact on the analysis increment). JRA-25 uses a 3D-Var method for data assimilation under the former assumption, in which all observations within the assimilation window are treated as valid at the analysis time. In practice, many 3D-Var systems estimate observation increments at observation times rather than analysis times (**Figure 2.6b**). This approach is referred to as 3D-FGAT ('first guess at the appropriate time'; *Lawless*, 2010). The implementation of 3D-FGAT in reanalysis systems varies. For example, *R1* and *R2* are 'semi-FGAT' systems in that observation increments are estimated relative to a linear interpolation between the initial and final states of the forecast before the analysis time and relative to a constant state after the analysis time (*i.e.*, these systems effectively use a pure 3D-Var approach for the portion of the assimilation window after the analysis time). The illustration provided in **Figure 2.6b** corresponds to this semi-FGAT approach. Other 3D-FGAT systems break each forecast into multiple piecewise segments of 30 minutes (ERA-40), one hour (CFSR), or three hours (MERRA and

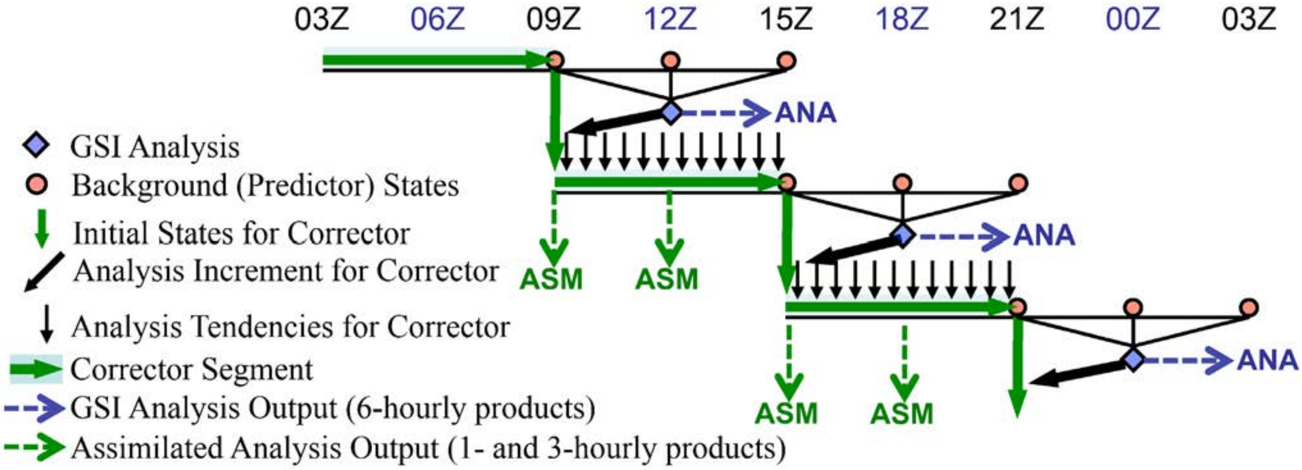


Figure 2.7: *A schematic illustration of the DAS procedure used to create ANA products and the IAU procedure used to create ASM products as implemented in MERRA and MERRA-2 (modified from Rienecker et al., 2011). See text for details.*

MERRA-2) in length. Observation increments are calculated by interpolating to observation times within each piecewise segment and then used to estimate analysis tendencies for each piecewise segment. These analysis tendencies are then combined to construct the full analysis increment.

MERRA and MERRA-2 include an additional step relative to other 3D-FGAT systems, and generate two separate sets of reanalysis products (designated 'ANA' for the analyzed state and 'IAU' for the incremental analysis update state) using an iterative predictor–corrector approach (*Rienecker et al.*, 2011). The ANA products are analogous to the analyses produced by other 3D-FGAT systems, and are generated by using the data assimilation scheme to adjust the background state produced by a 12-h 'predictor' forecast (from 9h before the analysis time to 3h after). The IAU products (also referred to as 'ASM') have no analogue among other 3D-FGAT reanalyses. These latter products are generated by conducting a 6-h 'corrector' forecast centred on the analysis time and using the IAU procedure (*Bloom*, 1996) to apply the previously calculated analysis increment gradually at each model time step rather than abruptly at the analysis time. The corrector forecast is then extended 6 h to generate the next predictor state. This iterative predictor–corrector procedure is illustrated in **Figure 2.7**. Note that the IAU state has only seen half of the analysis increment by the original analysis time, so that differences between the IAU and ANA states correspond to approximately half of the analysis increment. Moreover, the inclusion of the analysis increment as an additional tendency term may alter the physical tendency terms produced by the atmospheric model. For example, diabatic temperature tendencies produced by MERRA and MERRA-2 are archived during the corrector step rather than the predictor step. This arrangement applies to all tendency terms (moisture, momentum, ozone, *etc.*) and introduces a conceptual difference relative to the tendencies produced by other reanalyses (which are archived prior to the analysis during the initial forecast step), though it is important to emphasize that the analysis tendency is needed to close the budget in either case. For MERRA and MERRA-2, ANA products represent the closest match to assimilated observations, while the IAU products provide a more complete and consistent suite of atmospheric variables and tendency terms with reduced wind and tracer imbalances relative to the 3D-FGAT analyzed state (see also **Table 2.18** and associated discussion). IAU products should be used for transport simulations and other applications for which internal consistency is a priority (see also technical note on appropriate use of MERRA-2 products at **https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/docs/ANAvsASM.pdf**). MERRA and MERRA-2 analysis increments for temperature, winds, water vapor, and ozone are included in a subset of the data products provided by these systems.

Unlike 3D-Var and 3D-FGAT, which attempt to optimize the fit between assimilated observations and the atmospheric state at discrete analysis times, 4D-Var (**Figure 2.6c**) attempts to optimize the fit between assimilated observations and the time-varying forecast trajectory within the full assimilation window (*e.g.*, *Park and Županski*, 2003). 4D-Var makes more complete use of observations collected between analysis times than 3D-Var or 3D-FGAT, and has been shown to substantially improve the resulting analysis (*Talagrand*, 2010). However, the computational resources required to run a 4D-Var analysis are much greater than the computational resources required to run a 3D-Var or 3D-FGAT analysis, and the full implementation of 4D-Var is impractical for atmospheric reanalyses. Current reanalysis systems using 4D-Var (such as ERA-Interim, ERA-20C, ERA5, and JRA-55) therefore apply the simplified 'incremental 4D-Var' approach described by *Courtier et al.* (1994). Under this approach, the model state at the beginning of the assimilation window is iteratively adjusted to obtain progressively better fits between the assimilated observations and the forecast trajectory. This iterative adjustment process propagates information both forward and backward in time, which benefits the analysis but requires the derivation and maintenance of an adjoint model. The latter is a difficult and time-consuming process, and is a significant impediment to the implementation of 4D-Var. Incremental 4D-Var is tractable (unlike full 4D-Var), but it is still computationally expensive, and is therefore usually implemented in two nested loops for computational efficiency. Analysis increments are first tested and refined in an inner loop using the tangent linear model (and its adjoint) with reduced resolution and simplified physics. This approach takes advantage of the fact that the cost function for the tangent linear model is perfectly quadratic, thus permitting the use of efficient optimization algorithms designed especially for quadratic functions. The final analysis increments are then applied in an outer loop with full resolution and full physics after the inner loop converges.

Most implementations of variational methods in reanalysis systems are based on single deterministic forecasts. By contrast, EnKF (**Figure 2.6d**) uses an ensemble approach to evaluate and apply analysis increments, thus generating an ensemble of analysis states at each analysis time. Major advantages of the ensemble Kalman filter technique include ease of implementation (unlike 4D-Var, EnKF does not require an adjoint model) and the generation of useful estimates of analysis uncertainties, which are difficult to obtain when using variational techniques with single forecasts (ERA5 uses 4D-Var in a reduced-resolution 'ensemble of data assimilations', in part to address this issue). Although the assimilation of satellite radiances presents some unique challenges in EnKF (P*olavarapu and Pulido*, 2017 ; *Campbell et al.*, 2010), recent work provides approaches to overcome this problem (*Lei et al.*, 2018). *Whitaker et al.* (2009) found that 4D-Var and EnKF perform comparably well in the case of a reanalysis that assimilates only surface pressure observations, and that both 4D-Var and EnKF give more accurate results than 3D-Var in this case. 20CR uses an EnKF method for data assimilation.

As discussed in *Section 2.2.4*, some reanalyses use simpler methods (such as OI or Cressman interpolation) for certain types of data assimilation, especially analyses of screen-level meteorological variables or snow depth. In Cressman interpolation (*Cressman*, 1959), the analysis is iteratively 'corrected' toward the set of observed values, with weighted observation increments that reduce with distance according to a specified window function. The radius of influence defined by this window function is typically reduced on successive iterations so that the closest observations have the largest influence on the final analysis. OI (*Gandin*, 1963) is formulated as a multiple linear regression problem in which both the observations and the background state are assumed to be unbiased, with known random errors. Standard OI is a special case of two of the methods discussed above, and can be functionally equivalent to both 3D-Var (assuming linear observation operators and Gaussian errors) and to the Kalman filter (assuming constant background error covariance). Although the assumptions involved in Cressman interpolation and OI are rarely satisfied, they offer a flexibility in application that can be valuable for estimating analysis increments in variables with highly heterogeneous spatial distributions (such as surface air temperature).

Additional details regarding these methods, including relative advantages and disadvantages, have been discussed and summarized by *Park and Županski* (2003), L*orenc and Rawlins* (2005), *Kalnay et al.* (2007a; 2007b), *Gustafsson* (2007), and *Buehner et al.* (2010a; 2010b), among others.

The assimilation of observational data can introduce spurious artefacts into reanalyses of the state and variability of the upper troposphere, stratosphere, and mesosphere. For example, data assimilation can act to smooth sharp vertical gradients in the vicinity of the tropopause. The potential importance of this effect is illustrated by abrupt changes in vertical stratification near the tropopause at the beginning of the satellite era in *R1* (*Birner et al.*, 2006). Changes in data sources and availability can also lead to biases and artificial oscillations in temperature in various regions of the stratosphere, particularly in the polar and upper stratosphere where observations are sparse (*Lawrence et al.*, 2015; *Simmons et al.*, 2014; *Uppala et al.*, 2005; *Randel et al.*, 2004). Information and errors introduced by the input data and data assimilation system propagate upwards through the middle atmosphere in both resolved waves and parameterized gravity wave drag (*Polavarapu and Pulido*, 2017). The effects of this propagation are often but not always undesirable. The abrupt application of analysis increments can generate spurious gravity waves in systems that use intermittent data assimilation techniques (*Schoeberl et al.*, 2003), including most implementations of 3D-Var, 3D-FGAT, and EnKF, and may also generate instabilities that artificially enhance mixing and transport in the subtropical lower stratosphere (*Tan et al.*, 2004). Although most reanalysis systems use techniques to reduce these effects (see **Table 2.18** in the following section), reanalyses of the stratosphere and mesosphere are nonetheless quite sensitive to the details of the data assimilation scheme and input data at lower altitudes.

2.3.2 Data assimilation in reanalysis systems

Table 2.17 summarizes the schemes used for atmospheric data assimilation in the reanalysis systems, which include variations on the 3D-Var, 3D-FGAT, 4D-Var, and EnKF techniques.

As noted above, the application of analysis increments can generate spurious instabilities in the atmospheric state, particularly when these increments are applied intermittently (as in 3D-Var). Several methods have been developed to mitigate these effects, including nonlinear normal mode initialization techniques and the application of digital filters. Nonlinear normal mode initialization (*Daley*, 1981; *Machenhauer*, 1977) limits the impacts of spurious instabilities by reducing or eliminating the tendencies associated with all "fast-mode" disturbances (*i.e.*, gravity waves) in the vertical and horizontal domains. By contrast, digital filter initialization (*Lynch*, 1993) aims to reduce or eliminate high-frequency noise in the temporal domain. Both approaches can be applied as strong constraints (in which all potentially undesirable modes are eliminated) or as weak constraints (in which potentially undesirable modes are penalized rather than eliminated entirely).

Table 2.17: *List of assimilation schemes used for atmospheric analyses.*

Reanalysis System	Assimilation Schemes
ERA-40	3D-FGAT with a 9-h forecast step ending three hours after the analysis time and a 6-h assimilation window centred on the analysis time. Analysis tendencies are calculated in 30-minute windows and then combined to construct the analysis increment.
ERA-Interim	Incremental 4D-Var atmospheric analysis with 12-h assimilation windows extending from 03 UTC to 15 UTC and from 15 UTC to 03 UTC. Analysis increments are calculated on coarser grids that approach the model resolution over successive iterations.
ERA-20C	Incremental 4D-Var analysis with 24-h assimilation windows extending from 09 UTC to 09 UTC. Like earlier ECMWF reanalyses, assumed background error covariances are invariant in time; however, a scaling is applied for consistency with time-varying background errors produced by an earlier 10-member ensemble pilot reanalysis that also assimilated only surface observations (*Poli et al.*, 2013, 2016).
ERA5	Similar to ERA-Interim, but with assimilation windows extending from 09 UTC to 21 UTC and from 21 UTC to 09 UTC. A 10-member 'ensemble of data assimilations' is conducted on a coarser grid, providing more robust estimates of analysis uncertainties and background error covariances.
JRA-25 / JCDAS	3D-Var (not 3D-FGAT) with 6-h forecast steps. Observations from 3 hours before the analysis to 3 hours afterwards are considered.
JRA-55	Incremental 4D-Var with a 9-h forecast step that extends 3 h past the analysis time and a 6-h assimilation window centred on the analysis time. Analysis increments are calculated on a coarser T106/F80 inner grid (rather than the TL319/N160 outer grid used in the forecast model) to limit computational expense.
MERRA	3D-FGAT using the gridpoint statistical interpolation (GSI; *Wu et al.*, 2002; *Kleist et al.*, 2009) scheme with incremental analysis update (IAU; *Bloom et al.*, 1996) and 6-h assimilation windows centred on each analysis time. The IAU procedure (illustrated in **Figure 2.7**) is summarized in the text.
MERRA-2	GSI with IAU as in MERRA, but with updated background error specifications. A global constraint is imposed on the analysis increment of total water (*Takacs et al.*, 2015).
NCEP-NCAR R1	Spectral statistical interpolation (SSI; Parrish and Derber, 1992) in a 3D-Var 'semi-FGAT' configuration (see text) with a 6-hour assimilation window centred on each analysis time. For times before the analysis time, first guesses are based on linear interpolation between the initial and final model states. For times after the analysis time, first guesses are estimated as the first guess at the analysis time.
NCEP-DOE R2	Same as NCEP-NCAR R1.
CFSR / CFSv2	GSI with 9-h forecasts (from 6 h before to 3 h after each analysis time) and 6-h assimilation windows (centred on each analysis time). The implementation of GSI in CFSR is a form of 3D-FGAT with hourly first guesses.
NOAA-CIRES 20CR v2	Ensemble Kalman filter (EnKF) with a 6-h window centred on each analysis time. Observations from 3 hours before the analysis to 3 hours afterwards are used. The EnKF implementation in 20CR uses a window that straddles the analysis time, and is therefore technically an Ensemble Kalman Smoother (*Compo et al.*, 2011).

Table 2.18: *Initialization procedures used to mitigate assimilation-driven instabilities.*

Reanalysis System	Initialization procedure
ERA-40	Nonlinear normal mode initialization
ERA-Interim	Weak constraint digital filter
ERA-20C	Weak constraint digital filter
ERA5	Weak constraint digital filter
JRA-25 / JCDAS	Nonlinear normal mode initialization
JRA-55	None
MERRA	IAU
MERRA-2	IAU
NCEP-NCAR R1	None
NCEP-DOE R2	None
CFSR / CFSv2	6-h digital filter (*Lynch and Huang*, 1992)
NOAA-CIRES 20CR v2	None

Table 2.19: *List of radiative transfer schemes used for assimilating satellite radiances.*

Reanalysis System	Radiative transfer scheme used for assimilating satellite radiances
ERA-40	RTTOV-5 is used for assimilating satellite radiances.
ERA-Interim	RTTOV-7 is used for assimilating satellite radiances.
ERA-20C	Satellite radiances are not assimilated (see also **Table 2.21**).
ERA5	RTTOV-11 is used for assimilating satellite radiances. Note that where ERA-40 and ERA-Interim only assimilated clear-sky radiances (see also **Table 2.23**), ERA5 assimilates all-sky radiances from certain sensors.
JRA-25 / JCDAS	RTTOV-6 is used for assimilating TOVS radiances and RTTOV-7 is used for assimilating ATOVS radiances.
JRA-55	RTTOV-9 is used for assimilating satellite radiances.
MERRA	The GLATOVS radiative transfer model is used for assimilating SSU radiances; the CRTM is used for assimilating all other satellite radiances.
MERRA-2	All radiances are assimilated using version 2.1.3 of the CRTM.
NCEP-NCAR R1	Satellite radiances are not assimilated (see also **Table 2.21**).
NCEP-DOE R2	Satellite radiances are not assimilated (see also **Table 2.21**).
CFSR / CFSv2	CFSR uses the CRTM developed at NOAA/NESDIS and the JCSDA for assimilating satellite radiances.
NOAA-CIRES 20CR v2	Satellite radiances are not assimilated (see also **Table 2.21**).

Certain data assimilation techniques also aim to reduce the impacts of spurious instabilities and/or eliminate the need for initialization techniques. For example, one of the benefits of the SSI analysis technique (*Parrish and Derber*, 1992) developed at NCEP and used in R1 and R2 was that it imposed a global balance constraint on the analysis that eliminated the need for nonlinear normal mode initialization (*Kalnay et al.*, 1996). It should be noted, however, that balance constraints and filters (particularly those applied as strong constraints) may eliminate real information along with spurious noise. The loss of this information can have particularly detrimental effects in the middle atmosphere, where gravity waves that propagate upward from lower levels play important roles in the dynamics (*Polavarapu and Polido*, 2017). The application of IAU, as in MERRA and MERRA-2, can help to eliminate spurious instabilities without affecting other “fast-mode” disturbances in the model atmosphere. The use of IAU has been shown to improve the representation of the mesosphere in data assimilation systems (*e.g.*, *Sankey et al.*, 2007).

The assimilation of observed satellite radiances by a reanalysis system requires the use of a radiative transfer scheme. This scheme typically differs from that used in the forecast model (**Table 2.4**). **Table 2.19** lists the radiative transfer schemes used by each reanalysis system for assimilating satellite radiances.

2.4 Observational Data

2.4.1 Summary of basic information

This section provides information on key observational data assimilated in the reanalysis systems. Reanalysis systems assimilate observational data from a variety of sources. These sources are often grouped into two main categories: conventional data (*e.g.* surface records, radiosonde profiles, and aircraft measurements) and satellite data (*e.g.* microwave and infrared radiances, atmospheric motion vectors inferred from satellite imagery, and various retrieved quantities).

The densities and distributions of both types of observational data have changed considerably over time. **Figure 2.8** shows examples of the spatial distributions of observations assimilated by JRA-55 in the 1980s (00 UTC, 22 September 1983), while **Figure 2.9** shows examples of the spatial distributions of observations assimilated by the same reanalysis system in the 2010s (00 UTC, 23 June 2010). These two sets of examples are representative of the distribution and number of observations assimilated in most recent reanalysis systems (with the notable exception of ERA-20C and 20CR, which do not assimilate upper-air observations). **Figures 2.10** through **2.13** summarize the availability of different

Table 2.20: *List of codes/acronyms for selected observations assimilated by reanalysis systems.*

SYNOP (conventional)	Surface meteorological observation reported by manned and automated weather stations.
SHIP (conventional)	Surface meteorological observations reported by ships.
BUOY (conventional)	Surface meteorological observations reported by buoys.
PAOBS (conventional)	Surface pressure bogus data for the southern hemisphere. This was a product of human analysts in the Australian Bureau of Meteorology who estimated sea level pressure based on satellite imagery, conventional data and temporal continuity. Production and distribution of PAOBS ceased in mid-August 2010.
AMV (satellite)	Atmospheric motion vectors derived by tracing the movement of individual cloud or water vapour features in successive images from geostationary and polar-orbiting satellites.

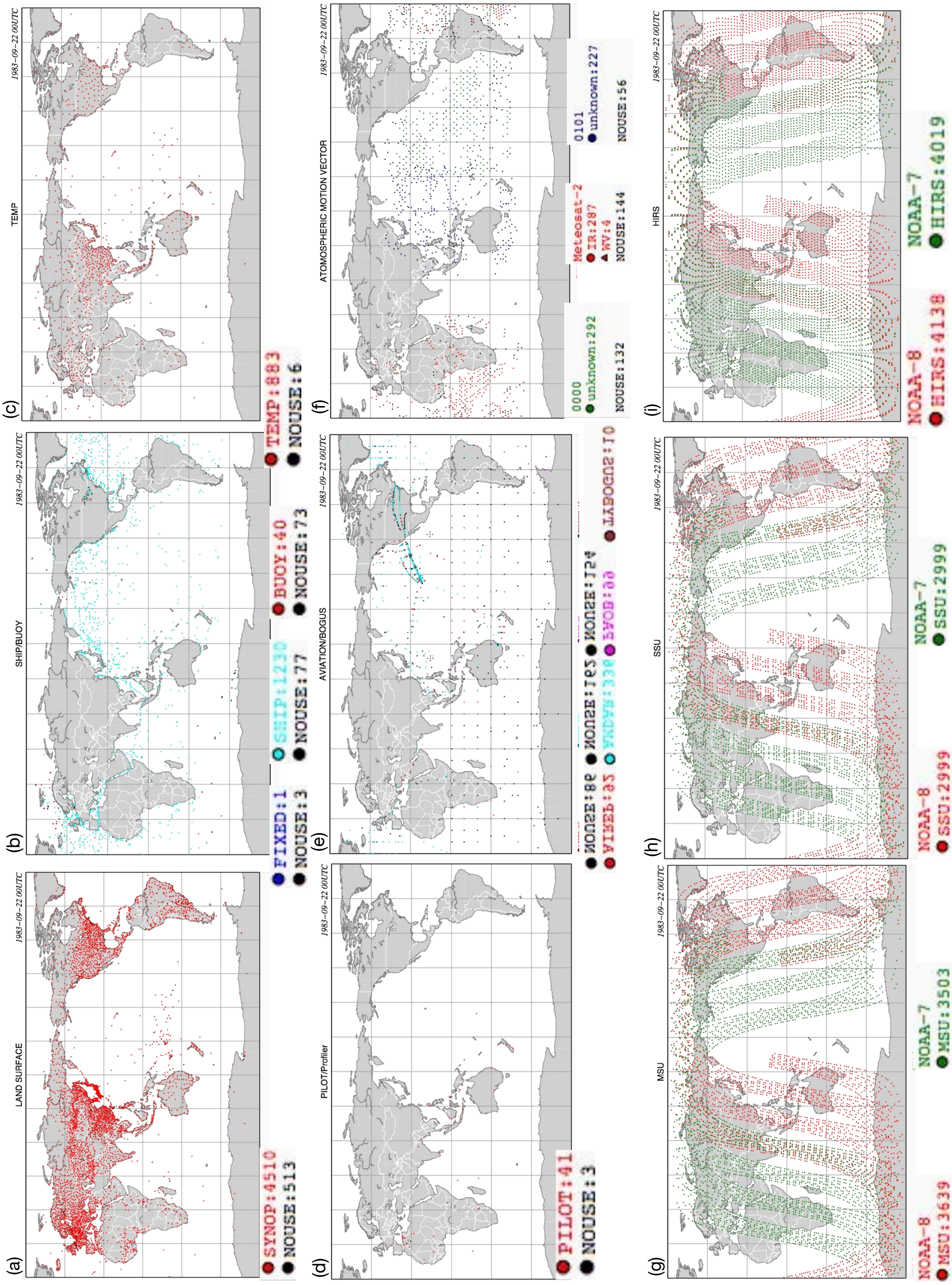
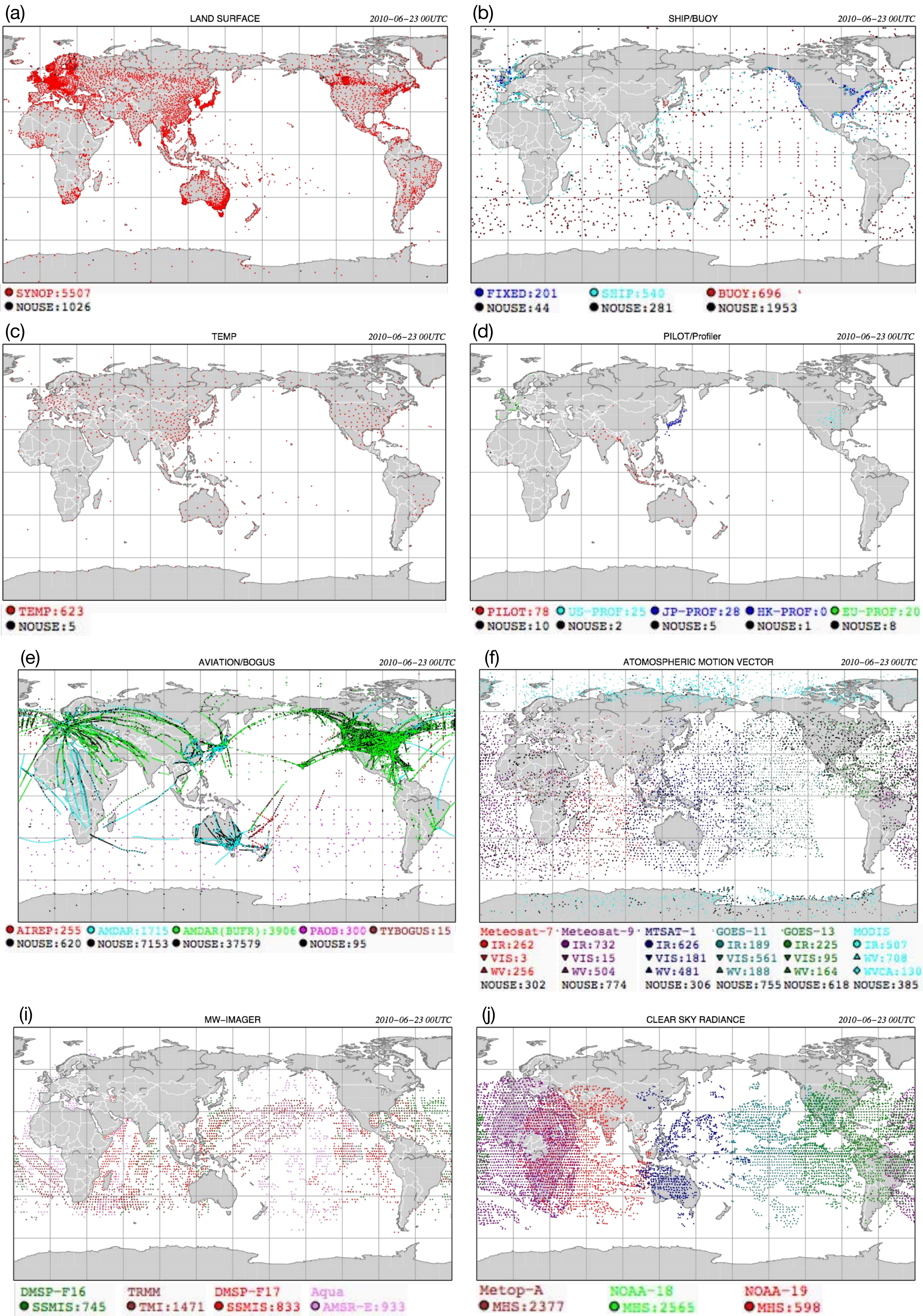


Figure 2.8: *Observations assimilated by JRA-55 at 00UTC 22 September 1983 (±3 hours): (a) land surface data, (b) surface meteorological data reported by ships and buoys, (c) radiosonde profiles, (d) pilot balloons, (e) aircraft, PAOBS, and tropical cyclone wind retrievals, and (f) atmospheric motion vectors from METEOSAT, GMS, and GOES satellites, (g) Microwave temperature sounder radiances from NOAA satellites, (h) stratospheric temperature sounder radiances from NOAA satellites, and (i) infrared sounder radiances (sensitive to temperature and moisture) from NOAA satellites.*


(a)
LAND SURFACE
2010-06-23 00UTC
SYNOP:5507
NOUSE:1026
(b)
SHIP/BUOY
2010-06-23 00UTC
FIXED:201
NOUSE:44
SHIP:540
NOUSE:281
BUOY:696
NOUSE:1953
(c)
TEMP
2010-06-23 00UTC
TEMP:623
NOUSE:5
(d)
PILOT/Profiler
2010-06-23 00UTC
PILOT:78
NOUSE:10
US-PROF:25
NOUSE:2
JP-PROF:28
NOUSE:5
HK-PROF:0
NOUSE:1
EU-PROF:20
NOUSE:8
(e)
AVIATION/BOGUS
2010-06-23 00UTC
AIREP:255
NOUSE:620
AMDAR:1715
NOUSE:7153
AMDAR(BUFR):3906
NOUSE:37579
PAOB:300
NOUSE:95
TYBOGUS:15
(f)
ATOMOSPHERIC MOTION VECTOR
2010-06-23 00UTC
Meteosat-7
IR:262
VIS:3
WV:256
NOUSE:302
Meteosat-9
IR:732
VIS:15
WV:504
NOUSE:774
MTSAT-1
IR:626
VIS:181
WV:481
NOUSE:306
GOES-11
IR:189
VIS:561
WV:188
NOUSE:755
GOES-13
IR:225
VIS:95
WV:164
NOUSE:618
MODIS
IR:507
WV:708
WVCA:130
NOUSE:385
(i)
MW-IMAGER
2010-06-23 00UTC
DMSP-F16
SSMIS:745
TRMM
TMI:1471
DMSP-F17
SSMIS:833
Aqua
AMSR-E:933
(j)
CLEAR SKY RADIANCE
2010-06-23 00UTC
Metop-A
MHS:2377
NOAA-18
MHS:2565
NOAA-19
MHS:598

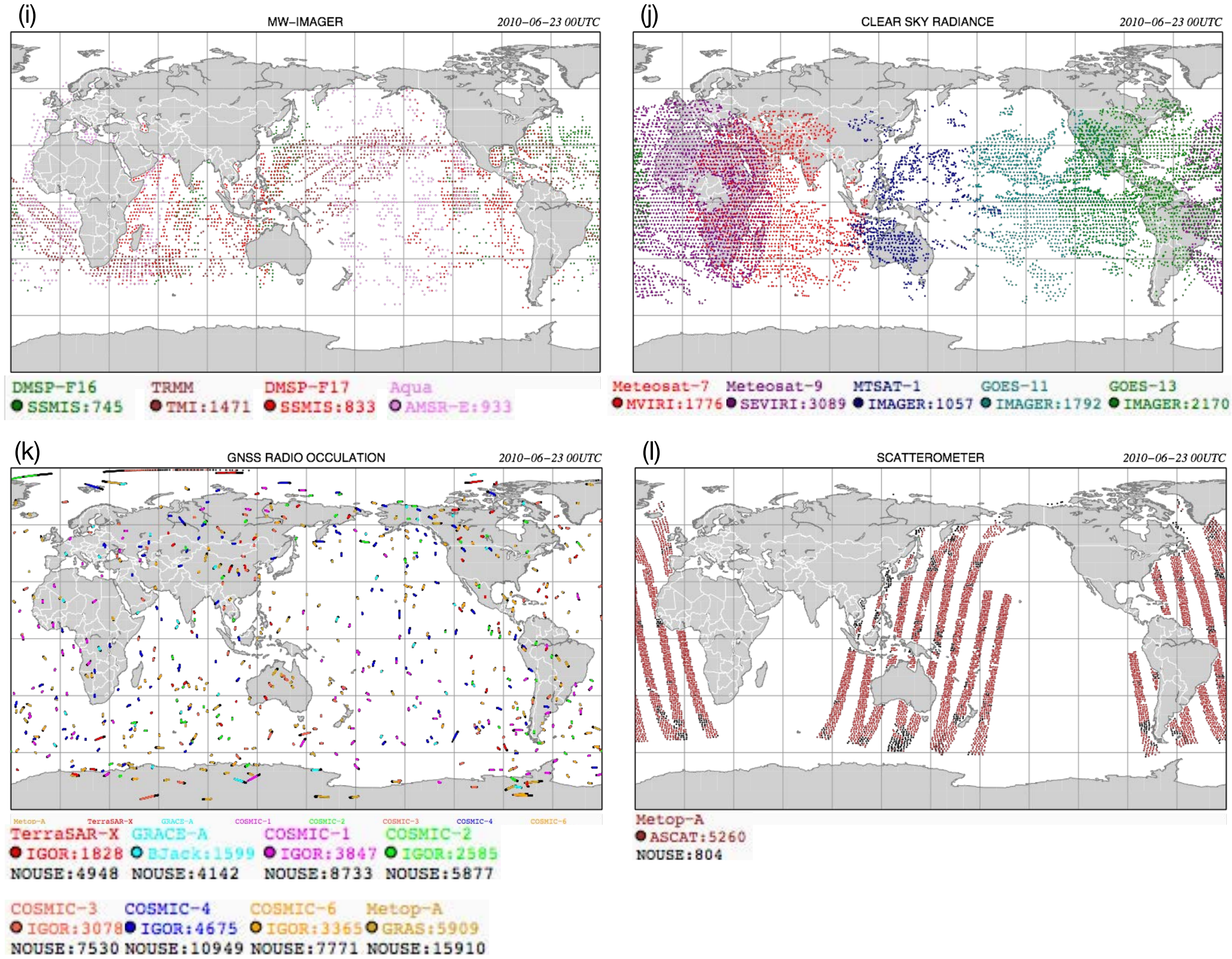


Figure 2.9: *Observations assimilated by JRA-55 at 00UTC 23 June 2010 (±3 hours): (a) land surface data, (b) surface meteorological data reported by ships and buoys, (c) radiosonde profiles, (d) pilot balloons and wind profilers, (e) aircraft, PAOBS, and tropical cyclone wind retrievals, and (f) atmospheric motion vectors from the METEOSAT, MTSAT, GOES, Aqua, and Terra satellites (g) microwave temperature sounder radiances from the NOAA, MetOp, and Aqua satellites, (h) microwave humidity sounder radiances from NOAA and MetOp satellites, (i) microwave imager radiances (sensitive to moisture) from the DMSP, TRMM, and Aqua satellites, (j) clear-sky radiances from METEOSAT, MTSAT, and GOES satellites, (k) GNSS-RO refractive index data (sensitive to temperature and moisture) from the COSMIC, GRACE, MetOp, and TerraSAR-X satellites, and (l) ocean surface winds from MetOp (ASCAT scatterometer).*

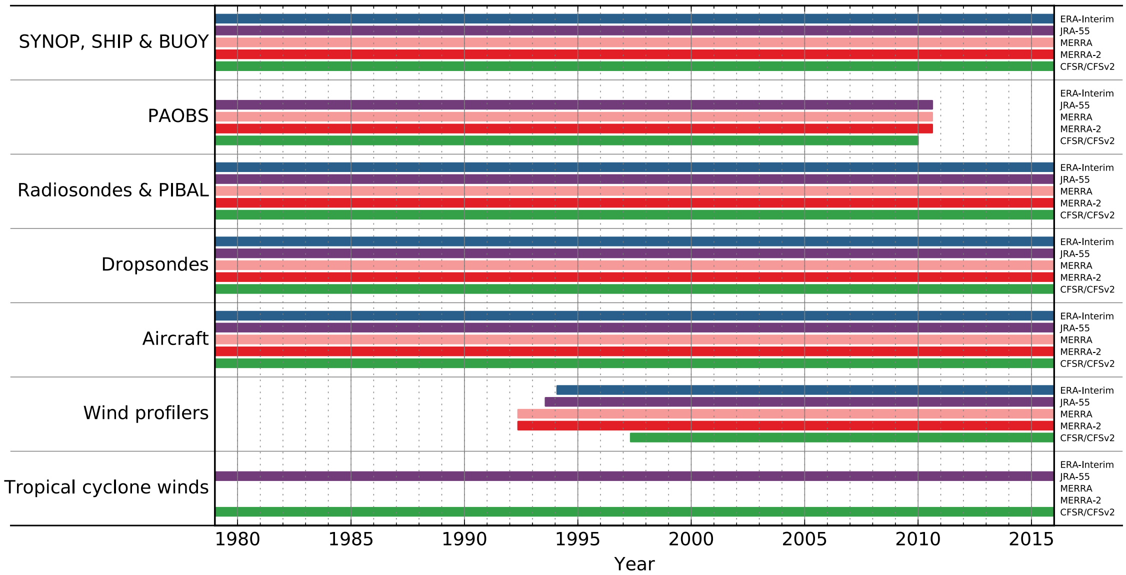


Figure 2.10: *Availability of conventional observations assimilated by ERA-Interim (blue), JRA-55 (purple), MERRA (dark red), MERRA-2 (light red), and CFSR (green) reanalysis systems as a function of time. See **Table 2.20** and Appendix B for acronym definitions. Reproduced from Fujiwara et al. (2017).*

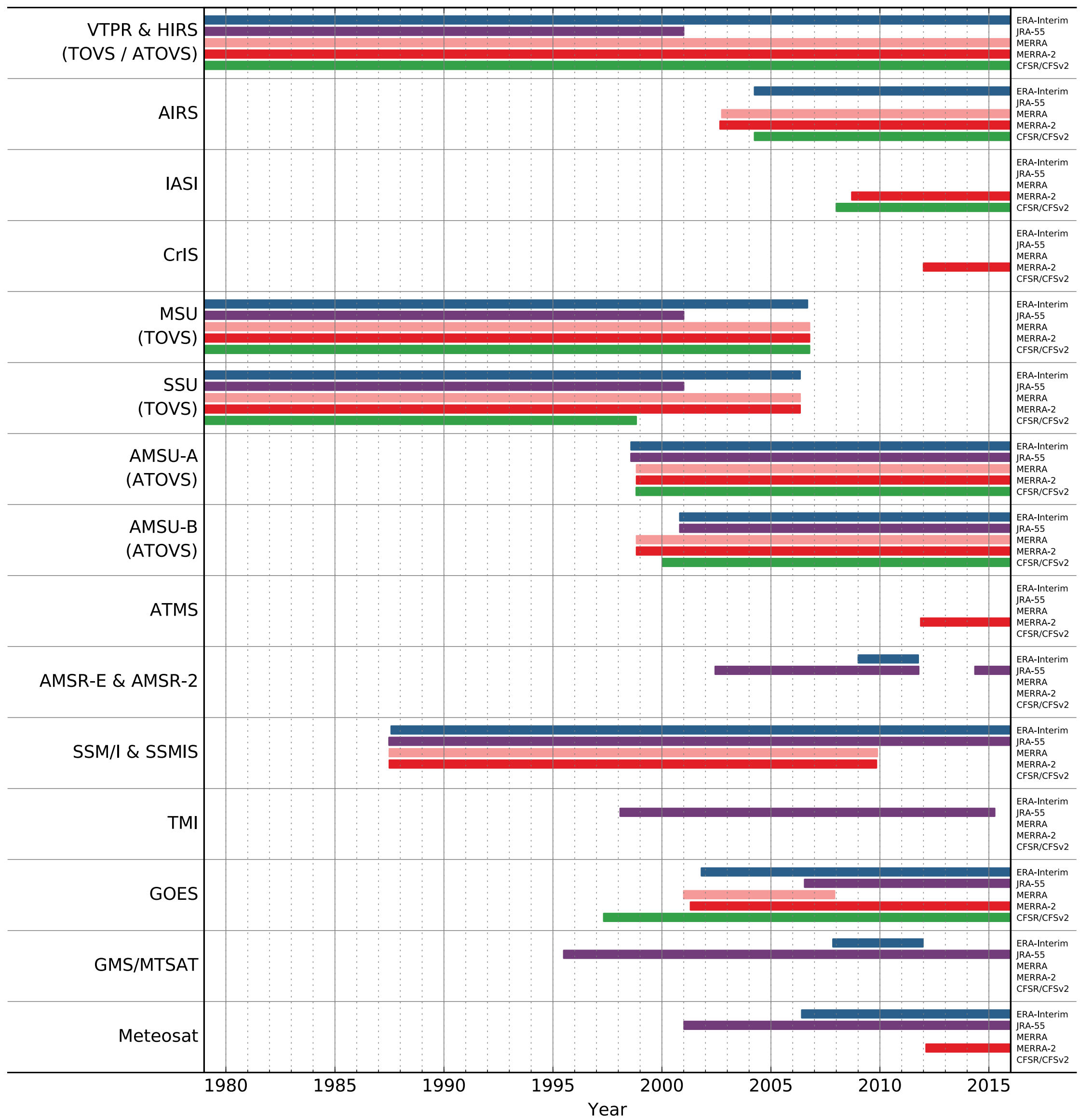


Figure 2.11: *As in **Figure 2.10**, but for satellite radiances assimilated by the reanalysis systems. Reproduced from Fujiwara et al. (2017).*

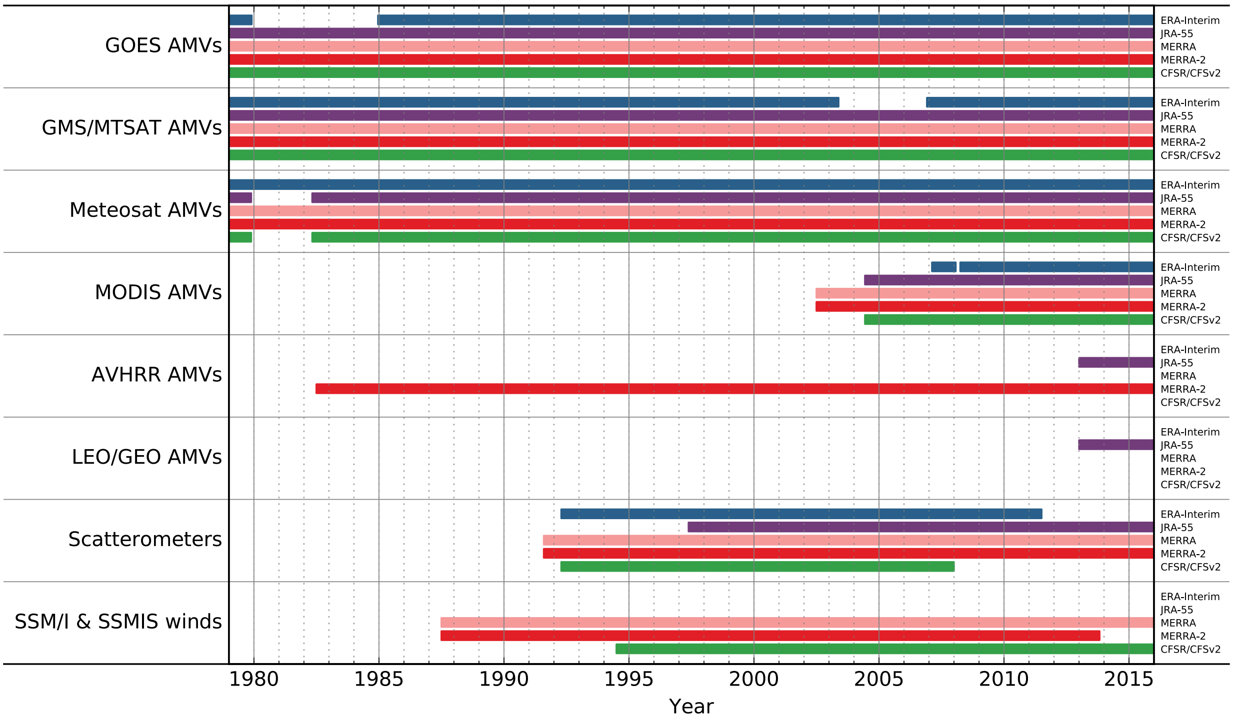


Figure 2.12: *As in **Figure 2.10**, but for AMVs and ocean surface wind products derived from satellites and assimilated by the reanalysis systems. Reproduced from Fujiwara et al. (2017).*

types of observations assimilated in five of the most recent reanalysis systems as a function of time. **Figure 2.14** provides a more detailed look at how the availability of radiances observed by certain instruments changes as satellites are launched and retired. Common codes and terminology for assimilated observations are listed in **Table 2.20**.

Several key details are apparent in **Figures 2.8** through **2.14**. First, conventional in-situ data (such as surface, radiosonde, and aircraft data) are unevenly distributed in space. Second, satellite data (microwave and infrared sounder data, air motion vector data from geostationary and polar satellites, *etc.*) are often more evenly distributed but still inhomogeneous in space. Third, none of these datasets are continuous and homogeneous in time. For example, microwave and infrared sounders (*i.e.*, the TOVS suite) were introduced in 1979, while advanced sounders (*i.e.*, the ATOVS suite) were introduced in 1998. Such changes in the availability of observational data for assimilation have strong impacts on the quality of the reanalysis datasets that assimilate them, so that discontinuities in reanalysis data should be carefully evaluated and checked for coincidence with changes in the input observations. The quality of a given type of measurement is also not necessarily uniform in time; for example, virtually all radiosonde sites have adopted different instrument packages over time (see *Section 2.4.2.1*), while TOVS and ATOVS data were collected using several different sounders on several different satellites with availability that changed over time (see **Figure 2.14** and *Section 2.4.2.2*). Finally, **Figures 2.10** through **2.13** show that, although modern reanalysis systems assimilate observations from many common sources, different reanalysis

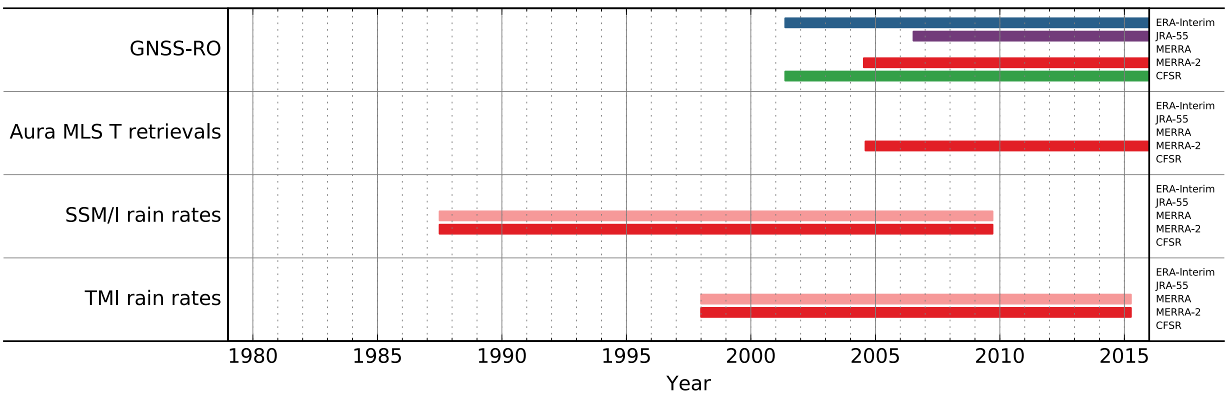


Figure 2.13: *As in **Figure 2.10**, but for other types of satellite observations assimilated by the reanalysis systems. Timelines of satellite retrievals of total column ozone and ozone profiles assimilated by the reanalysis systems are provided in Chapter 4 of this report (**Figures 4.1** and **4.2**). Timelines of GNSS-RO observations assimilated from different sets of sensors are provided in **Figure 2.17**. Reproduced from Fujiwara et al. (2017).*

systems assimilate different subsets of the available observations. Such discrepancies are particularly pronounced for certain categories of satellite observations and, like differences in the underlying forecast models, are an important potential source of inter-reanalysis differences.

Timelines of conventional data assimilated by reanalyses are quite consistent among modern full input reanalyses (**Figure 2.10**), as well as the conventional input JRA-55C (not shown). All of the reanalysis systems discussed in this chapter assimilate records of surface pressure from manned and automated weather stations, ships, and buoys, while all but 20CR assimilate at least some records of surface winds over oceans. All but ERA-Interim, ERA5, ERA-20C, 20CR, and JRA-55C assimilated synthetic surface pressure data for the Southern Hemisphere (PAOBS) through at least 2009. PAOBS are subjective analyses of surface pressure produced by the Australian BOM based on available observations and temporal continuity, which are used to compensate for the scarcity of direct observations in the Southern Hemisphere. The influence of these data in reanalysis systems has waned in recent years, as the availability of direct observations covering the Southern Hemisphere has expanded. All of the full input reanalyses and JRA-55C assimilate upper-air observations made by radiosondes, dropsondes, and wind profilers. JRA-25, JRA-55, and JRA-55C assimilate wind speed profiles in tropical cyclones, while 20CR assimilates records of tropical cyclone central pressures. CFSR uses the NCEP tropical storm relocation package (*Liu et al.*, 1999) to relocate tropical storm vortices to observed locations. ERA5 assimilates PAOBS before 1979 to improve its representation of tropical cyclones during the pre-satellite era. ERA-40, ERA-Interim, MERRA, MERRA-2, NCEP-NCAR R1 and NCEP-DOE R2 have no special treatment for tropical cyclones.

Timelines of satellite data assimilated by current reanalysis systems are more varied (**Figures 2.11** through **2.13**; see also **Figure 2.17** and **Figures 4.1** and **4.2**), but still include many commonalities. The core satellite data assimilated by most reanalyses are microwave and infrared radiances from a variety of instruments. All of the full input reanalyses (including NCEP-NCAR R1 and NCEP-DOE R2) also assimilate atmospheric motion vector (AMV) data derived from geostationary and polar-orbiting satellite imagery. Many of the more recent systems assimilate GNSS-RO data, while MERRA-2 assimilates

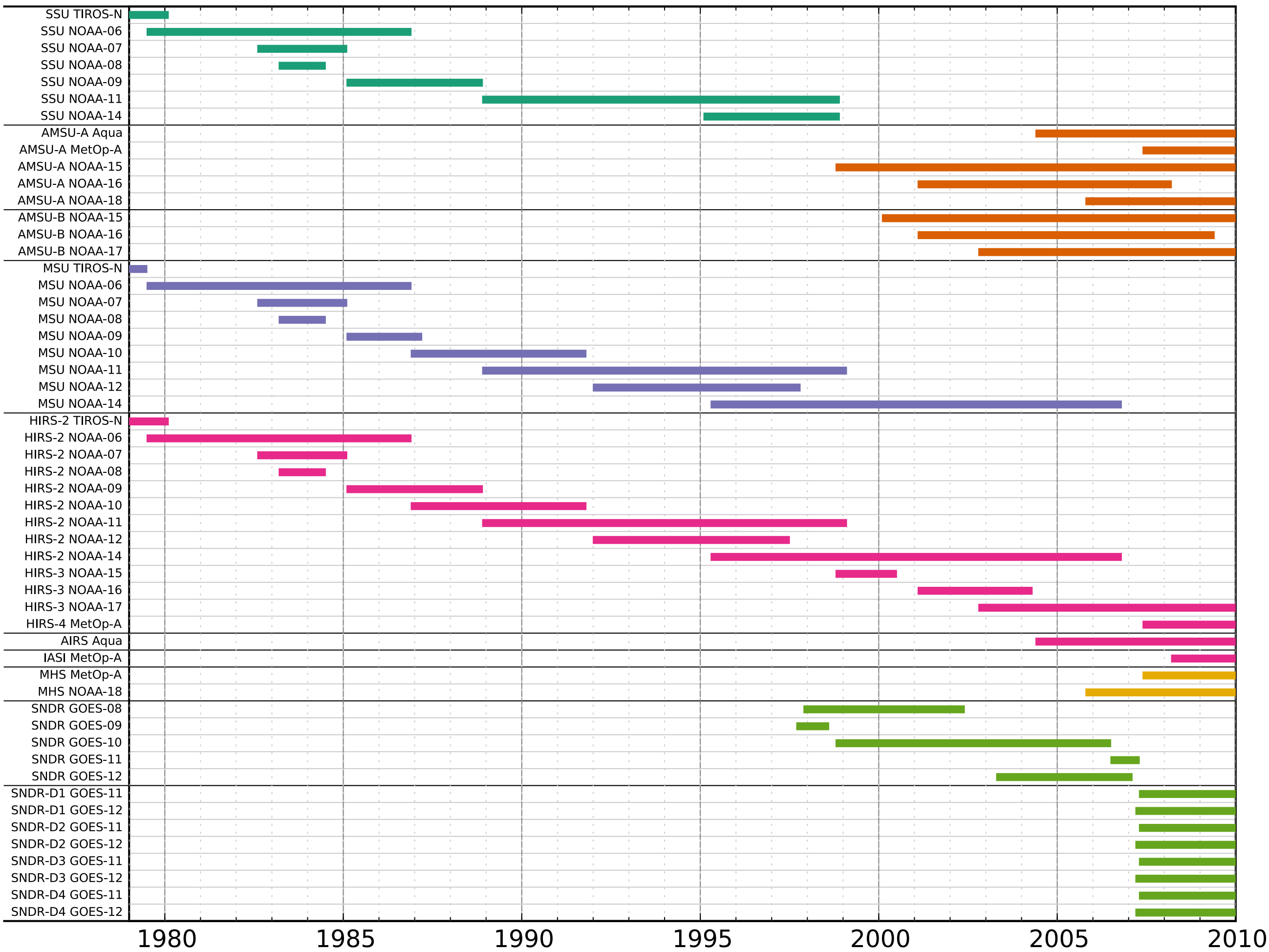


Figure 2.14: *Usage of satellite instruments with radiances assimilated by CFSR as a function of time. Adapted from Saha et al. (2010). Original © American Meteorological Society. Used with permission.*

***Table 2.21:** Special features regarding observational data assimilated in each reanalysis system (see also **Figures 2.10** through **2.13** for five recent full input reanalyses).*

Reanalysis System	Special features of assimilated observational data
ERA-40	SSM/I total column water vapor and surface wind retrievals were assimilated. Neither GNSS-RO data nor AIRS radiances were assimilated (ERA-40 effectively predates these data types). No special treatment for tropical cyclones was included.
ERA-Interim	GNSS-RO bending angles and AIRS radiances are assimilated. Unlike ERA-40, SSM/I radiances are assimilated directly (in place of TCWV and surface wind retrievals). No special treatment for tropical cyclones was included.
ERA-20C	ERA-20C assimilated surface pressure observations from ISPD (*Cram et al.,* 2015) and surface pressure and surface wind observations from ICOADS (*Woodruff et al.*, 2011). Reports that appear in both the IPSD and ICOADS databases were taken from ICOADS, with the IPSD report discarded. Tropical cyclone best track data were assimilated, but with relatively large rejection rates during quality control (*Poli et al.*, 2016).
ERA5	GNSS-RO bending angles are assimilated. AIRS radiances are assimilated, as are hyperspectral radiances observed by IASI and CrIS, microwave soundings from ATMS, and infrared and microwave radiances from several sounding instruments on the Chinese FY-3 series of meteorological satellites. Radiances from several microwave imagers are assimilated directly, including SSM/I and SSMIS, TMI, and GMI, as well as visible and infrared radiances from AHI. Variational bias corrections have been added for ozone, aircraft measurements, and surface pressure. PAOBS are assimilated to improve the representation of tropical cyclones in the pre-satellite era.
JRA-25 / JCDAS	Total column water vapor retrievals from SSM/I and AMSR-E were assimilated, as were wind profile retrievals in tropical cyclones. SSM/I surface winds, GNSS-RO data, and AIRS radiances were not assimilated.
JRA-55	GNSS-RO refractivity data are assimilated, as are wind profile retrievals in tropical cyclones. Clear-sky radiances from selected channels of microwave imagers such as SSM/I, TMI, and AMSR-E are assimilated over ocean (*Kobayashi et al.*, 2015). Neither SSM/I surface winds nor hyperspectral radiances were assimilated. Variational bias corrections have been added for non-blacklisted satellite radiances.
MERRA	AIRS radiances were assimilated, as were rain rates from SSM/I and TMI. SSM/I radiances were assimilated through late 2009, and surface winds were assimilated throughout. GNSS-RO data were not assimilated and no special treatment for tropical cyclones was included.
MERRA-2	GNSS-RO bending angles are assimilated up to 30 km. AIRS radiances are assimilated, as are hyperspectral radiances observed by IASI, CrIS and ATMS. MLS temperature retrievals are assimilated above 5 hPa (version 3.3 through 31 May 2015; version 4.2 from 1 June 2015). A new adaptive bias correction scheme is applied to aircraft observations (see also *Section 2.4.2.3*). Assimilated aerosol optical depths are also bias-corrected. Rain rates from SSM/I and TMI and satellite observations of AOD are assimilated, as are SSM/I surface wind retrievals. SSM/I radiances were assimilated through late 2009. No special treatment for tropical cyclones was included.
NCEP-NCAR R1	Temperature retrievals from microwave and infrared sounders are assimilated, rather than radiances. The horizontal and vertical resolutions of temperature retrievals are downgraded to reduce the weight given to satellite data in recent analyses. Satellite moisture retrievals and SSM/I surface winds are not assimilated.
NCEP-DOE R2	Same as NCEP-NCAR R1.
CFSR / CFSv2	GNSS-RO bending angles and radiances from AIRS and IASI are assimilated. SSM/I radiances are not assimilated, but surface wind retrievals are. The NCEP tropical storm relocation package is applied to relocate tropical storm vortices to observed locations.
NOAA-CIRES 20CR v2	Only observations of surface pressure, sea level pressure, and tropical cyclone central pressure were assimilated. No upper-air or satellite data were assimilated.

temperature retrievals from Aura MLS at pressures 5 hPa and less. Timelines of satellite ozone retrievals assimilated by reanalyses are discussed in Chapter 4 of this report (**Figures 4.1** and **4.2**).

Table 2.21 lists special features of each reanalysis system regarding observational data assimilated. Note that NCEP-NCAR R1 and NCEP-DOE R2 assimilated temperature retrievals from microwave and infrared sounders (*e.g.*, *Reale*, 2001), while the other reanalysis systems (except for surface-input reanalyses) assimilated radiance observations directly. Some systems use bias correction procedures. These are described in *Section 2.4.3*.

2.4.2 Quality control procedures

The observations assimilated by reanalyses are subjected to rigorous quality control procedures that are intended to prevent the introduction of errors into the analysis. Key steps in the quality control algorithm for each reanalysis are listed in **Table 2.22**. Common quality control procedures are briefly described in the following paragraphs (see also *Kalnay, 2003*).

A typical first step in quality control is a preliminary screening. This step eliminates observations with incomplete or duplicate data records, as well as observations that have previously been 'blacklisted' by either the data provider or the reanalysis center.

Many data assimilation systems include automated procedures that try to correct incomplete data records to reduce the number of observations that are eliminated at this stage. The preliminary screening is typically followed by tests to identify and exclude data with physically unreasonable values. The latter may take several different forms. The simplest, the 'gross check', involves comparison against climatological values. Observations are excluded from the analysis if the gross check indicates that they differ from the expected value by more than a specified threshold amount. This type of test may be supplemented (or superseded) by comparison to other reasonable expected values, such as the average of other nearby observations (*i.e.*, a 'buddy check') or the forecast background

Table 2.22: *Standard quality control procedures applied in the reanalysis systems.*

Reanalysis System	Quality control procedure
ERA-40	• Preliminary screening and exclusion of incomplete, duplicate, and blacklisted data • Thinning of selected observation types • Check that the departure from the first-guess is below a threshold that depends on expected error statistics • Variational quality control applied during the analysis step
ERA-Interim	Similar to ERA-40, but with updated thresholds.
ERA-20C	• Preliminary screening and exclusion of incomplete, duplicate, and blacklisted data • In the case of duplicates, precedence is given to ICOADS over ISPD • Wind observations over land and near coasts are excluded • Data are excluded if more than three constant values are reported within a five-day window • Background check eliminates data with departures large (more than seven times expected) relative to the combined error variance from the pilot ensemble • Variational quality control applied during the analysis step
ERA5	Similar to ERA-Interim, but with updated thresholds and additional information from the reduced-resolution ensemble of data assimilations.
JRA-25 / JCDAS	• Preliminary screening and exclusion of incomplete, duplicate, and blacklisted data • Gross check against climatology for most observation types, with thresholds determined using the "dynamic" method proposed by *Onogi* (1998) • Track checks against expected locations for ships, buoys, and aircraft • Complex quality control for radiosondes • Data thinning is applied to AMVs and some TOVS radiances to make the data distribution more uniform
JRA-55	Similar to JRA-25, but thresholds have been reviewed and updated (*Sakamoto*, 2009)
MERRA	• Preliminary screening and exclusion of incomplete, duplicate, and blacklisted data • Check that the departure from the first-guess background state is below a threshold that depends on observation type • Data thinning is applied to all radiance data
MERRA-2	Similar to MERRA, but with revised thresholds for departures from the first-guess background state.
NCEP-NCAR R1	• Complex quality control, including a hydrostatic check and correction • Data exclusion for unrealistic values, duplicate records, ship measurements over land, and blacklisted data • Thinning of selected observation types • Aircraft rejected during certain phases of flight • Background and buddy checks to eliminate observations with large departures • Quality control based on observations within ±24 hours rather than only the assimilation window • Horizontal and vertical thinning of satellite temperature retrievals to reduce the impact of resolution improvements over time
NCEP-DOE R2	Similar to NCEP-NCAR R1.
CFSR / CFSv2	• Complex quality control, including a hydrostatic check and correction • Data exclusion for unrealistic values, duplicate records, ship measurements over land, and blacklisted data • Thinning of selected observation types • Aircraft rejected during certain phases of flight • Variational quality control penalizes observations based on magnitude of departure from the preliminary analysis
NOAA-CIRES 20CR v2	• Pressure observations reduced to sea level and subjected to a gross check against the plausible range 880 to 1060 hPa • Background check eliminates data with departures large (more than three times expected) relative to the combined error variance • Buddy check against nearby observations; can override the results of the background check • Data thinning eliminates observations with weak impacts on the analysis; has the added effect of capping assimilated observations at near mid-20th century levels • Correction of systematic biases (recalibrated every 60 days)

state itself. These comparisons may also be combined, for instance by performing a simple OI analysis using nearby observations (except for the observation being evaluated) and then checking for consistency between the observation and the result of the OI analysis. One benefit of this kind of approach is that it can applied iteratively, rescuing data that might have been excluded by comparison to the initial background state or eliminating data that passed the initial checks but is too far from the OI analysis. In addition to expected values, observations may be checked for consistency with expected balance criteria. For example, height measurements might be compared against heights calculated from virtual temperature measurements via the hypsometric equation. Complex quality control refers to the common practice of applying these checks in combination, and then using an algorithm to decide whether each observation should be included or excluded.

The quality control procedures described above are used to pre-select observational data for use in the analysis. Many 3D-Var and 4D-Var data assimilation systems use variational quality control (*Anderson and Järvinen*, 1999), in which observations that are far from the expected value are penalized in the analysis rather than eliminated entirely. This means that observations that fail to meet the desired criteria have less impact on the analysis, but may still be influential, especially in regions where observations are sparse. Data pre-selection and variational quality control are not mutually exclusive. For example, ERA-Interim conducts a preliminary screening for incomplete, duplicate, and blacklisted data records before starting the incremental 4D-Var assimilation. The initial iterations of the assimilation (see *Section 2.3*) are then conducted without variational quality control, so that all observations that meet the pre-selection criteria are weighted equally. Variational quality control is then turned on for the later iterations of the assimilation to limit the impacts of outlier observations on the final analysis state.

In addition to consistency checks, data may be thinned to reduce redundancy in regions where many observations are available. This procedure can have several benefits, including identifying previously undetected duplicates and reserving an independent set of observations for validating the analysis (*Compo et al.*, 2011). Quality control criteria are also intimately connected to bias correction procedures. Bias corrections may be applied to certain observations either before or during the analysis step to keep otherwise good observations with known biases from being excluded from the analysis. Some typical bias correction procedures for radiosonde, satellite, and aircraft measurements are described in the following section.

2.4.3 Summary of key upper air observations and known issues

This section discusses a selection of upper air observational data that are assimilated in one or more of the reanalysis systems and are key for SPARC sciences. Radiosondes provide high vertical resolution profiles of temperature, horizontal wind, and humidity worldwide; however, most radiosonde stations are located in the Northern Hemisphere at middle and high latitudes over land (**Figure 2.15**). The typical vertical coverage of radiosonde data extends from the surface up to ~30 hPa for temperature and wind and from the surface up to 300~200 hPa for humidity. Operational satellite radiance measurements provide constraints for temperature and moisture with more homogeneous spatial coverage, but at the cost of coarse vertical resolution (*e.g.*, **Figure 2.16**). Moreover, the majority of these measurements were not available before 1978, and no radiance data have been assimilated prior to late 1972 in these reanalyses. Both observing systems have known biases, as well as jumps and drifts in the time series that may cause the quality of reanalysis products to change over time. Bias corrections prior to and/or within the assimilation step are therefore essential for creating more reliable reanalysis products (see below for examples). In addition to radiosonde and satellite data, atmospheric motion vector (AMV) data created from geostationary and polar-orbiter satellite images and wind and temperature observations collected by aircraft are influential in the upper troposphere and lower stratosphere.

2.4.3.1 Radiosonde data

The main source of systematic errors in radiosonde temperature measurements is the effect of solar radiative heating and (to a lesser extent) infrared cooling on the temperature sensor (*Nash et al.*, 2011). This issue, which is sometimes called the 'radiation error', can cause particularly pronounced warm biases in raw daytime stratospheric measurements. These biases may be corrected onsite in the ground data receiving system before reporting, and further corrections may be applied at each reanalysis centre before assimilation. The major issue with radiosonde humidity measurements is that the sensor response is too slow at cold temperatures (*Nash et al.*, 2011). Recent advances in radiosonde instrumentation are beginning to improve this issue, particularly in the upper troposphere; however, radiosonde observations of humidity at pressures less than 300 hPa are typically not assimilated by reanalysis systems. Other issues include frequent (and often undocumented) changes in radiosonde instrumentation and observing methods at radiosonde stations, which may cause jumps in the time series of temperature and relative humidity. Several 'homogenization' activities for radiosonde temperature data exist to support climate monitoring and trend analyses (see, *e.g.*, *Seidel et al.*, 2009). Although some of these activities have been effectively independent of reanalysis activities, cooperation between the two groups has increased substantially in recent years. Particularly notable is the production of RAOBCORE (*Haimberger et al.*, 2008, 2012), which was conducted with reanalysis applications in mind. One or more versions of RAOBCORE are used in ERA-Interim (v1.3), MERRA and MERRA-2 (v1.4 through 2005), and JRA-55 (v1.4 through 2005; v1.5 thereafter). ERA5 uses the RICH dataset (v1.5.1) rather than RAOBCORE. Further efforts on data rescue, reprocessing, homogenization, and uncertainty evaluation by the broader research community are likely to be an essential part of the next generation of reanalyses (*e.g., ACRE (Allan et al., 2011), and GRUAN (Bodeker et al., 2016)*).

The following example describes a 'homogenization' (or bias correction) of radiosonde temperature measurements for assimilation in a reanalysis system:

i. Radiosonde temperatures are corrected for estimated biases from 1980 onwards;
ii. Stations are separated into groups representing different countries or regions (because stations within the same country often use the same type of radiosonde from the same manufacturer);
iii. Mean differences between background forecasts and observations are accumulated for each group of stations;
iv. The mean error for all groups is subtracted from the bias computed for each group to provide a correction for radiation effects;

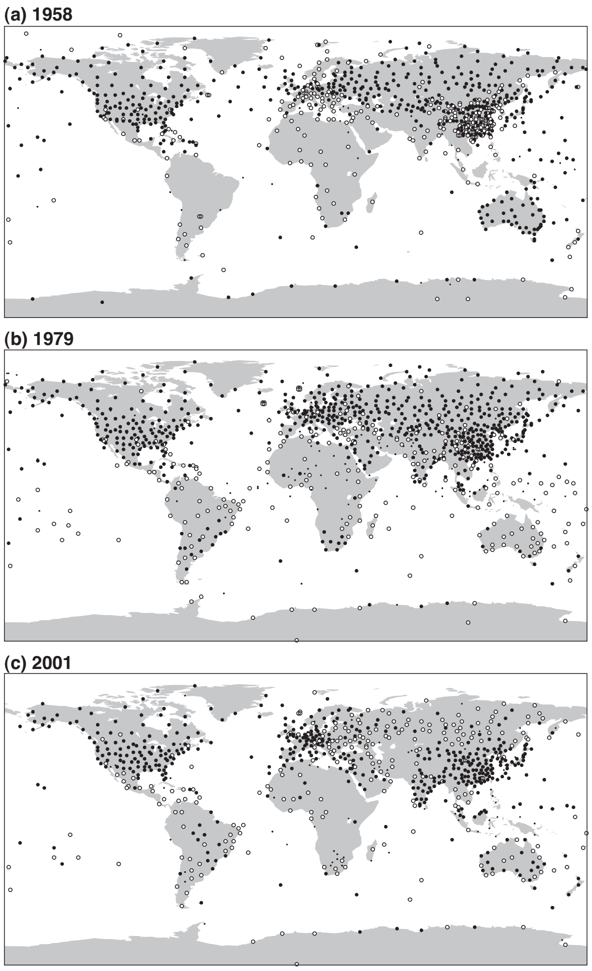


***Figure 2.15:** Frequency of radiosonde reports assimilated by ERA-40 during (a) 1958, (b) 1979, and (c) 2001. Solid circles denote stations reporting three times every 2 days on average, open circles denote stations reporting at least once every 2 days, and small dots denote stations reporting at least once per week (from Uppala et al., 2005). ©Royal Meteorological Society. Used with permission.*

This approach corrected for many daily and seasonal variations of the biases but did not account for variations in annual mean biases. Radiosonde temperature measurements homogenized using this approach were assimilated in both ERA-40 and JRA-25 (*Onogi et al.*, 2007; *Uppala et al.*, 2005; *Andrae et al.*, 2004). The homogenizations applied to produce the RAOBCORE temperatures assimilated by many later reanalyses (including ERA-Interim, JRA-55, MERRA, and MERRA-2, as discussed above) have been conducted using updated versions of this procedure. Although radiosonde humidity measurements are also known to suffer from biases, current reanalysis systems do not include schemes to correct for biases in radiosonde humidities.

Major quality control criteria for radiosonde profiles (and other conventional data) include checks for completeness, physical and climatological consistency, and duplicate reports (*Section 2.4.2*). Data may also be filtered using locally compiled blacklists or blacklists acquired from other data providers and reanalysis centres. Further information on the quality control criteria applied by different reanalysis is available in the text and supporting material of the publications listed in **Table 2.1**.

Radiosonde and other upper-air in situ data are also often shared among different reanalysis centres. For example, *Rienecker et al.* (2011) listed the sources for historical radiosonde, dropsonde, and PIBAL data used by MERRA as:

i. NCEP–NCAR: Office Note 20, Office Note 29, NMC/NCEP/GTS ingest;
ii. ECMWF: ECMWF/FGGE, ECMWF/MARS/GTS ingest;
iii. JMA: Japan Meteorological Agency GTS ingest;
iv. NCAR: International archives from Argentina, Australia, Brazil, Canada, China, Dominica, France, India, Japan, NCDC, New Zealand, Russia, Singapore, South Africa, United Kingdom Research sets: PermShips, RemoteSites, Ptarmigan, Scherhaug, LIE, GATE and BAS;
v. NCDC: U.S. military and academic sources, including TD52, TD53, TD54, TD90, USCNTRL, USAF, U.S. Navy, CCARDS and MIT.

These data sources overlap substantially with those used in ERA-40 and ERA-Interim (*Tavolato and Isaksen*, 2011; *Uppala et al.*, 2005, their Appendix B), JRA-25 and JRA-55 (*Kobayashi et al.*, 2015, their Table A1; *Onogi et al.*, 2007, their section 2.1a), MERRA-2 (*McCarty et al.*, 2016), NCEP-NCAR R1 (*Kalnay et al.*, 1996, their Section 3a), and CFSR (*Saha et al.*, 2010, their section "Conventional observing systems in the CFSR"); however, individual reanalyses may supplement standard data sets with data from unique sources. A detailed intercomparison of the conventional data used in each reanalysis is beyond the scope of this chapter; however, we note that at least four of the reanalyses (ERA-40, ERA-Interim, JRA-25, and JRA-55) use the ERA-40 ingest as a starting point, and that the ERA-40 ingest has much in common with the conventional data archives used by NCEP (*R1*, *R2*, and CFSR) and the NASA GMAO (MERRA and MERRA-2). More recent updates in data holdings at ECMWF, JMA, GMAO, and NCEP rely heavily on near-real-time data gathered from the WMO GTS, which also contributes to the use of a largely (but not completely) common set of conventional data among reanalysis systems.

2.4.3.2 Satellite data

Reanalysis systems assimilate data from several different types of satellite instruments, most notably the microwave and infrared sounders in the TOVS suite (1979–2006 on several satellites) and the ATOVS suite (1998–present on several satellites). The TOVS suite included the Stratospheric Sounding Unit (SSU), the Microwave Sounding Unit (MSU), and the High-resolution Infrared Sounder-2 (HIRS/2). The ATOVS suite includes the Advanced MSU-A (AMSU-A) and HIRS/3 (updated to HIRS/4 starting with NOAA-18). NCEP-NCAR R1 and NCEP-DOE R2 assimilate temperature retrievals from these instruments (see, *e.g.*, *Reale*, 2001). All of the other full input reanalyses described in this chapter assimilate microwave and infrared radiances from the TOVS and ATOVS suites. ERA-Interim, ERA5, MERRA, MERRA-2, and CFSR also assimilate radiances from AIRS, the first hyperspectral infrared sounder with data assimilated in reanalyses (2002–present). ERA5, MERRA-2, and CFSR assimilate hyperspectral infrared radiances from IASI (2008-present), while ERA5 and MERRA-2 also assimilate radiances from the hyperspectral infrared sounder CrIS and the most recent generation of microwave sounder ATMS (late 2011-present). ERA-Interim, ERA5, JRA-55, MERRA-2, and CFSR assimilate data from GNSS-RO instruments (CHAMP: 2001–2008; FORMOSAT-3/COSMIC: 2006–present; MetOp-A: 2008–present; and several other recent missions), in the form of bending angles or refractivity at the tangent point rather than temperature or water vapour retrievals.

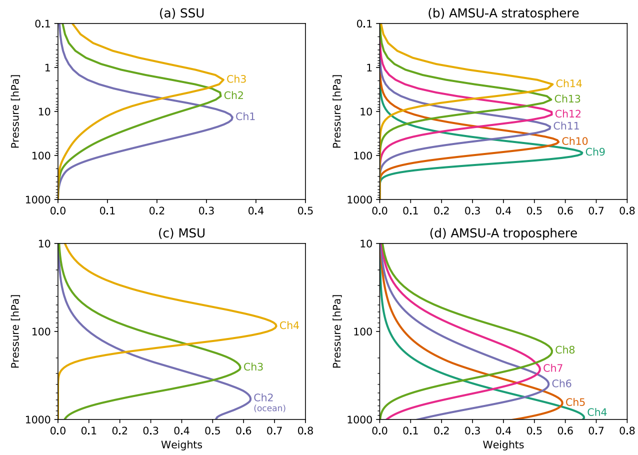


Figure 2.16: *Vertical weighting functions of radiance measurements for (a) SSU (1979–2005) channels 1–3, (b) AMSU-A (1998–present) stratospheric temperature channels 9–14, (c) MSU (1979–2006) channels 2–4, and (d) AMSU-A tropospheric temperature channels 4–8. Weighting functions are for nadir or near-nadir scan positions and have been normalized as described by Zou and Qian (2016).*

Satellite sounding instruments often have several channels with different vertical weighting functions (see, *e.g.*, **Figure 2.16**). Even when using the same satellite instrument, different reanalysis systems may assimilate data from different sets of channels. Bias corrections and quality control criteria for satellite radiances may also vary by channel. **Table 2.23** lists details of satellite data usage for five of the full input reanalysis systems considered in this chapter.

Radiances observed by the SSU instruments, which covered the period 1979–2005, represent an important archive of stratospheric temperatures (*e.g.*, *Nash*

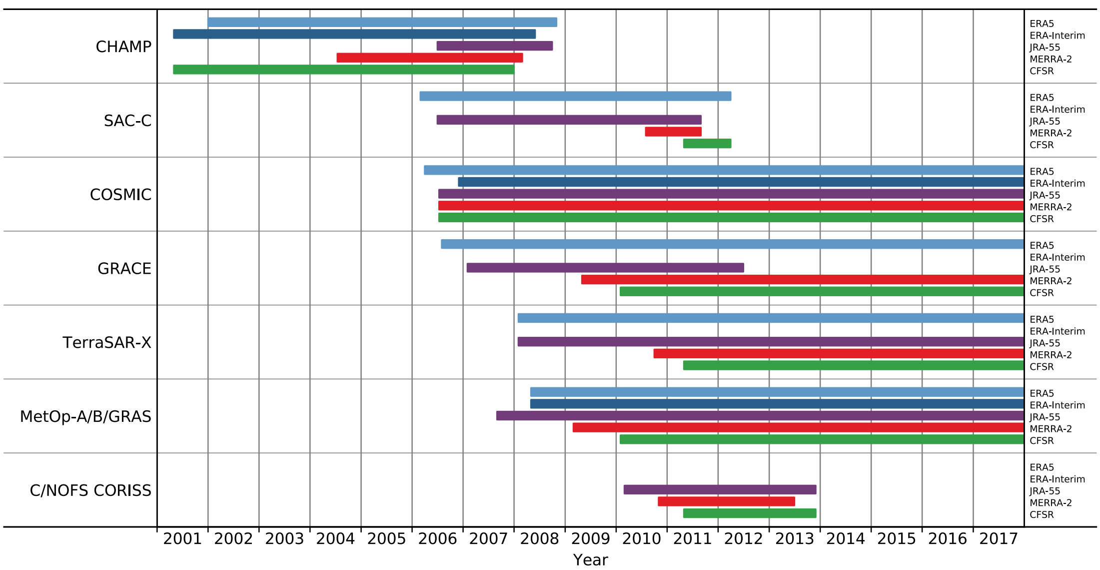


Figure 2.17: *Assimilation of GNSS-RO observations from different campaigns by five recent reanalyses.*

***Table 2.23:** Overview of satellite data usage in five of the most recent full input reanalysis systems. Adapted and updated from **http://reanalyses.org/observations/satellite-1**. Refer to the website for source information and the latest version of this table (including information for JRA-25/JCDAS). See Appendix for acronym definitions.*

Instrument (observable)	CFSR / CFSv2	MERRA	MERRA-2	JRA-55	ERA-Interim
MSU (radiances)	Channels 1,2,3,4 **Notes**: NESDIS SNO corrected calibration coefficients applied (NOAA-10 to -14) **Exclusions**: • More restrictive QC in tropics and over high terrain • Window test ch. 2	Channels 1,2,3,4 **Notes**: NESDIS SNO corrected calibration coefficients applied **Exclusions**: • Snow, ice, mixed surfaces for ch. 1 – 2	Channels 2,3,4 **Notes**: NESDIS SNO corrected calibration coefficients applied **Exclusions**: • Restrictive QC over snow, ice and mixed surfaces • Observation errors inflated over non-water surfaces	Channels 2,3,4 **Exclusions**: • Land or rain for ch. 2 • Land for ch. 3	Channels 2,3,4 **Exclusions**: • Land or rain for ch. 2 • Land for ch. 3
AMSU-A (radiances)	Channels 1 – 13, 15 **Exclusions**: • Estimated cloud liquid water large for ch. 1 – 5, 15 • Scattering index large for ch. 1 – 6, 15 • Ch. 4 gross check large for ch. 1 – 5, 15 • Ch. 6 gross check large for ch. 1 – 6, 15 • High terrain for ch. 1 – 5, 15 • Fit to emissivity or surface temp large for ch. 1 – 5, 15	Channels 1 – 15 **Exclusions**: • Snow, ice, mixed surfaces for ch. 1– 6, 15 • No offset bias correct for ch. 14	Channels 4 – 14 **Exclusions**: • Restrictive QC • Observation errors inflated for ch. 4 – 6 over non-water surfaces	Channels 4 – 14 **Exclusions**: • Sea ice or land for ch. 4 – 5 • High terrain for ch. 6 – 7 • Rain for ch. 4 – 8	Chanels 5 – 14 **Exclusions**: • High terrain for ch. 5 – 6 • Rain for ch. 5 – 7 • No offset bias correct for ch. 14
AMSU-B / MHS (radiances)	Channels 1 – 5 **Exclusions**: • Scattering index too large • Channel 1 fit too large • Any channel failing gross check • High terrain	Channels 1 – 5 **Exclusions**: • Snow, ice, mixed surfaces for ch. 1, 2, 5	Channels 1 – 5 **Exclusions**: • Restrictive gross check • Observation errors inflated for all channels over non-water surfaces	Channels 3 – 5 **Exclusions**: • Land, sea-ice, rain	Channels 3–5 **Exclusions**: • Sea ice, rain, high terrain for ch. 3 – 4 • Land for ch. 5
SSM/I (radiances)		Channels 1 – 7 **Exclusions**: • Land	Channels 1 – 7 **Exclusions**: • All non-water surfaces	Channels 1,3,4,6 **Exclusions**: • Land, rain	Channels 1 – 7 **Exclusions**: • Land, rain
HIRS (radiances)	Channels 2 – 15 **Exclusions**: • Over water wavenumbers >2400 during day • High terrain • Above model top • Channels without signal over clouds • Surface sensing channels with large difference.	Channels 2 – 15 **Exclusions**: • Land for ch. 5 – 8	Channels 2 – 12 **Exclusions**: • Surface-sensitive channels • Observation errors inflated over non-water surfaces	Ch. 2 – 7,11,12,14,15 **Exclusions**: • Land for ch. 4-7, 11,14,15 • High terrain for ch. 12 • Clouds for ch. 3 and above	Ch. 2 – 7,11,12,14,15 **Exclusions**: • Clouds, land for ch. 4 – 7,11,14,15 • High terrain for ch. 12
SSU (radiances)	Channels 1 – 3 **Notes**: • All channels bias-corrected.	Channels 1 – 3 **Notes**: • No offset bias correction for ch. 3	Channels 1 – 3 **Notes**: • Only ch. 1–2 after onset of NOAA-15 AMSU-A (1 Nov 1998) • No offset bias correction for ch. 3	Channels 1 – 3	Channels 1 – 3 **Notes**: • No offset bias correction for ch. 3
GEO (radiances)	GOES sounder **Notes**: • 5°×5° 1993–2007 • 1°×1° 2007–present	GOES sounder	GOES, Meteosat (after early 2012)	GOES, METEOSAT, GMS, MTSAT imagers	GOES, METEOSAT, MTSAT imagers
SSM/I (retrievals)	• Surface wind speed over oceans	• Surface wind speed ov. oceans • Rain rate	• Surface wind speed over oceans • Rain rate	• Snow cover	• Total column water vapor (rainy areas over oceans)
Imager (upper-air winds)	GOES, METEOSAT, GMS, MTSAT, MODIS	GOES, METEOSAT, GMS, MTSAT, MODIS	GOES, METEOSAT, GMS, MTSAT, MODIS	GOES, METEOSAT, GMS, MTSAT, MODIS	GOES, METEOSAT, GMS, MTSAT, MODIS
Scatterometer (winds over ocean surface)	ERS, Quikscat, ASCAT	ERS, Quikscat	ERS, Quikscat, ASCAT	ERS, Quikscat, ASCAT	ERS, Quikscat
Ozone sensors (retrievals)	SBUV V8 retrievals	SBUV V8 retrievals	SBUV V8 retrievals, OMI, MLS (v2.2 through 31 May 2015, switching to v4.2 from 1 June 2015; 261 hPa switched off from 1 May 2016)	TOMS, OMI (nudging)	TOMS, SBUV, GOME, MIPAS, SCIAMACHY, MLS, OMI
Other notable elements	• AIRS • IASI • GNSS-RO • AMSR-E • Reprocessed ERS • Reprocessed GMS • AMSU-B (NOAA-15 only)	• TMI rain rate • AIRS • NOAA-15 AMSU-B	• TMI rain rate • AIRS • IASI • CrIS • GNSS-RO • NOAA-15 AMSUB • ATMS • SEVIRI • MLS temperature retrievals (v3.3 through 31 May 2015, switching to v4.2 from 1 June 2015) above 5 hPa • AOD from MISR, MODIS, AVHRR and AERONET	• Reprocessed winds from GMS, GOES-9, MTSAT (revised) and METEOSAT • Reprocessed radiances from GMS, GOES-9, MTSAT • TMI (NASA) • AMSR-E (JAXA) • GNSS-RO • SSM/I-S • VTPR • Exclude HIRS from NOAA-15 and later	• GNSS-RO • AIRS • SSM/I-S • AMSR-E • HIRS NOAA-18

and Saunders, 2015; *Zou et al.*, 2014; *Wang et al.*, 2012) and serve as a useful illustration of the types of issues that may be encountered in assimilating satellite data. The SSU was a pressure-modulated radiometer with an on-board CO_2 cell for spectral filtering at 15 µm. The calibration of SSU radiances is affected by the following known issues:

i. Space-view anomalies due to electrical interference;
ii. CO2 gas leakage and cell pressure changes;
iii. Changes in atmospheric CO2 concentrations;
iv. Satellite orbital drift and diurnal sampling biases;
v. Short overlap periods between successive instruments.

Raw radiance data from SSU include drifts and jumps in the time series due to these issues (*e.g.*, **Figure 2.18**), which must be accounted for in the data assimilation system. Drifts and jumps of this type are not unique to SSU, and other long-term satellite radiance archives are also affected by issues specific to individual instruments. For example, *Simmons et al.* (2014; their Figure 13) have shown that estimated biases for certain MSU, HIRS, and AMSU-A channels can be of similar orders of magnitude to those for SSU, while trends in atmospheric CO_2 concentrations also cause long-term drifts in estimated biases for HIRS, AIRS, and IASI radiances unless accounted for in the observation operator. Biases in radiances observed by MSU and AMSU-A can be attributed mainly to inaccurate calibration offsets and non-linearity (*Zou et al.*, 2006).

Post-launch inter-satellite calibration (or “homogenization”) efforts by the satellite remote sensing community, such as the WMO GSICS (*Goldberg et al.*, 2011) have substantially reduced inter-satellite differences in some cases, including MSU (*Zou et al.*, 2006), AMSU-A (*Zou and Wang*, 2011), and SSU (*Zou et al.*, 2014). In practice, this type of inter-satellite calibration is usually performed by reanalysis systems internally via bias correction terms applied during the data assimilation step. It is therefore not strictly necessary for satellite data to be homogenized prior to its assimilation in a reanalysis system, although it is beneficial to assimilate data with biases as small as possible.

The use of externally homogenized data has been found to improve some aspects of recent reanalyses. For example, homogenized MSU data (NESDIS SNO corrected calibration coefficients; *Zou et al.*, 2006) assimilated by CFSR, MERRA and MERRA-2 (**Table 2.23**) have been found to improve temporal consistency in bias correction patterns (*Rienecker et al.*, 2011), and may have helped MERRA to produce a more realistic stratospheric temperature response following the eruption of Mount Pinatubo (*Simmons et al.*, 2014). In situations where conventional data are unavailable or insufficient to provide a reference for satellite bias correction, such as SSU in the middle and upper stratosphere, homogenized radiance data may be even more effective in eliminating artificial drifts and jumps in the analysis state. Homogenized satellite radiance time series only represent a relatively small fraction of the satellite data ingested by current reanalysis systems (several of which do not assimilate homogenized data at all); however, the availability of homogenized satellite radiance time series is increasing and these data are likely to become more influential in future reanalysis efforts.

Bias corrections for assimilated satellite data often vary by satellite platform and/or reanalysis system. Although bias corrections are intended to limit the impacts of changing satellite biases within the reanalysis, these impacts may still manifest as spurious trends or discontinuities in the time series of temperature and other reanalysis variables. In older reanalyses that assimilated satellite radiances, such as ERA-40 and JRA-25, bias corrections were often (but not always) based on a fixed regression that spanned the lifetime of the instrument (*Sakamoto and Christy*, 2009; *Onogi et al.*, 2007; *Uppala et al.*, 2005). This approach, which occasionally required the reanalysis to be interrupted for manual retuning of bias correction terms, has been replaced by adaptive (or variational) bias correction schemes in recent reanalysis systems. Adaptive bias corrections for satellite radiances are based on differences between observed radiances and expected radiances calculated from model-generated background states. Some early implementations of adaptive bias corrections, such as that applied to TOVS data in JRA-25, left the reanalysis vulnerable to jumps and drifts inherited from the assimilated radiances (*Sakamoto and Christy*, 2009). These problems are addressed in most recent reanalysis systems by defining observational “anchors” that are regarded as unbiased and are therefore allowed to contribute directly to the background state (*Dee*, 2005). A key example is the use of homogenized radiosonde data to anchor bias corrections for satellite radiances (*e.g., Aulígné et al.*, 2007). Versions of this approach have been implemented in ERA-Interim, ERA5, JRA-55, MERRA, and MERRA-2. GNSS-RO observations are also useful for anchoring bias corrections (*e.g.*, *Poli et al.*, 2010), and are used in this capacity in ERA-Interim, ERA5, JRA-55, and MERRA-2; however, GNSS-RO data are only available after May 2001, and in useful numbers only from 2006. The approach to bias correction taken by CFSR and CFSv2

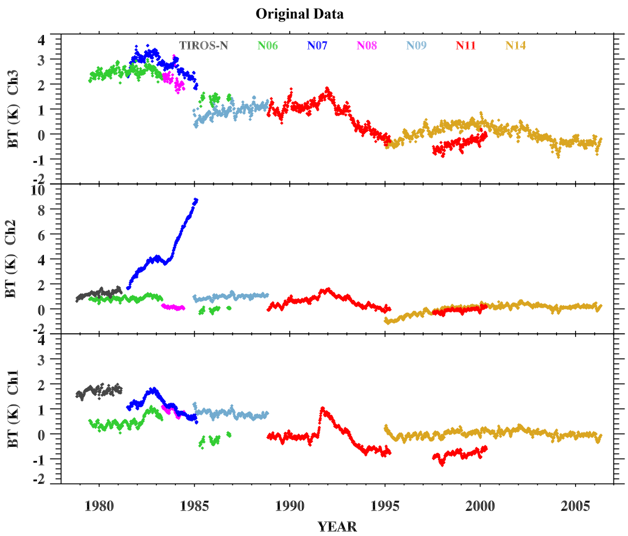


Figure 2.18: *Global mean pentad brightness temperature anomalies based on raw SSU radiances from different satellites. Anomalies are calculated relative to the 1995–2005 mean NOAA-14 annual cycle (from Wang et al., 2012). ©American Meteorological Society. Used with permission.*

(*Saha et al.*, 2010; *Derber and Wu*, 1998) differs from that taken by other systems in that anchor observations are not used. Instead, initial bias corrections are determined for each new satellite instrument via a three-month spin-up assimilation and then allowed to evolve slowly. The effects of satellite-specific drifts and jumps are kept small by assigning very low weights to the most recent biases between the observed and expected radiances, and by accounting for known historical variations in satellite performance as catalogued by multiple research centres. One byproduct of this procedure is an oscillating warm bias in CFSR in the upper stratosphere (see *Chapter 3* of this report). This bias, which is intrinsic to the forecast model, largely disappears when a new execution stream is introduced, only to slowly return as the model bias is imprinted on the observational bias correction terms.

A further example of the type of temporal discontinuities that can result from changes in satellite instrumentation is the cold bias (~2 K) in middle stratospheric temperature in JRA-25 between 1979 and 1998 (*Onogi et al.*, 2007). This feature resulted from a known cold bias in the radiative transfer model used by JRA-25. The SSU had only three channels sensitive to stratospheric temperature (too few to correct the model bias). The AMSU-A instruments, first launched in 1998, have more channels (*i.e.*, higher vertical resolution) in the stratosphere (see also **Figure 2.16**). Assimilation of the higher-resolution AMSU-A radiances effectively corrected the model bias. The JRA-55 system uses an improved radiative transfer model, and produces more realistic stratospheric temperatures during 1979 – 1998 (*Kobayashi et al.*, 2015; *Ebita et al.*, 2011).

A final illustrative example concerns temperatures in the upper stratosphere. MERRA shows artificial annual cycles in the upper stratosphere (*Rienecker et al.*, 2011; their Figure 16), which probably arise because the forward radiative transfer model used to assimilate SSU radiances did not consider variations in atmospheric CO_2. These issues have been corrected in MERRA-2, which uses version 2.1.3 of the CRTM to assimilate SSU radiances (**Table 2.19**). Several reanalyses also show jumps in upper stratospheric temperature in or around 1998 (the sign varies by vertical level and reanalysis) due to the introduction of AMSU-A, which includes channels that peak higher in the stratosphere. See *Chapter 3* of this report for further details and additional examples.

2.4.3.3 Aircraft data

Measurements made by aircraft, such as the AMDAR data collection, are influential inputs in many atmospheric analyses and reanalyses (*Petersen*, 2016). Horizontal wind data from aircraft are assimilated in all of the reanalysis systems but ERA-20C and 20CR, while temperature data from aircraft are assimilated in all of the reanalysis systems except for ERA-20C, JRA-55, JRA-25, and 20CR. In principle, aircraft data were assimilated from the outset by ERA- 40 (September 1957; *Uppala et al.*, 2005), JRA-55 (January 1958; *Kobayashi et al.*, 2015), and NCEP-NCAR R1 (January 1958; *Kalnay et al.*, 1996; see also *Moninger et al.*, 2003), although many of the data from these early years do not meet the necessary standards for assimilation. The volume of aircraft data suitable for assimilation increased substantially after January 1973 (*Uppala et al.*, 2005; *Kobayashi et al.*, 2015).

Aircraft temperature data have been reported to have a warm bias with respect to radiosonde observations (*Ballish and Kumar*, 2008). This type of discrepancy among ingested data sources can have important impacts on the analysis. For example, *Rienecker et al.* (2011) and *Simmons et al.* (2014) have shown that an increase in the magnitude of the temperature bias at 300 hPa in MERRA with respect to radiosondes in the middle to late 1990s coincides with a large increase in the number of aircraft observations assimilated by the system. Moreover, they conclude that differences in temperature trends at 200 hPa between MERRA and ERA-Interim reflect the different impacts of aircraft temperatures in these two reanalysis systems. MERRA-2 applies adaptive bias corrections to AMDAR observations that may help to reduce the uncertainties associated with assimilating these data (*McCarty et al.*, 2016): after each analysis step the updated bias is estimated as a weighted running mean of the aircraft observation increments from preceding analysis times. These adaptive bias corrections are calculated and applied for each aircraft tail number in the database separately.

2.4.4 Water vapour

The assimilation of radiosonde and satellite observations of humidity fields is problematic in the upper troposphere and above, where water vapour mixing ratios are very low and measurement uncertainties are relatively large. The impact of saturation means that humidity probability density functions are often highly non-Gaussian (*Ingleby et al.*, 2013). These issues are particularly pronounced near the tropopause, where sharp temperature gradients complicate the calculation and application of bias corrections for humidity variables during the assimilation step. Reanalysis systems therefore often do not assimilate observations of water vapour provided by radiosondes and/or microwave and infrared sounders (mostly in the form of radiances; see *Section 2.4.2.2*) above a specified upper bound, which is typically between ~300 hPa and ~100 hPa. In regions of the atmosphere that lie above this upper bound (*i.e.,* the uppermost troposphere and stratosphere), the water vapour field is typically determined by the forecast model alone. In this case, water vapour in the stratosphere is determined mainly by transport from below, turbulent mixing, and dehydration in the vicinity of the tropical cold point tropopause (*e.g.*, *Gettelman et al.*, 2010). **Table 2.24** provides brief descriptions of special treatments and caveats affecting reanalysis estimates of water vapour in the upper troposphere and stratosphere. A more detailed discussion and assessment of reanalysis estimates of water vapour is provided in *Chapter 4* of this report.

2.5 Execution streams

2.5.1 What is an 'execution stream'?

The production of reanalyses often must be completed under strict deadlines determined by external factors. To meet these deadlines, most reanalyses have been executed in two or more distinct 'streams', which are then combined. Discontinuities in the time series of some analyzed variables may occur when streams are joined. These potential discontinuities should be considered (along with the changes in assimilated observations described in *Section 2.4*) when reanalysis variables are used for assessments of climate variability and/or trends.

2.5.2 Summary of stream execution

Table 2.25 and **Figure 2.19** briefly summarize the streams used for generating each set of reanalysis products. Refer to the reference papers listed in **Table 2.1** for the procedures used to transition between streams in creating the final data product, as different reanalysis systems may use different approaches. Certain periods have been reprocessed to correct errors in the input data. The reprocessed periods and associated potential discontinuities listed in **Table 2.25** and shown in F**igure 2.19** may be incomplete, and are also likely to change subsequent to the publication of this report. Users are therefore recommended to contact the reanalysis centres directly if they encounter unexplained shifts or jumps in reanalysis products.

Table 2.24: *Notes on treatment of water vapour in the upper troposphere and stratosphere. Additional information is provided in Chapter 2E.*

Reanalysis System	Special treatments and caveats affecting reanalysis estimates of water vapour
ERA-40	No adjustments due to data assimilation are applied in the stratosphere (above the diagnosed tropopause. Methane oxidation is included via a simple parameterization in the stratosphere.
ERA-Interim	The ERA-Interim system contains a parameterization that allows supersaturation with respect to ice in the cloud-free portions of grid cells with temperatures less than 250 K. As in ERA-40, no adjustments due to data assimilation are applied in the stratosphere, and methane oxidation is included via a simple parameterization.
ERA-20C	ERA-20C does not assimilate any water vapour observations. Supersaturation with respect to ice is permitted in cloud-free portions of grid cells with temperatures less than 250 K, and methane oxidation is included via a simple parameterization in the stratosphere.
ERA5	Similar to ERA-Interim, but the parameterization of supersaturation with respect to ice in cloud-free portions of grid cells has been extended to all temperatures less than 273 K (as opposed to only temperatures less than 250 K as in ERA-Interim) and a more consistent treatment of potentially negative values in the stratosphere has been added.
JRA-25 / JCDAS	Observations of humidity are not assimilated and analyses of moisture variables are not provided at pressures less than 100 hPa. Vertical correlations of humidity background errors are set to zero at pressures less than 50 hPa to prevent spurious analysis increments above this level. No moisture source due to methane oxidation is applied to water vapour in the stratosphere. The radiation scheme assumes a constant volume mixing ratio of 2.5 ppmv in the stratosphere.
JRA-55	Analyses of moisture variables are not provided at pressures less than 100 hPa in the pressure-level analysis (anl_p), although analyses of moisture variables are provided for all model levels in the model-level analysis (anl_mdl). Observations of humidity are not assimilated at pressures less than 100 hPa, and vertical correlations of humidity background errors are set to zero at pressures less than 5 hPa to prevent spurious analysis increments above this level. No moisture source due to methane oxidation is applied to water vapour in the stratosphere. The radiation scheme uses climatological annual mean mixing ratios observed by HALOE and UARS MLS during 1991 – 1997 (without seasonal variations) in the stratosphere.
MERRA	The MERRA system tightly constrains stratospheric water vapour to a specified profile, which is based on zonal mean climatologies from HALOE and Aura MLS (*Rienecker et al.*, 2011; *Jiang et al.*, 2010). Water vapour does not undergo physically meaningful variations at pressures less than ~50 hPa.
MERRA-2	Essentially the same as MERRA.
NCEP-NCAR R1	Analyses of moisture variables are not provided at pressures less than 300 hPa. Satellite humidity retrievals are not assimilated.
NCEP-DOE R2	Satellite humidity retrievals are not assimilated.
CFSR / CFSv2	Although there is no upper limit to assimilated GNSS-RO data, radiosonde humidities are only assimilated at pressures 250 hPa and greater. Moisture variables are provided in the stratosphere, but dehydration processes in the tropopause layer may yield negative values, which are replaced by very small positive values for the radiation calculations, but are not replaced in the analysis. Methane oxidation is not included.
NOAA-CIRES 20CR v2	Moisture variables are provided in the stratosphere, but dehydration processes in the tropopause layer may yield negative values, which are artificially replaced by very small positive values for the radiation calculations, but are not replaced in the output fields. Methane oxidation is not included.

Table 2.25: *Information on the execution streams for each reanalysis system.*

Reanalysis System	Execution sreams
ERA-40	ERA-40 was planned for execution in three streams covering 1989–2002, 1957–1972, and 1972–1988. In practice, a small number of parallel-running sub-streams bridging gaps between the main streams had to be run in order to meet the production deadline.
ERA-Interim	ERA-Interim was carried out in two main streams, the first from 1989 to present and the second from 1979 to 1988. The period of the first stream covering January 1989 to August 1993 was rerun to include from the outset all changes made on the fly in the original production for this period; these changes were also included in the second main production stream. The second stream was extended to the end of 1989 to check consistency during the overlap period (see also discussion by *Simmons et al.*, 2014).
ERA-20C	The reanalysis consists of 22 streams, all but the last of which are six years in length. The first stream starts on 1 January 1899 and extends through 31 December 1904. Each subsequent stream starts on 1 January in years ending in 4 or 9 and ends on 31 December of the next year ending in 4 or 9. The final stream starts on 1 January 2004 and extends seven years through the end of the reanalysis. The first year of each stream is discarded from the final product.
ERA5	ERA5 comprises one high-resolution (31-km) analysis (HRES) and a 10-member reduced-resolution (62-km) ensemble of data assimilations (EDA). Seven production streams were run between 1979 and the present for the EDA, and additional shorter streams were run for the HRES to resolve, where practicable, issues encountered in the original production streams. Details are given in Table 3 of *Hersbach et al.* (2020). A further four streams have been run to provide analyses from 1950 to 1978. In addition to these streams for ERA5 core production, a rerun covering the period 2000–2006 has been conducted and is now publicly available under the name ERA5.1. This rerun offers improved representations of temperature and humidity in the stratosphere but differs little from ERA5 in the lower and middle troposphere.
JRA-25 / JCDAS	JRA-25 was conducted in two main streams: the first covers January 1979–December 1990, and the second covers January 1991–January 2014. Note also the transition from JRA-25 (conducted jointly by JMA and CRIEPI) to JCDAS (conducted by JMA only) in January 2005. The execution of JCDAS was conducted entirely in real time. Two periods (January 1994–December 1999 and January 2000–January 2002) were recalculated and replaced to fix problems with data quality; these two periods may be considered as separate sub-streams in addition to the two main streams.
JRA-55	JRA-55 has been executed in two streams. Stream A covers January 1958 through August 1980, while stream B covers September 1980 through the present. Three periods have also been reprocessed after errors were identified: January to June 1958, December 1974 to August 1980 and June 1987 to September 1992 (see also *Kobayashi et al.*, 2015; their Figure 7). JRA-55C has been executed in three streams: Stream A covers 1 November 1972 through 31 August 1980, Stream B covers 1 September 1980 through 31 August 2005, and Stream C covers 1 September 2005 through 31 December 2012. JRA-55AMIP has been executed in one continuous stream.
MERRA	MERRA was executed in three streams. Stream 1 covers January 1979–December 1992, stream 2 covers January 1993–December 2000, and stream 3 covers January 2001–present. Each stream was spun up in two stages: a 2-year analysis at $2^{\circ}\times2.5^{\circ}$ followed by a 1-year analysis on the native MERRA grid (see **Table 2.2**). The production version of stream 2 (after spin-up) overlaps with the final four years of stream 1 (January 1989–December 1992), while the production version of stream 3 overlaps with the final three years of stream 2 (January 1998–December 2000).
MERRA-2	MERRA-2 was executed in four streams covering January 1980–December 1991, January 1992–December 2000, January 2001–December 2010, and January 2011–present. Each stream was spun up for one year on the full MERRA-2 system.
NCEP-NCAR R1	NCEP-NCAR R1 was run in three streams. The first stream, which produced data covering 1982–present, was started in December 1978. The second stream, covering 1958–1981 (post-IGY), was started second. For the third and final stream, which covers 1948–1957 (pre-IGY), the analyses were conducted at 03Z, 09Z, 15Z and 21Z (rather than 00Z, 06Z, 12Z and 18Z). There may be additional discontinuities involving updates. For example, the original analyses may have been affected by a problem with the sea ice boundary condition. A second simulation with an improved sea ice boundary condition may be run for a few months, and then replace the original analyses. Transitions between the original product and these "patches" may cause discontinuities.
NCEP-DOE R2	NCEP-DOE R2 was executed in one continuous stream; however, like NCEP-NCAR R1, there may be discontinuities involving updates.
CFSR / CFSv2	CFSR was produced by running six simultaneous streams covering the following periods: • Stream 1: 1 December 1978 to 31 December 1986 • Stream 2: 1 November 1985 to 31 December 1989 • Stream 5: 1 January 1989 to 31 December 1994 • Stream 6: 1 January 1994 to 31 March 1999 • Stream 3: 1 April 1998 to 31 March 2005 • Stream 4: 1 April 2004 to 31 December 2009 A full 1-year overlap between the streams was used to address spinup issues concerning the deep ocean, the upper stratosphere and the deep soil. The entire CFSR thus covers 31 years (1979–2009) plus five overlap years. Each earlier stream is used to its end, so that the switch to the next stream occurs at the end of the overlap period. A separate one-year stream was run for 2010, after which the analysis system was updated to CFSv2 (with an increase in horizontal resolution from T382 to T574). For most applications, CFSR can be extended through the present using output from CFSv2.
NOAA-CIRES 20CR v2	20CR v2 was executed in 28 streams. With some exceptions, each stream typically produced five years of data with 14 months of spinup. The following text gives the data coverage provided by each stream (the streams are numbered sequentially), with the spin-up start year provided in parentheses: 1871–1875 (1869), 1876–1880 (1874), 1881–1885 (1879), 1886–1890 (1884), 1891–1895 (1889), 1896–1900 (1894), 1901–1905 (1899), 1906–1910 (1904), 1911–1915 (1909), 1916–1920 (1914), 1921–1925 (1919), 1926–1930 (1924), 1931–1935 (1929), 1936–1940 (1934), 1941–1945 (1939), 1946–1951 (1944), 1952–1955 (1949), 1956–1960 (1954), 1961–1965 (1959), 1966–1970 (1964), 1971–1975 (1969), 1976–1980 (1974), 1981–1985 (1979), 1986–1990 (1984), 1991–1995 (1989), 1996–2000 (1994), and 2001–2012 (1999). The spin-up start date for each stream was 00 UTC 1 November, the production start date was 00 UTC 1 January, and the production end date was 21 UTC 31 December.

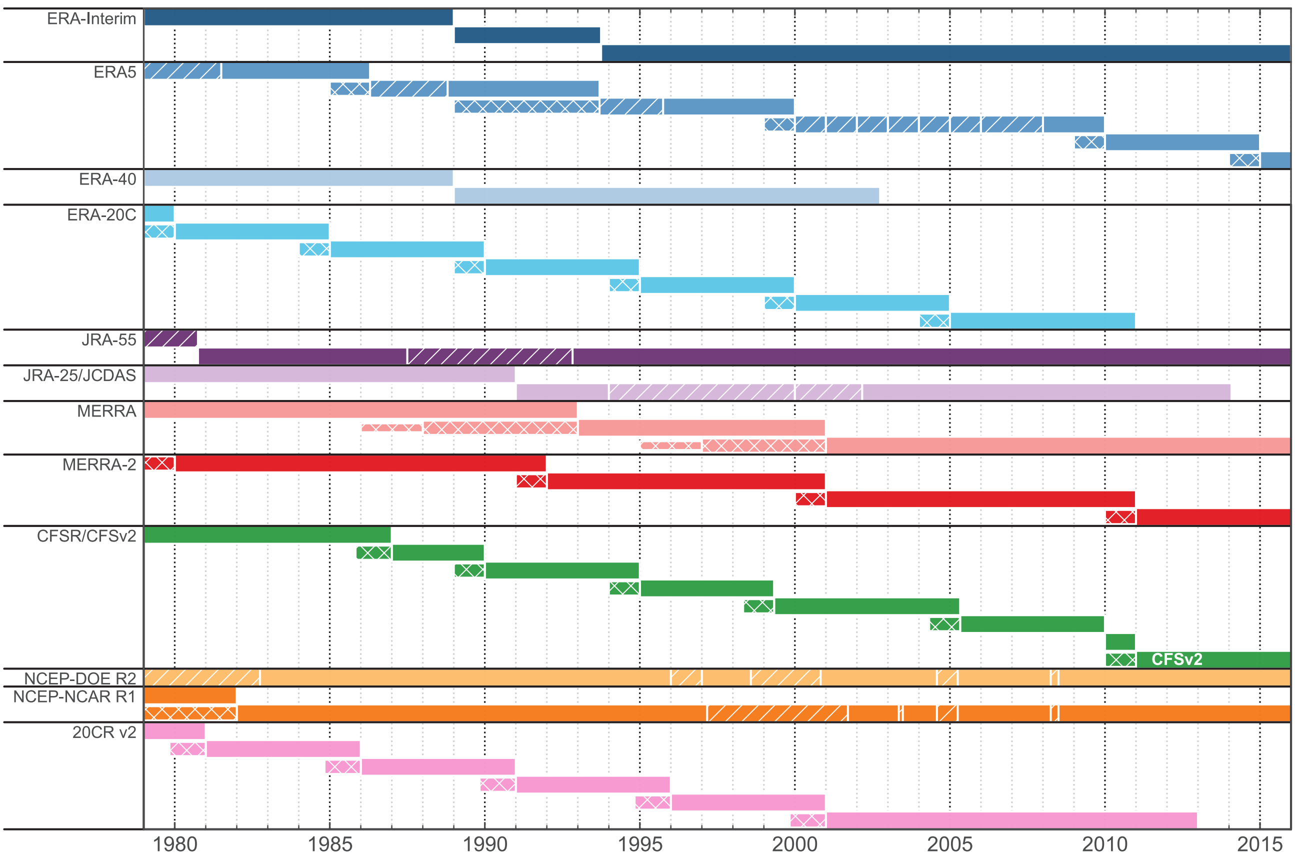


***Figure 2.19:** Summary of the execution streams of the reanalyses for the period 1979–2016. Hatching indicates known reprocessed 'patches' or 'repair runs'. The narrowest cross-hatched segments indicate known spin-up periods, while the medium-narrow cross-hatched segments indicate overlap periods. See also **Table 2.25**. Reproduced from Fujiwara et al. (2017).*

2.6 Archived data

The original data at model resolution and model levels (**Table 2.2**) are converted by each reanalysis centre to data on regular horizontal grids (sometimes at multiple resolutions) and on pressure levels (see *Appendix A*) for public release. The converted data (and sometimes the original data) can typically be obtained via the reanalysis centre websites (see the S-RIP website for links). Some other institutes or projects, such as the NCAR Research Data Archive (RDA), have also constructed public archives of one or more of the reanalysis datasets. Such institutes may have used independent conversions for the data grid, levels, and/or units. Pre-processed data sets have also been produced for the S-RIP activity, including zonal-mean data sets containing dynamical (*Martineau*, 2017) and diabatic (*Wright*, 2017) diagnostics on pressure levels. These pre-processed data are stored in the S-RIP archive at CEDA (**http://data.ceda.ac.uk/badc/srip/**), together with detailed documentation (see also *Martineau et al.*, 2018). Additional data produced for S-RIP include supplementary data files for this chapter (many also provided as a supplement to *Fujiwara et al.*, 2017) and common grid files containing basic variables (*Davis*, 2020). CFSR/CFSv2 products on model levels have also been converted to netCDF format for S-RIP using the High-Resolution Initial Conditions binary files and forecast files archived by NOAA NCEI (**https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/climate-forecast-system-version2-cfsv2**). Data users of these or any other public release of reanalysis or reanalysis-based products should always read the documentation for that release carefully.

It is particularly important to check unit information, as different reanalysis centres or public archives may use different units for the same variable. For example, temperature may be provided in units of °C or K. Some centres provide geopotential height in meters (or "gpm"), while others provide geopotential in $m^2\,s^{-2}$. For water vapour, specific humidity (not volume mixing ratio) is provided in most cases, in units of either $kg\,kg^{-1}$ or $g\,kg^{-1}$. Some reanalyses do not provide vertical pressure velocity (ω, in $Pa\,s^{-1}$) and/or specific humidity data in the stratosphere. Ozone is provided as mass mixing ratio (not volume mixing ratio) in most cases, in units of either $kg\,kg^{-1}$ or $mg\,kg^{-1}$ (*i.e.*, ppmm). Care is also recommended when using precipitation or other 'flux' data, because the integration time period may not be explicitly documented in the data file. Precipitation data may also be divided into multiple categories (such as anvil, convective, and large-scale), the exact definitions of which vary by reanalysis.

Monthly mean products may also differ across different reanalyses, and even for different variables from the same reanalysis, owing to differences in the sampling times or intervals (hourly, 3-hourly, or 6-hourly; instantaneous or time-average). Such differences can be especially impactful for variables with distinct diurnal or sub-diurnal signals (*e.g.*, land-sea breezes in the boundary layer and atmospheric tides in the upper stratosphere and above). In the following, we describe the exact definitions of monthly means for major variable groups in recent reanalyses.

- ERA-Interim and ERA5 divide variables into "instantaneous", "forecast", and "accumulated" products. For ERA-Interim, monthly means of instantaneous products are calculated from 6-hourly data valid at 00, 06, 12, and 18 UTC throughout the month. For ERA5, monthly means of instantaneous products are calculated from hourly fields valid from 00 through 23 UTC throughout the month. Monthly means of accumulated products account for all forecast time steps, although in some cases (*e.g.*, temperature and moisture tendencies) these products are not provided and must be calculated by the user. In cases where ECMWF does provide a monthly mean, partial time steps have been accounted for so that only time steps within the specified month have been included in the average.

- For JRA-55, monthly means of upper-air winds, temperature, geopotential height, and other core analysis fields are calculated from instantaneous analyses at 00, 06, 12, and 18 UTC. Distinctions between instantaneous and time-averaged forecast diagnostics also apply for JRA-55. Instantaneous forecast products are output either every three hours (for two-dimensional fields) or every six hours (for three-dimensional fields), and the monthly means reflect this sampling. Time-averaged fields are designated by the fcst_phy collections, with monthly means representing all time steps.

- For MERRA and MERRA-2, monthly means for each product are calculated by averaging the corresponding instantaneous or time-averaged data. For example, inst3_3d_asm_Np (3-hourly instantaneous data) gets averaged over a month to produce instM_3d_asm_Np and inst6_3d_asm_Np (6-hourly instantaneous data) gets averaged over a month to produce instM_3d_asm_Np. By contrast, the tavg files contain fields averaged from all the (15-min) model time steps within a given time window. Like inst3_3d_asm, these fields are forecast model outputs from the IAU "corrector" step as described in section 2.3 above.

- For CFSR/CFSv2, monthly mean analysis fields are calculated from instantaneous values at 00, 06, 12, and 18 UTC. Monthly means of most forecast variables are also calculated from instantaneous outputs. Only radiation, precipitation, and other 'flux'-type variables are aggregated from averages over the forecast step. These distinctions are directly embedded in the metadata of original GRIB2 files for CFSR/CFSv2 (e.g., 'anl' for analysis variables, '6 hour fcst' for instantaneous forecast variables, and '0-6 hour ave fcst' for time-average forecast variables).

The file formats for archived data may include GRIB, GRIB2, NetCDF, and HDF. Grid boundaries and orientations, such as the starting point for longitude (0°E or 180°W), the order of latitudes (from the North Pole or from the South Pole), and the vertical orientation (from the surface or from the TOA) may also vary by reanalysis and/or data source.

After interpolation to pressure levels, most reanalyses (with the exception of MERRA and MERRA-2) provide data below the surface (*e.g.*, at 1000 hPa over the continents). These data are calculated via vertical extrapolation, and are provided for two reasons. First, they enable the use of a complete field when plotting or taking derivatives, and second, they allow data users to visualize variability over the whole globe (including features over mountains) using data from a single pressure surface. The extrapolation procedure may differ by variable and/or reanalysis system. Users of data in the lower part of the troposphere should be aware of this feature, particularly in regions of complex topography.

Acknowledgements

The materials contained in the tables and figures have been compiled from a variety of sources, for which we acknowledge the contributions of Santha Akella, Peter Bechtold, Michael Bosilovich, Dick Dee, John Derber, Ron Gelaro, Yun-Tai Hou, Robert Kistler, Daryl Kleist, Shinya Kobayashi, Shrinivas Moorthi, Eric Nielsen, Yvan Orsolini, Saroja Polavarapu, Paul Poli, Bill Putman, Suranjana Saha, Jack Woollen, Fanglin Yang, and Valery Yudin. Yoshio Kawatani prepared an early draft of the text on calculation of monthly means. We thank Wenhao Dong, Suqin Duan, and Jacob Smith for providing comments on advance drafts of this document, and an anonymous reviewer for detailed and insightful comments on the first submitted draft. For C.-Z. Zou, the views, opinions, and findings contained in this report are those of the authors and should not be construed as an official NOAA or U.S. Government position, policy, or decision.

References

Allan, P., *et al.*, 2011: The international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative. *Bull. Am. Meteor. Soc.*, **92**, 1421 - 1425, doi: 10.1175/2011BAMS3218.1.

Allen, R.J., and C.S. Zender, 2011: Forcing of the Arctic Oscillation by Eurasian snow cover. *J. Climate*, **24**, 6528 - 6539, doi: 10.1175/2011JCLI4157.1.

Andersson, E. and H. Järvinen, 1999: Variational quality control. *Q. J. R. Meteorol. Soc.*, **125**, 697 - 722, doi: 10.1002/qj.49712555416.

Andrae, U., N. Sokka, and K. Onogi, 2004: The radiosonde temperature bias corrections used in ERA-40. *ERA-40 Project Report Series*, No. **15**, ECMWF, Shinfield Park, Reading, UK, 34 pp. [available at **https://www.ecmwf.int/en/elibrary/7773-radiosonde-temperature-bias-corrections-used-era-40**; accessed August 2018].

Arakawa, A., 2004: The cumulus parameterization problem: past, present, and future. *J. Climate*, **17**, 2493 - 2525, doi: 10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2.

Arakawa, A., and W.H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, part I. *J. Atmos. Sci.*, **31**, 674 - 701, doi: 10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

Auligné, T., A.P. McNally, and D.P. Dee, 2007: Adaptive bias correction for satellite data in a numerical weather prediction system. *Q. J. R. Meteorol. Soc.*, **133**, 631 - 642, doi: 10.1002/qj.56.

Bacmeister, J.T., M.J. Suarez, and F.R. Robertson, 2006: Rain re-evaporation, boundary layer–convection interactions, and Pacific rainfall patterns in an AGCM. *J. Atmos. Sci.*, **63**, 3383 - 3403, doi: 10.1175/JAS3791.1.

Baines, P.G., and T.N. Palmer, 1990: Rationale for a new physically-based parametrization of subgrid-scale orographic effects. *Tech. Memo*. No. **169**, ECMWF, Shinfield Park, Reading, UK, 11 pp. [available at **https://www.ecmwf.int/en/elibrary/7875-rationale-new-physically-based-parametrization-subgrid-scale-orographic-effects**; accessed August 2018].

Ballish, B.A., and V.K. Kumar, 2008: Systematic differences in aircraft and radiosonde temperatures: Implications for NWP and climate studies. *Bull. Am. Meteor. Soc.*, **89**, 1689 - 1708, doi: 10.1175/2008BAMS2332.1.

Balsamo, G., *et al.*, 2015: ERA-Interim/Land: a global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389 - 407, doi: 10.5194/hess-19-389-2015.

Bechtold, P., *et al.*, 2014: Representing equilibrium and nonequilibrium convection in large-scale models. *J. Atmos. Sci.*, **71**, 734 - 753, doi:10.1175/JAS-D-13-0163.1.

Beljaars, A.C.M., 1995: The impact of some aspects of the boundary layer scheme in the ECMWF model. In *Proc. of ECMWF Seminar on Parameterization of Sub-grid Scale Physical Processes*, pp. 125 - 161, Shinfield Park, Reading, UK, 5 - 9 September 1994. [available at **https://www.ecmwf.int/en/elibrary/8035-impact-some-aspects-boundary-layer-scheme-ecmwf-model**; accessed August 2018].

Bengtsson, L., M. Kanamitsu, P. Kållberg, and S. Uppala, 1982: FGGE research activities at ECMWF. *Bull. Am. Meteor. Soc.*, **63**, 277 - 303, doi: 10.1175/1520-0477-63.3.277.

Birner, T., D. Sankey, and T.G. Shepherd, 2006: The tropopause inversion layer in models and analyses. *Geophys. Res. Lett.*, **33**, L14804, doi: 10.1029/2006GL026549.

Bloom, S., L. Takacs, A. DaSilva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256 - 1271, doi: 10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.

Bodeker, G.E., *et al.*, 2016: Reference upper-air observations for climate: From concept to reality. *Bull. Am. Meteor. Soc.*, **97**, 123 - 135, doi: 10.1175/BAMS-D-14-00072.1.

Boussetta, S., *et al.*, 2013: Impact of a satellite-derived leaf area index monthly climatology in a global numerical weather prediction model. *Int. J. Remote Sens.*, **34**, 3520 - 3542, doi: 10.1080/01431161.2012.716543.

Bouttier, F., and P. Courtier, 1999: Data assimilation concepts and methods. *Meteorological Training Course Lecture Series*, ECMWF, Shinfield Park, Reading, UK. [available at **https://www.ecmwf.int/en/elibrary/16928-data-assimilation-concepts-and-methods**; accessed August 2018].

Briegleb, B.P., 1992: Delta-Eddington approximation for solar radiation in the NCAR community climate model. *J. Geophys. Res. Atmos.*, **97**, 7603 - 7612, doi:10.1029/92JD00291.

Buchard, V., *et al.*, 2015: Using the OMI aerosol index and absorption aerosol optical depth to evaluate the NASA MERRA Aerosol Reanalysis. *Atmos. Chem. Phys.*, **15**, 5743 - 5760, doi:10.5194/acp-15-5743-2015.

Buchard, V., *et al.*, 2017: The MERRA-2 Aerosol Reanalysis, 1980–onward, Part II: Evaluation and Case Studies. *J. Climate*, **30**, 6851 - 6872, doi: 10.1175/JCLI-D-16-0613.1.

Buehner, M., *et al.*, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: description and single-observation experiments. *Mon. Wea. Rev.*, **138**, 1550 - 1566, doi: 10.1175/2009MWR3157.1.

Buehner, M., *et al.*, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: one-month experiments with real observations. *Mon. Wea. Rev.*, **138**, 1567 - 1586, doi: 10.1175/2009MWR3158.1.

Campbell, W.F., C.H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Wea. Rev.*, **138**, 282 - 290, doi: 10.1175/2009MWR3017.1.

Cariolle, D. and M. Déqué, 1986: Southern hemisphere medium-scale waves and total ozone disturbances in a spectral general circulation model. *J. Geophys. Res. Atmos.*, **91**, 10825 - 10846, doi: 10.1029/JD091iD10p10825.

Cariolle, D. and H. Teyssèdre, 2007: A revised linear ozone photochemistry parameterization for use in transport and general circulation models: multi-annual simulations. *Atmos. Chem. Phys.*, **7**, 2183 - 2196, doi: 10.5194/acp-7-2183-2007.

Charnock, H., 1955: Wind stress on a water surface. *Q. J. R. Meteorol. Soc.*, **81**, 639 - 640, doi: 10.1002/qj.49708135027.

Chou, M.-D., and K.-T. Lee, 1996: Parameterizations for the absorption of solar radiation by water vapor and ozone. *J. Atmos. Sci.*, **53**, 1203 - 1208, doi: 10.1175/1520-0469(1996)053<1203:PFTAOS>2.0.CO;2.

Chou, M.-D., and M.J. Suarez, 1999: A solar radiation parameterization for atmospheric studies. *NASA Tech. Rep. Series on Global Modeling and Data Assimilation*, Vol. **15**, NASA/TM-1999-104606, Greenbelt, Maryland, 40 pp. [available at **https://ntrs.nasa.gov/search.jsp?R=19990060930**; accessed August 2018].

Chou, M.-D., M.J. Suarez, X.-Z. Liang, and M.M.-H. Yan, 2001: A thermal infrared radiation parameterization for atmospheric studies. *NASA Tech. Rep. Series on Global Modeling and Data Assimilation*, Vol. **19**, NASA/TM-2001-104606, Greenbelt, Maryland, 56 pp. [available at **https://ntrs.nasa.gov/search.jsp?R=20010072848**; accessed August 2018].

Chun, H.-Y, and J.-J. Baik, 1998: Momentum flux by thermally induced internal gravity waves and its approximation for large-scale models. *J. Atmos. Sci.*, **55**, 3299 - 3310, doi: 10.1175/1520-0469(1998)055<3299:MFBTII>2.0.CO;2.

Cionni, I., *et al.*, 2011: Ozone database in support of CMIP5 simulations: results and corresponding radiative forcing. *Atmos. Chem. Phys.*, **11**, 11267 - 11292, doi: 10.5194/acp-11-11267-2011.

Clough, S.A., *et al.*, 2005: Atmospheric radiative transfer modeling: A summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transfer*, **91**, 233 - 244, doi: 10.1016/j.jqsrt.2004.05.058.

Cohen, J., and D. Entekhabi, 1999: Eurasian snow cover variability and Northern Hemisphere climate variability. *Geophys. Res. Lett.*, **26**, 345 - 348, doi: 10.1029/1998GL900321.

Cohen, J., *et al.*, 2014: Linking Siberian snow cover to precursors of stratospheric variability. *J. Climate*, **27**, 5422 - 5432, doi: 10.1175/JCLI-D-13-00779.1.

Colarco, P.R., A. da Silva, M. Chin, and T. Diehl, 2010: Global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth. *J. Geophys. Res. Atmos.*, **115**, D14207, doi: 10.1029/2009JD012820.

Compo, G.P., *et al.*, 2011: The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.*, **137**, 1 - 28, doi: 10.1002/qj.776.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367 - 1387, doi: 10.1002/qj.49712051912.

Cram, T.A., *et al.*, 2015: The International Surface Pressure Databank version 2. *Geosci. Data J.*, **2**, 31 - 46, doi: 10.1002/gdj3.25.

Cressman, G.P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367 - 374, doi: 10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2.

Daley, R., 1981: Normal mode initialization. *Rev. Geophys.*, **19**, 450 - 468, doi: 10.1029/RG019i003p00450.

Davis, S., 2020: SPARC Reanalysis Intercomparison Project (S-RIP) reanalysis common grid files. doi: 10.5281/zenodo.3754752.

de Rosnay, P., *et al.*, 2014: Initialisation of land surface variables for numerical weather prediction. *Surv. Geophys.*, **35**, 607 - 621, doi: 10.1007/s10712-012-9207-x.

de Rosnay, P., L. Isaksen, and L. Dahoui, 2015: Snow data assimilation at ECMWF. *ECMWF Newsletter* No. **143**, Spring 2015. [available at **https://www.ecmwf.int/en/elibrary/17328-snow-data-assimilation-ecmwf**; accessed September 2018].

Dee, D.P., 2005: Bias and data assimilation. *Q. J. R. Meteorol. Soc.*, **131**, 3323 - 3343, doi: 10.1256/qj.05.137.

Dee, D.P., *et al.*, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, **137**, 553 - 597, doi: 10.1002/qj.828.

Derber, J.C., and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287 - 2299, doi: 10.1175/1520-0493(1998)126<2287:TUOTCC>2.0.CO;2.

Dethof, A., and E. V. Hólm, 2004: Ozone assimilation in the ERA-40 reanalysis project. *Q. J. R. Meteorol. Soc.*, **130**, 2851 - 2872, doi: 10.1256/qj.03.196.

Donlon, C., *et al.*, 2007: The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bull. Am. Meteor. Soc.*, **88**, 1197 - 1213, doi: 10.1175/BAMS-88-8-1197.

Dorman, J.L., and P.J. Sellers, 1989: A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the Simple Biosphere model (SiB). *J. Appl. Meteorol.*, **28**, 833 - 855, doi: 10.1175/1520-0450(1989)028<0833:AGCOAR>2.0.CO;2.

Dragani, R., 2011: On the quality of the ERA-Interim ozone reanalyses: Comparisons with satellite data. *Q. J. R. Meteorol. Soc.*, **137**, 1312 - 1326, doi: 10.1002/qj.821.

Dutra, E., et al., 2010: An improved snow scheme for the ECMWF land surface model: Description and offline validation. *J. Hydrometeorol.*, **11**, 899 - 916, doi: 10.1175/2010JHM1249.1.

Ebita, A., *et al.*, 2011: The Japanese 55-year reanalysis "JRA-55": An interim report. *SOLA*, **7**, 149 - 152, doi: 10.2151/sola.2011-038.

Ek, M.B., *et al.*, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res. Atmos.*, **108**, 8851, doi: 10.1029/2002JD003296.

Evensen, G., and P.J. van Leeuwen, 2000: An Ensemble Kalman Smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852 - 1867, doi: 10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Evensen, G., 2009: Data Assimilation: The Ensemble Kalman Filter. *Springer*, Berlin, 307 pp., doi: 10.1007/978-3-642-03711-5.

Fels, S., and M.D. Schwarzkopf, 1975: The simplified exchange approximation: A new method for radiative transfer calculations. *J. Atmos. Sci.*, **32**, 1475 - 1488, doi: 10.1175/1520-0469(1975)032<1475:TSEAAN>2.0.CO;2.

Flannaghan, T.J., and S. Fueglistaler, 2014: Vertical mixing and the temperature and wind structure of the tropical tropopause layer. *J. Atmos. Sci.*, **71**, 1609 - 1622, doi: 10.1175/JAS-D-13-0321.1.

Flemming, J., *et al.*, 2017: The CAMS interim Reanalysis of carbon monoxide, ozone and aerosol for 2003 - 2015, *Atmos. Chem. Phys.*, **17**, 1945 - 1983, doi: 10.5194/acp-17-1945-2017.

Fortuin, J.P., and U. Langematz, 1995: An update on the current ozone climatology and on concurrent ozone and temperature trends. *Proc. SPIE*, **2311**, 207 - 216, doi: 10.1117/12.198578.

Fouquart, Y., and B. Bonnel, 1980: Computation of solar heating of the Earth's atmosphere: a new parameterization. *Beitr. Phys.*, **53**, 35 - 62.

Freidenreich, S.M., and V. Ramaswamy, 1999: A new multiple-band solar radiative parameterization for general circulation models. *J. Geophys. Res. Atmos.*, **104**, 31389 - 31409, doi: 10.1029/1999JD900456.

Fujiwara, M., *et al.*, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp-17-1417-2017.

Gandin, L.S., 1963: Objective analysis of meteorological fields. *Gidrometerologischeskoe Izdatelstvo*, Leningrad. [English translation by Israeli Program for Scientific Translations, Jerusalem, 1965].

Garcia, R.R. and B.A. Boville, 1994: Downward control of the mean meridional circulation and temperature distribution of the polar winter stratosphere. *J. Atmos. Sci.*, **51**, 2238 - 2245, doi: 10.1175/1520-0469(1994)051<2238:COTMMC>2.0.CO;2.

Gelaro, R., *et al.*, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate*, **30**, 5419 - 5454, doi: 10.1175/JCLI-D-16-0758.1.

Gettelman, A., *et al.*, 2010: Multimodel assessment of the upper troposphere and lower stratosphere: Tropics and global trends. *J. Geophys. Res. Atmos.*, **115**, D00M08, doi: 10.1029/2009JD013638.

Gibson, J.K., *et al.*, 1997: ERA description, *ERA-15 Rep. Series*, No. **1**, ECMWF, Shinfield Park, Reading, UK, 72 pp. [available at **https://www.ecmwf.int/en/elibrary/9584-era-description**; accessed August 2018].

Goldberg, M., *et al.*, 2011: The Global Space-based Inter-Calibration System. *Bull. Am. Meteor. Soc.*, **92**, 467 - 475, doi: 10.1175/2010BAMS2967.1.

Goody, R.M., 1952: A statistical model for water-vapour absorption. *Quart. J. R. Meteor. Soc.*, **78**, 165 - 169, doi: 10.1002/qj.49707833604.

Gustafsson, N., 2007: Discussion on '4D-Var or EnKF?', *Tellus A*, **59**, 774 - 777, doi: 10.1111/j.1600-0870.2007.00262.x.

Haimberger, L., C. Tavolato, and S. Sperka, 2008: Towards elimination of the warm bias in historic radiosonde temperature records - some new results from a comprehensive intercomparison of upper air data. *J. Climate*, **21**, 4587 - 4606, doi: 10.1175/2008JCLI1929.1.

Haimberger, L., C. Tavolato, and S. Sperka, 2012: Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighboring stations. *J. Climate*, **25**, 8108 - 8131, doi: 10.1175/JCLI-D-11-00668.1.

Harada, Y., *et al.*, 2016: The JRA-55 Reanalysis: representation of atmospheric circulation and climate variability. *J. Meteor. Soc. Jpn.*, **94**, 269 - 302, doi:1 0.2151/jmsj.2016-015.

Helfand, H.M., *et al.*, 1987: The effect of a gravity wave drag parameterization scheme on GLA fourth-order GCM forecasts. Collection of papers presented at the *WMO/IUGG NWP Symposium*, Tokyo, 4 - 8 August 1986, 729 - 742. [available at **https://ntrs.nasa.gov/search.jsp?R=19880050777**; accessed August 2018].

Helfand, H.M. and S.D. Schubert, 1995: Climatology of the simulated Great Plains low-level jet and its contribution to the continental moisture budget of the United States. *J. Climate*, **8**, 784 - 806, doi: 10.1175/1520-0442(1995)008<0784:COTSGP>2.0.CO;2.

Hering, W.S., and T.R. Borden, Jr., 1965: Mean distributions of ozone density over North America, 1963–1964. T*ech. Rep. AFCRL-65-913*, Air Force Cambridge Research Laboratories.

Hersbach, H., *et al.*, 2015: ERA-20CM: a twentieth-century atmospheric model ensemble. *Q. J. R. Meteorol. Soc.*, **141**, 2350 - 2375, doi: 10.1002/qj.2528.

Hersbach, H., 2020: The ERA5 Global Reanalysis, *Q. J. R. Meteorol. Soc.*, **146**, 1999 - 2049, doi: 10.1002/qj.3803.

Holton, J.R., and W.M. Wehrbein, 1980: A numerical model of the zonal mean circulation of the middle atmosphere. *Pure Appl. Geophys.*, **118**, 284 - 306, doi: 10.1007/BF01596455.

Hong, S.-Y. and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322 - 2339, doi: 10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2.

Hortal, M., and A.J. Simmons, 1991: Use of reduced Gaussian grids in spectral models. *Mon. Wea. Rev.*, **119**, 1057 - 1074, doi: 10.1175/1520-0493(1991)119<1057:UORGGI>2.0.CO;2.

Iacono, M.J., 2008: Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. *J. Geophys. Res. Atmos.*, **113**, D13103, doi:10.1029/2008JD009944.

Ingleby, N.B., 2013: Improved variational analyses using a nonlinear humidity control variable. *Q. J. R. Meteorol. Soc.*, **139**, 1875 - 1887, doi: 10/1002/qj.2073.

IPCC, 1996: Climate Change 1995: The Science of Climate Change. Contribution of Working Group I to the *Second Assessment Report of the Intergovernmental Panel on Climate Change*, Houghton, J.T., L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell (eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 572 pp.

Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the KOBE collection. *Int. J. Climatol.*, **25**, 865 - 879, doi: 10.1002/joc.1169.

Iwasaki, T., S. Yamada, and K. Tada, 1989a: A parameterization scheme of orographic gravity wave drag with the different vertical partitionings, Part I: Impact on medium range forecasts. *J. Meteor. Soc. Jpn.*, **67**, 11 - 27, doi: 10.2151/jmsj1965.67.1_11.

Iwasaki, T., S. Yamada, and K. Tada, 1989b: A parameterization scheme of orographic gravity wave drag with the different vertical partitionings, Part II: Zonally averaged budget analyses based on transformed Eulerian-mean method. *J. Meteor. Soc. Jpn.*, **67**, 29 - 41, doi: 10.2151/jmsj1965.67.1_29.

Janssen, P.A.E.M., *et al.*, 2001: Impact and feedback of ocean waves on the atmosphere. *Tech. Memo.* No. **341**, ECMWF, Shinfield Park, Reading, UK, 34 pp. [available at **https://www.ecmwf.int/en/elibrary/10225-impact-and-feedback-ocean-waves-atmosphere**; accessed August 2018].

Jiang, J. H., *et al.*, 2010: Five year (2004–2009) observations of upper tropospheric water vapor and cloud ice from MLS and comparisons with GEOS-5 analyses. *J. Geophys. Res. Atmos.*, **115**, D15103, doi: 10.1029/2009JD013256.

JMA, 2007: Outline of the operational numerical weather prediction at the Japan Meteorological Agency. *WMO Technical Progress Report on the Global Data-processing and Forecasting System (GDPFS) and Numerical Weather Prediction (NWP) Research*, JMA, Japan. [available at **http://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2007-nwp/index.htm**; accessed August 2018].

JMA, 2013: Outline of the operational numerical weather prediction at the Japan Meteorological Agency. *WMO Technical Progress Report on the Global Data-processing and Forecasting System (GDPFS) and Numerical Weather Prediction (NWP) Research*, JMA, Japan. [available at **http://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2013-nwp/index.htm**; accessed August 2018].

Kållberg, P., A. Simmons, S. Uppala, and M. Fuentes, 2007: The ERA-40 Archive. *ECMWF ERA-40 Project Report Series*, No. **17**, 32 pp. (originally published in September 2004, revised in October 2007).

Kalman, R.E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35 - 45, doi: 10.1115/1.3662552.

Kalnay, E., *et al.*, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteor. Soc.*, **77**(3), 437 - 471, doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kalnay, E., 2003: Atmospheric Modeling, Data Assimilation, and Predictability. *Cambridge University Press*, Cambridge, 341 pp.

Kalnay, E., *et al.*, 2007a: 4-D-Var or ensemble Kalman filter?, *Tellus A*, **59**, 758 - 773, doi: 10.1111/j.1600-0870.2007.00261.x.

Kalnay, E., *et al.*, 2007b: Response to the discussion on “4-D-Var or EnKF?” by Nils Gustafsson, *Tellus A*, **59**, 778 - 780, doi: 10.1111/j.1600-0870.2007.00263.x.

Kanamitsu, M., *et al.*, 2002: NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteor. Soc.*, **83**(11), 1631 - 1643, doi: 10.1175/BAMS-83-11-1631.

Kawai, H. and T. Inoue, 2006: A simple parameterization scheme for subtropical marine stratocumulus. *SOLA*, **2**, 17 - 20, doi: 10.2151/sola.2006-005.

Kim, Y.-J., and A. Arakawa, 1995: Improvement of orographic gravity wave parameterization using a mesoscale gravity wave model. *J. Atmos. Sci.*, **83**, 1875 - 1902, doi: 10.1175/1520-0469(1995)052<1875:IOOGWP>2.0.CO;2.

Kistler, R., *et al.*, 2001: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Am. Meteor. Soc.*, **82**(2), 247 - 267, doi: 10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2.

Kleist, D.T., *et al.*, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691 - 1705, doi: 10.1175/2009WAF2222201.1.

Kobayashi, C., *et al.*, 2014: Preliminary results of the JRA-55C, an atmospheric reanalysis assimilating conventional observations only. *SOLA*, **10**, 78 - 82, doi: 10.2151/sola.2014-016.

Kobayashi, S., *et al.*, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteorol. Soc. Jpn.*, **93**, 5 - 48, doi: 10.2151/jmsj.2015-001.

Koepke, P., M. Hess, I. Schult, and E.P. Shettle, 1997: Global Aerosol Data Set. *Report* No. **243**, Max-Planck-Institut für Meteorologie, Hamburg, ISSN 0937-1060, 44 pp. [available at **http://opac.userweb.mwn.de/radaer/gads.html**; accessed September 2018].

Köhler, M., M. Ahlgrimm, and A. Beljaars, 2011: Unified treatment of dry convective and stratocumulus-topped boundary layers in the ECMWF model. *Q. J. R. Meteorol. Soc.*, **137**, 43 - 57, doi: 10.1002/qj.713.

Komen, G.J., *et al.*, 1994: Dynamics and Modelling of Ocean Waves. *Cambridge University Press*, Cambridge, 532 pp.

Kondo, J., 1975: Air–sea bulk transfer coefficients in diabatic conditions. *Bound.-Lay. Meteorol.*, **9**, 91 - 112, doi: 10.1007/BF00232256.

Koren, V., *et al.*, 1999: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *J. Geophys. Res. Atmos.*, **104**, 19,569 - 19,585, doi: 10.1029/1999JD900232.

Koster, R.D., *et al.*, 2000: A catchment-based approach to modeling land surface processes in a general circulation model. Part 1: Model structure. *J. Geophys. Res. Atmos.*, **105**, 24,809 - 24,822, doi: 10.1029/2000JD900327.

Krishnamurti, T.N., and L. Bounoua, 1996: An Introduction to Numerical Weather Prediction Techniques. *CRC Press*, Florida, USA, 293 pp.

Lacis, A.A., and J.E. Hansen, 1974: A parameterization for the absorption of solar radiation in the Earth’s atmosphere. *J. Atmos. Sci.*, **31**, 118 - 133, doi: 10.1175/1520-0469(1974)031<0118:APFTAO>2.0.CO;2.

Laloyaux, P., *et al.*, 2016: A coupled data assimilation system for climate reanalyses. *Q. J. R. Meteorol. Soc.*, **142**, 65 - 78, doi: 10.1002/qj.2629.

Lamarque, J.-F., *et al.*, 2010: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmos. Chem. Phys.*, *10*, 7017 - 7039, doi: 10.5194/acp-10-7017-2010.

Large, W.G., and S. Pond, 1981: Open ocean momentum flux measurements in moderate to strong winds,.*J. Phys. Oceanogr.*, **11**, 324 - 336, doi: 10.1175/1520-0485(1981)011<0324:OOMFMI>2.0.CO;2.

Lawless, A.S., 2010: A note on the analysis error associated with 3D-FGAT. *Q. J. R. Meteorol. Soc.*, **136**, 1094 - 1098, doi: 10.1002/qj.619.

Lawrence, Z.D., *et al.*, 2015: Comparisons of polar processing diagnostics from 34 years of the ERA-Interim and MERRA reanalyses. *Atmos. Chem. Phys.*, **15**, 3873 - 3892, doi: 10.5194/acp-15-3873-2015.

Lei, L., J.S. Whitaker, and C. Bishop, 2018: Improving assimilation of radiance observations by implementing model space localization in an ensemble Kalman filter. J. *Adv. Model. Earth Syst.*, **10**, 3221 - 3232, doi: 10.1029/2018MS001468.

Lin, S.-J., 2004: A “vertically Lagrangian” finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293 - 2307, doi: 10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2.

Lindzen, R.S., E.S. Batten, and J.W. Kim, 1968: Oscillations in atmospheres with tops. *Mon. Wea. Rev.*, **96**, 133 - 140, doi: 10.1175/1520-0493(1968)096<0133:OIAWT>2.0.CO;2.

Liu, Q., *et al.*, 1999: Improvements in hurricane initialization and forecasting at NCEP with global and regional (GFDL) models. *NOAA Tech. Proc. Bull.*, No. **472**, National Weather Service, Office of Meteorology, Silver Spring, MD, 1 - 7. [available at **http://www.nws.noaa.gov/om/tpb/472.htm**; accessed August 2017].

Lock, A.P., *et al.*, 2000: A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Wea. Rev.*, **128**, 3187- 3199, doi: 10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2.

London, J., 1962: Mesospheric dynamics, part III. The distribution of total ozone in the Northern Hemisphere. Final Report, Dept. of Meteorology and Oceanography, New York University, 99 pp.

Lorenc, A.C. and Rawlins, F., 2005: Why does 4D-Var beat 3D-Var? *Q. J. R. Meteorol. Soc.*, **131**, 3247 - 3257, doi: 10.1256/qj.05.85.

Lott, F., and M.J. Miller, 1997: A new subgrid-scale orographic drag parameterization: Its formulation and testing. *Q. J. R. Meteorol. Soc.*, **123**, 101 - 127, doi: 10.1002/qj.49712353704.

Louis, J.F., 1979: Parameteric model of vertical eddy fluxes in the atmosphere. *Bound.-Lay. Meteorol.*, **17**, 187 - 202, doi: 10.1007/BF00117978.

Louis, J.F., M. Tiedtke, and J.-F. Geleyn, 1982: A short history of the operational PBL parameterization at ECMWF. In *Proc. ECMWF Workshop on Boundary Layer Parameterization*, pp. 59 - 80, Reading, 25 - 27 November 1981 [available at **https://www.ecmwf.int/en/elibrary/10845-short-history-pbl-parameterization-ecmwf**; accessed 9 August 2017].

Lynch, P. and X. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.*, **120**, 1019 - 1034, doi: 10.1175/1520-0493(1992)120<1019:IOTHMU>2.0.CO;2.

Lynch, P., 1993: Digital filters for numerical weather prediction. *HIRLAM Technical Reports*, No. **10**, 55 pp. [available at **http://hirlam.org/index.php/publications-54/hirlam-technical-reports-a/doc_download/1563-hirlam-technical-report-no-10-digital-filters-for-numerical-weather-prediction**; accessed 7 October 2017].

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model with applications to normal mode initialization. *Contrib. Atmos. Phys.*, **50**, 253 - 271.

Machenhauer, B., 1979: The Spectral Method, in: Numerical Methods used in Atmospheric Models, Vol. II. *GARP Publication Series*, **17**, World Meteorological Organization, Geneva, 121 - 275.

Mahfouf, J.F., *et al.*, 1995: The land surface scheme ISBA within the Météo-France climate model ARPEGE. *J. Climate*, **8**, 2039 - 2057, doi: 10.1175/1520-0442(1995)008<2039:TLSSIW>2.0.CO;2.

Mahrt, L., and H. Pan, 1984: A two-layer model of soil hydrology. *Bound.-Lay. Meteorol.*, **29**, 1 - 20, doi: 10.1007/BF00119116.

Martineau, P., 2017: S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels [data set]. *Centre for Environmental Data Analysis (CEDA)*, doi: 10.5285/b241a7f536a244749662360bd7839312.

Martineau, P., J.S. Wright, N. Zhu, and M. Fujiwara, 2018: Zonal-mean data set of global atmospheric reanalyses on pressure levels. *Earth Syst. Sci. Data,* **10**, 1925 - 1941, doi: 10.5194/essd-10-1925-2018.

Matsumoto, T., M. Ishii, Y. Fukuda, and S. Hirahara, 2006: Sea ice data derived from microwave radiometer for climate monitoring. *Proceedings of the 14th Conference on Satellite Meteorology and Oceanography*, Atlanta, USA, P2.21.

McCarty, W., *et al.*, 2016: MERRA-2 input observations: Summary and assessment. *NASA Tech. Rep. Series on Global Modeling and Data Assimilation*, Vol. **46**, NASA/TM-2016-104606, Greenbelt, Maryland. [available at **https://ntrs.nasa.gov/search.jsp?R=20160014544**; accessed 9 August 2017]

McCormack, J.P., S.D. Eckermann, D.E. Siskind, and T. McGee, 2006: CHEM2D-OPP: A new linearized gas-phase ozone photochemistry parameterization for high-altitude NWP and climate models. *Atmos. Chem. Phys.*, **6**, 4943 - 4972, doi: 10.5194/acp-6-4943-2006.

McFarlane, N.A., 1987: The effect of orographically excited gravity-wave drag on the circulation of the lower stratosphere and troposphere. *J. Atmos. Sci.*, **44**, 1775 - 1800, doi: 10.1175/1520-0469(1987)044<1775:TEOOEG>2.0.CO;2.

Meinshausen, M., *et al.*, 2011: The RCP greenhouse gas concentrations and their extension from 1765 to 2300. *Clim. Change*, **109**, 213 - 241, doi: 10.1007/s10584-011-0156-z.

Mellor, G.L. and T. Yamada, 1974: A hierarchy of turbulence closure models for planetary boundary layers. *J. Atmos. Sci.*, **31**, 1791 - 1806, doi: 10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2.

Miller, M.J., T.N. Palmer, and R. Swinbank, 1989: Parameterization and influence of subgrid-scale orography in general circulation and numerical weather prediction models. *Meteorol. Atmos. Phys.*, **40**, 84 - 109, doi:10.1007/BF01027469.

Mlawer, E.J., *et al.*,1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the long wave. *J. Geophys. Res. Atmos.*, **102**, 16663 - 16682, doi: 10.1029/97JD00237.

Molod, A., 2012: Constraints on the profiles of total water PDF in AGCMs from AIRS and a high-resolution model. *J. Climate*, **25**, 8341 - 8352, doi: 10.1175/JCLI-D-11-00412.1.

Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA-2. *Geosci. Model Dev.*, **8**, 1339 - 1356, doi: 10.5194/gmd-8-1339-2015.

Moninger, W.R., R.D. Mamrosh, and P.M. Pauley, 2003: Automated meteorological reports from commercial aircraft. *Bull. Am. Meteor. Soc.*, **84**, 203 - 216, doi: 10.1175/BAMS-84-2-203.

Moorthi, S., and M.J. Suarez, 1992: Relaxed Arakawa-Schubert. A parameterization of moist convection for general-circulation models. *Mon. Wea. Rev.*, **120**, 978 - 1002, doi: 10.1175/1520-0493(1992)120<0978:RASAPO>2.0.CO;2.

Moorthi, S., H.-L. Pan, and P. Caplan, 2001: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. *Technical Procedures Bulletin* **484**, NOAA NWS, Silver Spring, MD, 14 pp. [available at **https://rda.ucar.edu/datasets/ds093.0/docs/484.pdf**; accessed August 2018].

Moorthi, S., R. Sun, H. Xiao, and C.R. Mechoso, 2010: Low-cloud simulation in the Southeast Pacific in the NCEP GFS: Role of vertical mixing and shallow convection. *NCEP Office Note* **463**, 28 pp. [available at **http://www.emc.ncep.noaa.gov/officenotes/FullTOC.html**; accessed August 2018].

Morcrette, J-J., *et al.*, 2008: Impact of a new radiation package, McRad, in the ECMWF Integrated Forecasting System. *Mon. Wea. Rev.*, **136**, 4773 - 4798, doi: 10.1175/2008MWR2363.1.

Murai, S., S. Yabu, and H. Kitagawa, 2005: Development of a new radiation scheme for the global atmospheric NWP model. *21st Conference on Weather Analysis and Forecasting and 17th Conference on Numerical Weather Prediction*, American Meteorological Society, Washington, D. C. [5pp., available at **https://ams.confex.com/ams/pdfpapers/94316.pdf**; accessed November 2019].

Nash, J., T. Oakley, H. Vömel, and W. Li, 2011: WMO intercomparison of high quality radiosonde systems, Yangjiang, China, 12 July–3 August 2010. *Instruments and Observing Methods Report* No. **107**, WMO/TD-No. 1580, 238 pp. [available at **https://library.wmo.int/index.php?lvl=notice_display&id=15531**; accessed 9 August 2017].

Nash, J., and R. Saunders, 2015: A review of Stratospheric Sounding Unit radiance observations for climate trends and reanalyses. *Q. J. R. Meteorol. Soc.*, **141**, 2103 - 2113, doi: 10.1002/qj.2505.

Onogi, K., 1998: A data quality control method using forecasted horizontal gradient and tendency in a NWP system: Dynamic QC. *J. Meteorol. Soc. Japan*, **76**, 497 - 516, doi: 10.2151/jmsj1965.76.4_497.

Onogi, K., *et al.*, 2007: The JRA-25 reanalysis. *J. Meteorol. Soc. Jpn.*, **85**(3), 369 - 432, doi: 10.2151/jmsj.85.369.

Orr, A., *et al.*, 2010: Improved middle atmosphere climate and forecasts in the ECMWF model through a nonorographic gravity wave drag parameterization. *J. Climate*, **23**, 5905 - 5926, doi: 10.1175/2010JCLI3490.1.

Orsolini, Y.J., and N. Kvamstø, 2009: The role of the Eurasian snow cover upon the wintertime circulation: decadal simulations forced with satellite observations. *J. Geophys. Res. Atmos.*, **114**, D19108, doi: 10.1029/2009JD012253.

Osborn, T.J., and P.D. Jones, 2014: The CRUTEM4 land-surface air temperature data set: construction, previous versions and dissemination via Google Earth. *Earth Syst. Sci. Data*, **6**, 61 - 68, doi: 10.5194/essd-6-61-2014.

Palmer, T.N., G.J. Schutts, and R. Swinbank, 1986: Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parameterization. *Q. J. R. Meteorol. Soc.*, **112**, 1001 - 1039, doi: 10.1002/qj.49711247406.

Paltridge, G.W., and C.M.R. Platt, 1976: Radiative Processes in Meteorology and Climatology. *Elsevier*, New York, 336 pp.

Pan, H.-L., and L. Mahrt, 1987: Interaction between soil hydrology and boundary-layer development. *Bound.-Lay. Meteorol.*, **38**, 185 - 202, doi: 10.1007/BF00121563.

Pan, H.-L., and W. Wu, 1995: Implementing a mass flux convective parameterization package for the NMC Medium-Range Forecast model. *NMC Office Note*, No. **409**, 43 pp. [available at **http://www.ncep.noaa.gov/officenotes/NOAA-NPM-NCEPON-0005/01408A42.pdf**; accessed August 2018].

Park, S.K., and D. Županski 2003: Four-dimensional variational data assimilation for mesoscale and storm-scale applications. *Meteorol. Atmos. Phys.*, **82**, 173 - 208, doi: 10.1007/s00703-001-0586-7.

Parrish, D.F., and J.C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747 - 1763, doi: 10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.

Petersen, R.A., 2016: On the impact and benefits of AMDAR observations in operational forecasting - Part I: A review of the impact of automated aircraft wind and temperature reports. *Bull. Am. Meteor. Soc.*, **97** 585 - 602, doi: 10.1175/BAMS-D-14-00055.1.

Pierrehumbert, R.T., 1986: An essay on the parameterization of orographic gravity wave drag. *Seminar/Workshop 1986 Observation, Theory and Modelling of Orographic Effects*, ECMWF, Shinfield Park, Reading, UK, 251 - 282. [available at **https://www.ecmwf.int/en/elibrary/11673-essay-parameterization-orographic-gravity-wave-drag**; accessed August 2018].

Polavarapu, S. , T.G. Shepherd, Y. Rochon, and S. Ren, 2005: Some challenges of middle atmosphere data assimilation. *Q. J. R. Meteorol. Soc.*, **131**, 3513 - 3527, doi: 10.1256/qj.05.87.

Polavarapu, S.M. and M. Pulido, 2017: Stratospheric and mesospheric data assimilation: The role of middle atmospheric dynamics. In: *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, vol. **3**, edited by: S. K. Park and L. Xu, Springer-Verlag, Berlin, pp. 429 - 454, doi: 10.1007/978-3-319-43415-5_19.

Poli, P., S.B. Healy, and D.P. Dee, 2010: Assimilation of Global Positioning System radio occultation data in the ECMWF ERA–Interim reanalysis. *Q. J. R. Meteorol. Soc.*, **136**, 1972 - 1990, doi: 10.1002/qj.722.

Poli, P., *et al.*, 2013: The data assimilation system and initial performance of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C), ERA Report Series, Vol. 14. [available at **https://www.ecmwf.int/en/elibrary/11699-data-assimilation-system-and-initial-performance-evaluation-ecmwf-pilot-reanalysis**; accessed August 2018].

Poli, P., *et al.*, 2016: ERA-20C: An atmospheric reanalysis of the 20th century, *J. Climate*, **29**, 4083 - 4097, doi: 10.1175/JCLI-D-15-0556.1.

Putman, W. and S.-J. Lin, 2007: Finite volume transport on various cubed sphere grids. *J. Comput. Phys.*, **227**, 55 - 78, doi: 10.1016/j.jcp.2007.07.022.

Randel, W., *et al.*, 2004: The SPARC intercomparison of middle-atmosphere climatologies, *J. Climate*, **17**, 986 - 1003, doi: 10.1175/1520-0442(2004)017<0986:TSIOMC>2.0.CO;2.

Randles, C.A., *et al.*, 2017: The MERRA-2 Aerosol Reanalysis, 1980–onward, Part I: System Description and Data Assimilation Evaluation. *J. Climate*, **30**, 6823 - 6850, doi: 10.1175/JCLI-D-16-0609.1.

Reale, A.L., 2001: NOAA operational sounding products from Advanced-TOVS polar orbiting environmental satellites. *NOAA Technical Report NESDIS* **102**, 59 pp. [available at **https://repository.library.noaa.gov/view/noaa/1123**; accessed August 2018].

Reichle, R.H., *et al.*, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, **24**, 6322 - 6338, doi: 10.1175/JCLI-D-10-05033.1.

Reichle, R.H., *et al.*, 2017a: Land surface precipitation in MERRA-2. *J. Climate*, **30**, 1643 - 1664, doi: 10.1175/JCLI-D-16-0570.1.

Reichle, R.H., *et al.*, 2017b: Assessment of MERRA-2 land surface hydrology estimates. *J. Climate*, **30**, 2937 - 2960, doi: 10.1175/JCLI-D-16-0720.1.

Reynolds, R.W., and T.M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 1609 - 1625, doi: 10.1175/1520-0442(1994)007<0929:IGSSTA>2.0.CO;2.

Reynolds, R.W., *et al.*, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609 - 1625, doi: 10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

Reynolds, R.W., *et al.*, 2007: Daily high-resolution blended analyses for sea surface temperature. *J. Climate*, **20**, 5473 - 5496, doi: 10.1175/2007JCLI1824.1.

Riehl, H., and J.S. Malkus, 1958: On the heat balance in the equatorial trough zone. *Geophysica*, **6**, 503 - 538.

Rienecker, M.M., *et al.*, 2008: The GEOS-5 Data Assimilation System—Documentation of versions 5.0.1 and 5.1.0, and 5.2.0. *NASA Tech. Rep. Series on Global Modeling and Data Assimilation*, Vol. **27**, NASA/TM-2008-104606, Greenbelt, Maryland, 92 pp. [available at **https://ntrs.nasa.gov/search.jsp?R=20120011955**; accessed August 2018].

Rienecker, M.M., *et al.*, 2011: MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Climate*, **24**, 3624 - 3648, doi: 10.1175/JCLI-D-11-00015.1.

Rosenfield, J.E., M.R. Schoeberl, and M.A. Geller, 1987: A computation of the stratospheric diabatic circulation using an accurate radiative transfer model. *J. Atmos. Sci.*, **44**, 859 - 876, doi: 10.1175/1520-0469(1987)044<0859:ACOTSD>2.0.CO;2.

Saha, S., *et al.*, 2010: The NCEP climate forecast system reanalysis. *Bull. Am. Meteor. Soc.*, **91**(8), 1015 - 1057, doi: 10.1175/2010BAMS3001.1.

Saha, S., *et al.*, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185 - 2208, doi: 10.1175/JCLI-D-12-00823.1.

Sakamoto, M., 2009: An update to the quality control thresholds of the conventional observing system for global data assimilation. *CAS/JSC WGNE Res. Activ. Atmos. Oceanic Modell.*, **39**, 33 - 34.

Sakamoto, M. and J.R. Christy, 2009: The influences of TOVS radiance assimilation on temperature and moisture tendencies in JRA-25 and ERA-40. *J. Atmos. Ocean. Tech.*, **26**, 1435 - 1455, doi:10.1175/2009JTECHA1193.1.

Sankey, D., *et al.*, 2007: Impact of data assimilation filtering methods on the mesosphere. *J. Geophys. Res. Atmos.*, **112**, D24104, doi: 10.1029/2007JD008885.

Sato N., *et al.*, 1989: Effects of implementing the simple biosphere model (SiB) in a general circulation model. *J. Atmos. Sci.*, **46**, 2757 - 2782, doi: 10.1175/1520-0469(1989)046<2757:EOITSB>2.0.CO;2.

Sato, M., J.E. Hansen, M.P. McCormick, and J.B. Pollack, 1993: Stratospheric aerosol optical depths. 1850 - 1990, *J. Geophys. Res. Atmos.*, **98**, 22987 - 22994, doi: 10.1029/93JD02553.

Schoeberl, M.R., A.R. Douglass, Z.Zhu, and S. Pawson, 2003: A comparison of the lower stratospheric age spectra derived from a general circulation model and two data assimilation systems. *J. Geophys. Res. Atmos.*, **108**, 4113, doi: 10.1029/2002JD002652.

Schubert, S.D., J. Pfaendtner, and R. Rood, 1993: An assimilated data set for Earth science applications. *Bull. Am. Meteor. Soc.*, **74**, 2331 - 2342, doi: 10.1175/1520-0477(1993)074<2331:AADFES>2.0.CO;2.

Schubert, S.D., *et al.*, 1995: A multiyear assimilation with the GEOS-1 system: Overview and results. *NASA Tech. Rep. Series on Global Modeling and Data Assimilation*, Vol. **6**, NASA/TM-104606, Greenbelt, Maryland, 201 pp. [available at **https://ntrs.nasa.gov/search.jsp?R=19950020003**; accessed August 2018].

Schwarzkopf, M.D., and S. Fels, 1991: The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *J. Geophys Res. Atmos.*, **96**, 9075 - 9096, doi: 10.1029/89JD01598.

Scinocca, J.F., 2003: An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *J. Atmos. Sci.*, **60**, 667 - 682, doi: 10.1175/1520-0469(2003)060<0667:AASNGW>2.0.CO;2.

Seidel, D.J., et al., 2009: Reference upper-air observations for climate: Rationale, progress, and plans. *Bull. Am. Meteor. Soc.*, **90**, 361 - 369, doi: 10.1175/2008BAMS2540.1.

Sellers, P.J., Y. Mintz, Y.C. Sud, and A. Dalcher, 1986: A simple biosphere model (SiB) for use within general circulation models. *J. Atmos. Sci.*, **43**, 505 - 531, doi: 10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2.

Shepherd, T.G., K. Semeniuk, and J.N. Koshyk, 1996: Sponge layer feedbacks in middle-atmosphere models. *J. Geophys. Res. Atmos.*, **101**, 23447 - 23464, doi: 10.1029/96JD01994.

Shepherd, T.G. and T.A. Shaw, 2004: The angular momentum constraint on climate sensitivity and downward influence in the middle atmosphere. *J. Atmos. Sci.*, **61**, 2899 - 2908, doi: 10.1175/JAS-3295.1.

Simmons, A.J., and D.M. Burridge, 1981: An energy and angular-momentum conserving finite-difference scheme and hybrid vertical coordinates. *Mon. Wea. Rev.*, **109**, 758 - 766, doi: 10.1175/1520-0493(1981)109<0758:AEAAMC>2.0.CO;2.

Simmons, A.J., *et al.*, 2004: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res. Atmos.*, **109**, D24115, doi: 10.1029/2004JD005306.

Simmons, A.J., *et al.*, 2014: Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. *Q. J. R. Meteorol. Soc.*, **140**, 329 - 353. doi: 10.1002/qj.2317.

Slivinski, L.C., *et al.*, 2019: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Q. J. R. Meteorol. Soc.,* **145**, 2876 - 2908, doi: 10.1002/qj.3598.

Smith, R.N.B., 1990: A scheme for predicting layer clouds and their water content in a general circulation model. *Q. J. R. Meteorol. Soc.*, **116**, 435 - 460. doi: 10.1002/qj.49711649210.

Spencer, J.W., 1971: Fourier series representation of the position of the sun. *Search*, **2**, 172 [available at **http://www.mail-archive.com/sundial@uni-koeln.de/msg01050.html**; accessed August 2018].

Stajner, I., *et al.*, 2008: Assimilated ozone from EOS-Aura: Evaluation of the tropopause region and tropospheric columns. *J. Geophys. Res.*, **113**, D16S32, doi: 10.1029/2007JD008863.

Stieglitz, M., A. Ducharne, R. Koster, and M. Suarez, 2001: The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales. *J. Hydrometeorol.*, **2**, 228 - 242, doi: 10.1175/1525-7541(2001)002<0228:TIODSP>2.0.CO;2.

Sundqvist, H., 1978: A parameterization scheme for non-convective condensation including prediction of cloud water content. *Q. J. R. Meteorol. Soc.*, **104**, 677 - 690, doi: 10.1002/qj.49710444110.

Takacs, L.L., M. Suarez, and R. Todling, 2015: Maintaining atmospheric mass and water balance within reanalysis. *NASA Tech. Rep. Series on Global Modeling and Data Assimilation*, Vol. **37**, NASA/TM-2014-104606, Greenbelt, Maryland, 46 pp. [available at **https://ntrs.nasa.gov/search.jsp?R=20150007977**; accessed August 2018].

Talagrand, O., 2010: Variational assimilation. In: *Data Assimilation: Making Sense of Observations*, edited by: W. Lahoz, B. Khattatov, and R. Menard, Springer-Verlag, Berlin, pp. 41 - 67, doi: 10.1007/978-3-540-74703-1_3.

Tan, W.W., M.A. Geller, S. Pawson, and A. da Silva, 2004: A case study of excessive subtropical transport in the stratosphere of a data assimilation system. *J. Geophys. Res. Atmos.*, **109**, D11102, doi: 10.1029/2003JD004057.

Tanré, D., J.-F. Geleyn, and J.M. Slingo, 1984: First results of the introduction of an advanced aerosol-radiation interaction in the ECMWF low resolution global model. In: *Aerosols and their Climatic Effects*, H. E. Gerber and A. Deepak (eds.), A. Deepak Publishing: Hampton, VA, USA, pp. 133 - 177.

Tavolato, C., and L. Isaksen, 2011: Data usage and quality control for ERA-40, ERA-Interim, and the operational ECMWF data assimilation system. *ERA Report Series*, No. **7**, ECMWF, Reading, UK, 44 pp. [available at **https://www.ecmwf.int/en/elibrary/12573-data-usage-and-quality-control-era-40-era-interim-and-operational-ecmwf-data**; accessed August 2018].

Taylor, K.E., D. Williamson, and F. Zwiers, 2000: The sea surface temperature and sea ice concentration boundary conditions for AMIP II simulations. *Program for Climate Model Diagnosis and Intercomparison (PCMDI) Report* **60**, Lawrence Livermore National Laboratory. [available at **https://pcmdi.llnl.gov/mips/amip/**; accessed August 2018].

Tegen, I., *et al.*, 1997: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *J. Geophys. Res. Atmos.*, **102**, 23895 - 23915, doi: 10.1029/97JD01864.

Tiedtke, M., 1983: The sensitivity of the time-mean large-scale flow to cumulus convection in the ECMWF model. *ECMWF Workshop on Convection in Large-Scale Models, Reading*, UK, ECMWF, 297 - 316. [available at **https://www.ecmwf.int/en/elibrary/12733-sensitivity-time-mean-large-scale-flow-cumulus-convection-ecmwf-model**; accessed 9 August 2017].

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779 - 1800, doi: 10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.

Tiedtke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121**, 3040 - 3061, doi: 10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2.

Titchner, H.A. and N.A. Rayner, 2014: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.*, **119**, 2864 - 2889, doi: 10.1002/2013JD020316.

Tompkins, A.M., K. Gierens, and G. Rädel, 2007: Ice supersaturation in the ECMWF integrated forecast system. *Q. J. R. Meteorol. Soc.*, **133**, 53 - 63, doi: 10.1002/qj.14.

Troen, I. and Mahrt, L., 1986: A simple model of the atmospheric boundary layer; sensitivity to surface evaporation. *Bound.-Lay. Meteorol.*, **37**, 129 - 148, doi: 10.1007/BF00122760.

Tsuyuki, T., and T. Miyoshi, 2007: Recent progress of data assimilation methods in meteorology, *J. Meteorol. Soc. Jpn.*, **85B**, 331 - 361, doi: 10.2151/jmsj.85B.331.

Uppala, S.M., *et al.*, 2005: The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.*, **131**(612), 2961 - 3012, doi:10.1256/qj.04.176.

USGS 2000: Global land cover characteristics data base, version 2.0. [Available at **https://lta.cr.usgs.gov/glcc/globdoc2_0**; accessed August 2018].

van den Hurk, B.J.J., P. Viterbo, A.C.M. Beljaars, and A. Betts, 2000: Offline validation of the ERA40 surface scheme. *Tech. Memo*. **295**, ECMWF, Shinfield Park, Reading, UK, 43 pp. [available at **https://www.ecmwf.int/en/elibrary/12900-offline-validation-era40-surface-scheme**; accessed August 2018].

van Vuuren D.P., *et al.*, 2011: The representative concentration pathways: An overview. *Clim. Change*, **109**, 5 - 31, doi: 10.1007/s10584-011-0148-z.

Walsh, J.E., and W.L. Chapman, 2001: 20th-century sea-ice variations from observational data. *Ann. Glaciol.*, **33**, 444 - 448, doi: 10.3189/172756401781818671.

Wang, L., C.-Z. Zou, and H. Qian, 2012: Construction of stratospheric temperature data records from Stratospheric Sounding Units. *J. Climate*, **25**, 2931 - 2946, doi: 10.1175/JCLI-D-11-00350.1.

Wang, W. and C.-Z. Zou, 2014: AMSU-A-only atmospheric temperature data records from the lower troposphere to the top of the stratosphere. *J. Atmos. Ocean. Tech.*, **31**, 808 - 825, doi: 10.1175/JTECH-D-13-00134.1.

Wargan, K., *et al.*, 2017: Evaluation of the ozone fields in NASA's MERRA-2 reanalysis. *J. Climate*, **30**, 2961 - 2988, doi: 10.1175/JCLI-D-16-0699.1.

Whitaker, J.S., G.P. Compo, and J.-N. Thépaut, 2009: A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Mon. Wea. Rev.*, **137**, 1991 - 1999, doi:10.1175/2008MWR2781.1.

WMO, 1986: A preliminary cloudless standard atmosphere for radiation computation. *World Climate Programme (WCP)*, **112**, 53 pp. [Available at **https://library.wmo.int/index.php?lvl=notice_display&id=11668**; accessed April 2021].

Woodruff, S.D., et al., 2011: ICOADS release 2.5: Extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.*, **31**, 951 - 967, doi: 10.1002/ joc.2103.

Wright, J., 2017: S-RIP: Zonal-mean heating rates of global atmospheric reanalyses on pressure levels [data set]. *Centre for Environmental Data Analysis (CEDA)*, doi: 10.5285/70146c789eda4296a3c3ab6706931d56.

Wu, W., R.J. Purser, and D.F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905 - 2916, doi: 10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2.

Xie, S., and M. Zhang, 2000: Impact of the convection triggering function on single-column model simulations, *J. Geophys. Res. Atmos.*, **105**, 14983 - 14996, doi: 10.1029/2000JD900170.

Xu, K.-M., and S.K. Krueger, 1991: Evaluation of cloudiness parameterizations using a cumulus ensemble model. *Mon. Wea. Rev.*, **119**, 342 - 367, doi: 10.1175/1520-0493(1991)119<0342:EOCPUA>2.0.CO;2.

Xu, K.-M., and D.A. Randall, 1996: A semiempirical cloudiness parameterization for use in climate models. *J. Atmos. Sci.*, **53**, 3084 - 3102, doi: 10.1175/1520-0469(1996)053<3084:ASCPFU>2.0.CO;2.

Zhao, Q.Y., and F.H. Carr, 1997: A prognostic cloud scheme for operational NWP models. *Mon. Wea. Rev.*, **125**, 1931 - 1953, doi: 10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2.

Zou, C.-Z., et al., 2006: Recalibration of microwave sounding unit for climate studies using simultaneous nadir overpasses. *J. Geophys. Res.*, **111**, D19114, doi: 10.1029/2005JD006798.

Zou, C.-Z., and W. Wang, 2011: Intersatellite calibration of AMSU-A observations for weather and climate applications. *J. Geophys. Res.*, **116**, D23113, doi: 10.1029/2011JD016205.

Zou, C.-Z., et al., 2014: Recalibration and merging of SSU observations for stratospheric temperature trend studies. *J. Geophys. Res.*, **119**, 13180 - 13205, doi: 10.1002/2014JD021603.

Zou, C.-Z., and H. Qian, 2016: Stratospheric temperature climate data record from merged SSU and AMSU-A observations. *J. Atmos. Ocean Tech.*, **33**, 1967 - 1984, doi: 10.1175/JTECH-D-16-0018.1.

Appendix A: Vertical levels of the models

A2.1 ERA-40 and ERA-Interim

ERA-40 and ERA-Interim both use hybrid sigma–pressure (hybrid σ–p) vertical coordinates (*Simmons and Burridge*, 1981), which are also sometimes referred to as eta (η) vertical coordinates (see also **http://rda.ucar.edu/datasets/ds627.0/docs/Eta_coordinate/**). Both systems use the same vertical resolution with 61 levels (*Kållberg et al.*, 2007). The pressure on each level is calculated as $p_k = A_k + B_k \times p_{srf}$, where p_{srf} is surface pressure. The following table provides example pressures at layer interfaces (k–1/2) and layer midpoints (k) for a surface pressure of 1013.25 hPa, from TOA to surface. Pressures at layer midpoints are defined as the average of pressures at layer interfaces. Pressure levels in brackets are used for ERA-Interim products but not for ERA-40 products.

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	p (hPa)
1	0.00	0.00000	0.00	0.10	
2	0.20	0.00000	0.20	0.29	
3	0.38	0.00000	0.38	0.51	
4	0.64	0.00000	0.64	0.80	
5	0.96	0.00000	0.96	1.15	1
6	1.34	0.00000	1.34	1.58	
7	1.81	0.00000	1.81	2.08	2
8	2.35	0.00000	2.35	2.67	3
9	2.98	0.00000	2.98	3.36	
10	3.74	0.00000	3.74	4.19	
11	4.65	0.00000	4.65	5.20	5
12	5.76	0.00000	5.76	6.44	7
13	7.13	0.00000	7.13	7.96	
14	8.84	0.00000	8.84	9.89	10
15	10.95	0.00000	10.95	12.26	
16	13.56	0.00000	13.56	15.19	
17	16.81	0.00000	16.81	18.81	20
18	20.82	0.00000	20.82	23.31	
19	25.80	0.00000	25.80	28.88	30
20	31.96	0.00000	31.96	35.78	
21	39.60	0.00000	39.60	44.33	
22	49.07	0.00000	49.07	54.62	50
23	60.18	0.00000	60.18	66.62	70
24	73.07	0.00000	73.07	80.40	
25	87.65	0.00008	87.73	95.98	100
26	103.76	0.00046	104.23	113.42	
27	120.77	0.00182	122.61	132.76	(125)
28	137.75	0.00508	142.90	154.00	150
29	153.80	0.01114	165.09	177.12	(175)
30	168.19	0.02068	189.15	202.09	200

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	p (hPa)
31	180.45	0.03412	215.03	228.84	(225)
32	190.28	0.05169	242.65	257.36	250
33	197.55	0.07353	272.06	287.64	300
34	202.22	0.09967	303.22	319.63	
35	204.30	0.13002	336.04	353.23	(350)
36	203.84	0.16438	370.41	388.27	400
37	200.97	0.20248	406.13	424.57	
38	195.84	0.24393	443.01	461.90	(450)
39	188.65	0.28832	480.79	500.00	500
40	179.61	0.33515	519.21	538.591	(550)
41	168.99	0.38389	557.97	577.38	
42	157.06	0.43396	596.78	616.04	600
43	144.11	0.48477	635.31	654.27	(650)
44	130.43	0.53571	673.24	691.75	700
45	116.33	0.58617	710.26	728.16	
46	102.10	0.63555	746.06	763.20	(750), 775
47	88.02	0.68327	780.35	796.59	(800)
48	74.38	0.72879	812.83	828.05	(825)
49	61.44	0.77160	843.26	857.34	850
50	49.42	0.81125	871.42	884.27	(875)
51	38.51	0.84737	897.11	908.65	(900)
52	28.88	0.87966	920.19	930.37	925
53	20.64	0.90788	940.55	949.35	(950)
54	13.86	0.93194	958.15	965.57	
55	8.55	0.95182	972.99	979.06	(975)
56	4.67	0.96765	985.14	989.95	
57	2.10	0.97966	994.75	998.39	1000
58	0.66	0.98827	1002.02	1004.64	
59	0.07	0.99402	1007.26	1009.06	
60	0.00	0.99763	1010.85	1012.05	
	0.00	1.00000	1013.25		

A2.2 ERA-20C

ERA-20C uses hybrid sigma–pressure (hybrid σ–p) vertical coordinates (*Simmons and Burridge*, 1981) with 91 levels. The pressure on each level is calculated as $p_k = A_k + B_k \times p_{srf}$, where p_{srf} is surface pressure. The following table provides example pressures at layer interfaces (k–1/2) and layer midpoints (k) for a surface pressure of 1013.25 hPa, from TOA to surface. Pressures at layer midpoints are defined as the average of pressures at layer interfaces.

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	p (hPa)
1	0	0	0	0.01	
2	0.02	0	0.02	0.03	
3	0.03980832	0	0.04	0.06	
4	0.07387186	0	0.07	0.10	
5	0.12908319	0	0.13	0.17	
6	0.21413612	0	0.21	0.28	
7	0.33952858	0	0.34	0.43	
8	0.51746601	0	0.52	0.64	
9	0.76167656	0	0.76	0.92	1
10	1.08715561	0	1.09	1.30	
11	1.50986023	0	1.51	1.78	2
12	2.04637451	0	2.05	2.38	
13	2.71356506	0	2.71	3.12	3
14	3.52824493	0	3.53	4.02	
15	4.50685791	0	4.51	5.09	5
16	5.66519226	0	5.67	6.34	7
17	7.01813354	0	7.02	7.80	
18	8.57945801	0	8.58	9.47	10
19	10.36166504	0	10.36	11.37	
20	12.37585449	0	12.38	13.50	
21	14.6316394	0	14.63	15.88	
22	17.13709595	0	17.14	18.52	
23	19.8987439	0	19.90	21.41	20
24	22.92155518	0	22.92	24.57	
25	26.20898438	0	26.21	27.99	
26	29.76302246	0	29.76	31.67	30
27	33.58425781	0	33.58	35.63	
28	37.67196045	0	37.67	39.85	
29	42.02416504	0	42.02	44.33	
30	46.63776367	0	46.64	49.07	50
31	51.50859863	0	51.51	54.07	
32	56.6315625	0	56.63	59.31	
33	61.99839355	0	62.00	64.80	
34	67.59727051	0	67.60	70.51	70
35	73.41469727	0	73.41	76.43	
36	79.4292627	0.000014	79.44	82.57	
37	85.64624023	0.000055	85.70	88.96	
38	92.08305664	0.000131	92.22	95.62	
39	98.73560547	0.000279	99.02	102.58	100
40	105.5888184	0.000548	106.14	109.89	
41	112.6248438	0.001	113.64	117.59	
42	119.8266211	0.001701	121.55	125.75	125
43	127.1389746	0.002765	129.94	134.40	
44	134.5322559	0.004267	138.86	143.59	
45	141.9200977	0.006322	148.33	153.35	150

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	p (hPa)
46	149.2268555	0.009035	158.38	163.72	
47	156.3805371	0.012508	169.05	174.72	175
48	163.2956055	0.01686	180.38	186.38	
49	169.9062305	0.022189	192.39	198.76	200
50	176.1328125	0.02861	205.12	211.87	
51	181.910293	0.036227	218.62	225.77	225
52	187.1696875	0.045146	232.91	240.48	
53	191.8454492	0.055474	248.05	256.07	250
54	195.8751367	0.067316	264.08	272.56	
55	199.1979688	0.080777	281.05	290.02	
56	201.7539453	0.095964	298.99	308.48	300
57	203.4891602	0.112979	317.97	327.99	
58	204.341582	0.131935	338.02	348.62	350
59	204.2621875	0.152934	359.22	370.42	
60	203.1901172	0.176091	381.61	393.44	400
61	201.0703125	0.20152	405.26	417.73	
62	197.8535742	0.229315	430.21	443.34	450
63	193.4877539	0.259554	456.48	470.17	
64	187.9882227	0.291993	483.85	497.96	500
65	181.4129688	0.326329	512.07	526.46	
66	173.855957	0.362203	540.86	555.40	550
67	165.4458594	0.399205	569.94	584.49	
68	156.3356641	0.436906	599.03	613.50	600
69	146.6564551	0.475016	627.97	642.29	650
70	136.5321973	0.51328	656.61	670.73	
71	126.0838379	0.551458	684.85	698.70	700
72	115.4316699	0.589317	712.56	726.07	
73	104.7131055	0.626559	739.57	752.67	750
74	94.05222656	0.662934	765.77	778.40	775
75	83.5625293	0.698224	791.04	803.16	800
76	73.35164551	0.732224	815.28	826.81	825
77	63.53920898	0.764679	838.35	849.25	850
78	54.22802734	0.795385	860.15	870.38	875
79	45.5021582	0.824185	880.61	890.13	
80	37.43464355	0.85095	899.66	908.44	900
81	30.10146973	0.875518	917.22	925.22	925
82	23.56202637	0.897767	933.22	940.44	
83	17.84854614	0.917651	947.66	954.09	950
84	12.97656128	0.935157	960.52	966.17	
85	8.95193542	0.950274	971.82	976.67	975
86	5.76314148	0.963007	981.53	985.63	
87	3.36772369	0.973466	989.73	993.30	
88	1.62043427	0.982238	996.87	999.84	1000
89	0.54208336	0.989153	1002.80	1005.12	
90	0.06575628	0.994204	1007.44	1009.15	
91	0.0000316	0.99763	1010.85	1012.05	
	0	1	1013.25		

A2.3 ERA5

ERA5 uses hybrid sigma–pressure (hybrid σ–p) vertical coordinates (*Simmons and Burridge*, 1981) with 137 levels. The pressure on each level is calculated as $p_k = A_k + B_k \times p_{\mathrm{srf}}$, where p_{srf} is surface pressure. The following table provides example pressures at layer interfaces (k–1/2) and layer midpoints (k) for a surface pressure of 1013.25 hPa, from TOA to surface. Pressures at layer midpoints are defined as the average of pressures at layer interfaces.

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	p (hPa)
1	0	0	0.00	0.01	
2	0.02000365	0	0.02	0.03	
3	0.03102241	0	0.03	0.04	
4	0.04666084	0	0.05	0.06	
5	0.06827977	0	0.07	0.08	
6	0.09746966	0	0.10	0.12	
7	0.13605424	0	0.14	0.16	
8	0.18608931	0	0.19	0.22	
9	0.24985718	0	0.25	0.29	
10	0.3298571	0	0.33	0.38	
11	0.42879242	0	0.43	0.49	
12	0.54955463	0	0.55	0.62	
13	0.69520576	0	0.70	0.78	
14	0.86895882	0	0.87	0.97	1
15	1.07415741	0	1.07	1.19	
16	1.31425507	0	1.31	1.45	
17	1.59279404	0	1.59	1.75	
18	1.91338562	0	1.91	2.10	2
19	2.27968948	0	2.28	2.49	
20	2.69539581	0	2.70	2.93	3
21	3.16420746	0	3.16	3.43	
22	3.68982361	0	3.69	3.98	
23	4.27592499	0	4.28	4.60	
24	4.92616028	0	4.93	5.29	5
25	5.64413452	0	5.64	6.04	
26	6.43339905	0	6.43	6.87	7
27	7.29744141	0	7.30	7.77	
28	8.23967834	0	8.24	8.75	
29	9.2634491	0	9.26	9.82	10
30	10.37201172	0	10.37	10.97	
31	11.56853638	0	11.57	12.21	
32	12.85610352	0	12.86	13.55	
33	14.23770142	0	14.24	14.98	
34	15.71622925	0	15.72	16.51	
35	17.29448975	0	17.29	18.13	
36	18.97519287	0	18.98	19.87	20
37	20.76095947	0	20.76	21.71	
38	22.65431641	0	22.65	23.66	
39	24.65770508	0	24.66	25.72	
40	26.77348145	0	26.77	27.89	
41	29.00391357	0	29.00	30.18	30
42	31.35119385	0	31.35	32.58	
43	33.81743652	0	33.82	35.11	
44	36.40468262	0	36.40	37.76	
45	39.11490479	0	39.11	40.53	
46	41.94930664	0	41.95	43.43	
47	44.90817383	0	44.91	46.45	
48	47.99149414	0	47.99	49.60	50
49	51.1989502	0	51.20	52.86	
50	54.52990723	0	54.53	56.26	
51	57.98344727	0	57.98	59.77	
52	61.56074219	0	61.56	63.42	
53	65.26946777	0	65.27	67.19	
54	69.11870605	0	69.12	71.12	70
55	73.11869141	0	73.12	75.20	
56	77.27412109	0.000007	77.28	79.45	
57	81.59354004	0.000024	81.62	83.88	
58	86.08525391	0.000059	86.15	88.51	
59	90.76400391	0.000112	90.88	93.35	
60	95.62682617	0.000199	95.83	98.42	
61	100.6597852	0.00034	101.00	103.71	100
62	105.8463184	0.000562	106.42	109.24	
63	111.1666211	0.00089	112.07	115.02	
64	116.6006738	0.001353	117.97	121.05	
65	122.1154785	0.001992	124.13	127.35	125
66	127.6687305	0.002857	130.56	133.92	
67	133.2466895	0.003971	137.27	140.77	
68	138.8133106	0.005378	144.26	147.91	150
69	144.3213965	0.007133	151.55	155.34	
70	149.7561523	0.009261	159.14	163.09	
71	155.0825684	0.011806	167.04	171.16	175
72	160.2611523	0.014816	175.27	179.55	
73	165.2732227	0.018318	183.83	188.29	
74	170.0878906	0.022355	192.74	197.37	200
75	174.6761328	0.026964	202.00	206.81	
76	179.0162109	0.032176	211.62	216.62	
77	183.0843359	0.038026	221.61	226.80	225
78	186.8571875	0.044548	232.00	237.38	
79	190.3128906	0.051773	242.77	248.36	250
80	193.4351172	0.059728	253.95	259.75	
81	196.2004297	0.068448	265.56	271.57	
82	198.5939063	0.077958	277.58	283.82	
83	200.5993164	0.088286	290.06	296.52	300
84	202.1966406	0.099462	302.98	309.67	
85	203.3786328	0.111505	316.36	323.29	
86	204.1230859	0.124448	330.22	337.39	
87	204.4207813	0.138313	344.57	351.99	350
88	204.2571875	0.153125	359.41	367.09	
89	203.6181641	0.16891	374.77	382.71	
90	202.4951172	0.185689	390.64	398.85	400
91	200.8708594	0.203491	407.06	415.54	
92	198.7402539	0.222333	424.02	432.78	
93	196.0857227	0.242244	441.54	450.59	450
94	192.9022656	0.263242	459.63	468.97	
95	189.1746094	0.285354	478.31	487.95	
96	184.8970703	0.308598	497.58	507.50	500
97	180.0692578	0.332939	517.42	527.57	
98	174.7183984	0.358254	537.72	548.03	550
99	168.886875	0.384363	558.34	568.77	
100	162.6204688	0.411125	579.19	589.68	600
101	155.9669531	0.438391	600.17	610.66	
102	148.9845313	0.466003	621.16	631.62	
103	141.7332422	0.4938	642.08	652.44	650
104	134.2776953	0.521619	662.81	673.03	
105	126.6825781	0.549301	683.26	693.30	700
106	119.0133984	0.576692	703.35	713.16	
107	111.3330469	0.603648	722.98	732.53	
108	103.7017578	0.630036	742.09	751.34	750
109	96.17515625	0.655736	760.60	769.53	775
110	88.80453125	0.680643	778.47	787.05	
111	81.63375	0.704669	795.64	803.86	800
112	74.7034375	0.727739	812.08	819.93	825
113	68.04421875	0.749797	827.78	835.24	
114	61.6853125	0.770798	842.70	849.77	850
115	55.64382813	0.790717	856.84	863.52	
116	49.93796875	0.809536	870.20	876.50	875
117	44.57375	0.827256	882.79	888.71	
118	39.55960938	0.843881	894.62	900.17	900
119	34.89234375	0.859432	905.71	910.90	
120	30.57265625	0.873929	916.08	920.92	925
121	26.59140625	0.887408	925.76	930.26	
122	22.94242188	0.8999	934.77	938.95	
123	19.615	0.911448	943.14	947.02	
124	16.59476563	0.922096	950.91	954.51	950
125	13.87546875	0.931881	958.10	961.43	
126	11.4325	0.94086	964.76	967.83	
127	9.26507813	0.949064	970.90	973.74	975
128	7.34992188	0.95655	976.57	979.19	
129	5.680625	0.963352	981.80	984.20	
130	4.24414063	0.969513	986.60	988.81	
131	3.02476563	0.975078	991.02	993.05	
132	2.02484375	0.980072	995.08	996.95	
133	1.22101563	0.984542	998.81	1000.52	1000
134	0.6278125	0.9885	1002.23	1003.79	
135	0.22835938	0.991984	1005.36	1006.79	
136	0.03757813	0.995003	1008.22	1009.54	
137	0	0.99763	1010.85	1012.05	
	0	1	1013.25		

A2.4 JRA-25/JCDAS

JRA-25 used a hybrid sigma–pressure (hybrid σ–p) vertical coordinate after *Simmons and Burridge* (1981). The pressure on each level is calculated as $p_k = A_k + B_k \times p_{srf}$, where p_{srf} is surface pressure. The following table provides example pressures at layer interfaces (k–1/2) and layer midpoints (k) for a surface pressure of 1013.25 hPa, from TOA to surface. Pressures at layer midpoints are defined as the average of pressures at layer interfaces.

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	p (hPa)
1	0.000000	0.000000	0.00	0.40	0.4
2	0.800000	0.000000	0.80	1.13	1
3	1.460000	0.000000	1.46	2.01	2
4	2.560000	0.000000	2.56	3.45	3
5	4.330000	0.000000	4.33	5.72	5
6	7.100000	0.000000	7.10	9.15	7
7	11.200000	0.000000	11.20	14.10	10
8	17.000000	0.000000	17.00	21.00	20
9	25.000000	0.000000	25.00	30.15	30
10	35.299999	0.000000	35.30	41.70	
11	48.099998	0.000000	48.10	55.55	50
12	62.634430	0.000366	63.01	71.53	70
13	76.105057	0.003895	80.05	89.60	
14	88.363998	0.010636	99.14	109.71	100
15	98.876595	0.021123	120.28	131.88	
16	107.299492	0.035701	143.47	156.10	150
17	113.447090	0.054553	168.72	182.38	
18	117.259979	0.077740	196.03	210.71	200
19	118.777374	0.105223	225.39	241.10	250
20	118.113609	0.136886	256.81	273.55	
21	115.438545	0.172561	290.29	308.05	300
22	110.961449	0.212039	325.81	344.09	
23	105.094887	0.253905	362.36	381.16	
24	98.151306	0.297849	399.95	419.76	400
25	90.192863	0.344807	439.57	460.40	
26	81.437820	0.394562	481.23	502.57	500
27	72.323532	0.445676	523.90	545.75	
28	63.056015	0.497944	567.60	589.95	600
29	53.811684	0.551188	612.30	635.16	
30	44.741348	0.605259	658.02	680.87	700
31	36.158020	0.658842	703.73	726.58	
32	28.130577	0.711869	749.43	771.77	
33	20.862747	0.763137	794.11	815.43	
34	14.485500	0.811514	836.75	856.55	850
35	9.064261	0.855936	876.34	894.10	
36	4.611954	0.895388	911.86	932.15	925
37	1.105610	0.938894	952.44	960.05	
38	0.000000	0.955000	967.65	977.79	
39	0.000000	0.975000	987.92	995.52	1000
40	0.000000	0.990000	1003.12	1008.18	
	0.000000	1.000000	1013.25		

A2.5 JRA-55

JRA-55 uses a hybrid sigma–pressure (hybrid σ–p) vertical coordinate after *Simmons and Burridge*, (1981). The pressure on each level is calculated as $p_k = A_k + B_k \times p_{srf}$ where p_{srf} is surface pressure. The following table provides example pressures at layer interfaces (k–1/2) and layer midpoints (k) for a surface pressure of 1013.25 hPa, from TOA to surface. Pressures at layer midpoints are defined as the average of pressures at layer interfaces.

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	p (hPa)
1	0.000000	0.000000	0.00	0.10	
2	0.200000	0.000000	0.20	0.30	
3	0.390000	0.000000	0.39	0.52	
4	0.650000	0.000000	0.65	0.81	
5	0.970000	0.000000	0.97	1.17	1
6	1.360000	0.000000	1.36	1.59	
7	1.820000	0.000000	1.82	2.10	2
8	2.370000	0.000000	2.37	2.69	
9	3.010000	0.000000	3.01	3.39	3
10	3.770000	0.000000	3.77	4.23	
11	4.690000	0.000000	4.69	5.25	5
12	5.810000	0.000000	5.81	6.51	7
13	7.200000	0.000000	7.20	8.07	
14	8.930000	0.000000	8.93	9.99	10
15	11.050000	0.000000	11.05	12.38	
16	13.700000	0.000000	13.70	15.35	
17	17.000000	0.000000	17.00	19.03	20
18	21.050000	0.000000	21.05	23.58	
19	26.100000	0.000000	26.10	29.20	30
20	32.300000	0.000000	32.30	36.15	
21	40.000000	0.000000	40.00	44.75	
22	49.500000	0.000000	49.50	55.25	50
23	60.886730	0.000113	61.00	67.77	70
24	72.015690	0.002484	74.53	81.81	
25	82.262449	0.006738	89.09	97.13	100
26	91.672470	0.013328	105.18	114.24	
27	100.146151	0.022854	123.30	133.39	125
28	107.299494	0.035701	143.47	154.58	150
29	112.854041	0.052146	165.69	177.82	175
30	116.633554	0.072366	189.96	203.12	200
31	118.554343	0.096446	216.28	230.46	225
32	118.612531	0.124387	244.65	259.35	250
33	116.953716	0.155046	274.05	289.78	300
34	113.696478	0.189304	305.51	321.75	
35	109.126384	0.225874	337.99	355.26	350
36	103.294362	0.265706	372.52	390.30	400
37	96.561819	0.307438	408.07	426.36	
38	89.140822	0.350859	444.65	463.45	450
39	81.221598	0.395778	482.24	501.55	500
40	72.974699	0.442025	520.86	540.16	550
41	64.767182	0.488233	559.47	578.77	
42	56.718242	0.534282	598.08	617.38	600
43	48.918808	0.580081	636.69	655.48	650
44	41.629564	0.62437	674.27	693.06	700
45	34.688715	0.668311	711.85	729.63	750
46	28.474848	0.709525	747.40	764.16	775
47	22.948417	0.748052	780.91	797.16	800
48	17.909074	0.785091	813.40	828.63	825
49	13.4768	0.819523	843.86	858.07	850
50	9.597972	0.851402	872.28	884.97	875
51	6.346027	0.879654	897.66	908.82	900
52	3.649041	0.904351	919.98	930.13	925
53	1.33051	0.926669	940.28	949.41	950
54	0	0.946	958.53	965.63	
55	0	0.96	972.72	978.80	975
56	0	0.972	984.88	989.95	
57	0	0.982	995.01	998.56	1000
58	0	0.989	1002.10	1004.64	
59	0	0.994	1007.17	1008.69	
60	0	0.997	1010.21	1011.73	
	0	1	1013.25		

A2.6 MERRA and MERRA-2

MERRA and MERRA-2 use identical hybrid sigma–pressure (hybrid σ–p) vertical coordinates after *Simmons and Burridge*, (1981). The pressure on each level is calculated as $p_k = A_k + B_k \times p_{srf}$, where p_{srf} is surface pressure. The following table provides example pressures at layer interfaces (*k*–1/2) and layer midpoints (*k*) for a surface pressure of 1013.25 hPa, from TOA to surface. Pressures at layer midpoints are defined as the average of pressures at layer interfaces. NASA GMAO is transitioning away from this vertical grid and recommends that data users use the three-dimensional pressure fields provided with MERRA and MERRA-2 instead.

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	*p* (hPa)
1	0.0100	0	0.01	0.015	
2	0.0200	0	0.02	0.026	
3	0.0327	0	0.03	0.040	
4	0.0476	0	0.05	0.057	
5	0.0660	0	0.07	0.078	
6	0.0893	0	0.09	0.105	0.1
7	0.1197	0	0.12	0.140	
8	0.1595	0	0.16	0.185	
9	0.2113	0	0.21	0.245	
10	0.2785	0	0.28	0.322	0.3
11	0.3650	0	0.37	0.420	0.4
12	0.4758	0	0.48	0.546	0.5
13	0.6168	0	0.62	0.706	0.7
14	0.7951	0	0.80	0.907	1
15	1.0194	0	1.02	1.160	
16	1.3005	0	1.30	1.476	
17	1.6508	0	1.65	1.868	2
18	2.0850	0	2.08	2.353	
19	2.6202	0	2.62	2.948	3
20	3.2764	0	3.28	3.677	4
21	4.0766	0	4.08	4.562	5
22	5.0468	0	5.05	5.632	
23	6.2168	0	6.22	6.918	7
24	7.6198	0	7.62	8.456	
25	9.2929	0	9.29	10.285	10
26	11.2769	0	11.28	12.460	
27	13.6434	0	13.64	15.050	
28	16.4571	0	16.46	18.124	
29	19.7916	0	19.79	21.761	20
30	23.7304	0	23.73	26.049	
31	28.3678	0	28.37	31.089	30
32	33.8100	0	33.81	36.993	40
33	40.1754	0	40.18	43.910	
34	47.6439	0	47.64	52.016	50
35	56.3879	0	56.39	61.496	
36	66.6034	0	66.60	72.558	70
37	78.5123	0	78.51	85.439	
38	92.3657	0	92.37	100.514	100
39	108.6630	0	108.66	118.250	
40	127.8370	0	127.84	139.115	150
41	150.3930	0	150.39	163.662	
42	176.9300	0	176.93	192.587	200
43	201.1920	0.006960	208.24	226.745	
44	216.8650	0.028010	245.25	267.087	250
45	224.3630	0.063720	288.93	313.966	300
46	223.8980	0.113602	339.01	358.038	350
47	218.7760	0.156224	377.07	396.112	400
48	212.1500	0.200350	415.15	434.212	450
49	203.2590	0.246741	453.27	472.335	
50	193.0970	0.294403	491.40	510.475	500
51	181.6190	0.343381	529.55	548.628	550
52	169.6090	0.392891	567.71	586.793	600
53	156.2600	0.443740	605.88	624.966	
54	142.9100	0.494590	644.05	663.146	650
55	128.6960	0.546304	682.24	694.969	700
56	118.9590	0.581041	707.70	720.429	725
57	109.1820	0.615818	733.16	745.890	750
58	99.3652	0.650635	758.62	771.355	775
59	89.0999	0.685900	784.09	796.822	800
60	78.8342	0.721166	809.56	819.742	825
61	70.6220	0.749378	829.93	837.570	
62	64.3626	0.770637	845.21	852.852	850
63	58.0532	0.791947	860.49	868.135	875
64	51.6961	0.813304	875.78	883.418	
65	45.3390	0.834661	891.06	898.701	900
66	38.9820	0.856018	906.34	913.984	
67	32.5708	0.877429	921.63	929.268	925
68	26.0920	0.898908	936.91	944.553	950
69	19.6131	0.920387	952.20	959.837	
70	13.1348	0.941865	967.48	975.122	975
71	6.5938	0.963406	982.76	990.408	
72	0.0480	0.984952	998.05	1005.650	1000
	0	1	1013.25		

A2.7 NCEP-NCAR R1 and NCEP-DOE R2

NCEP-NCAR R1 and NCEP-DOE R2 use a sigma vertical coordinate. The pressure on each level is calculated as $p_k = \sigma_k \times p_{srf}$, where p_{srf} is surface pressure. The following table provides example pressures at each level for a surface pressure of 1013.25 hPa, from TOA to surface.

Model Levels			Pressure Levels
k	σ_k	p_k (hPa)	*p* (hPa)
1	0.00273	2.77	3
2	0.01006	10.19	10
3	0.01834	18.58	20
4	0.02875	29.13	30
5	0.04179	42.34	
6	0.05805	58.82	50
7	0.07815	79.19	70
8	0.10278	104.14	100
9	0.13261	134.37	
10	0.16823	170.46	150
11	0.21006	212.84	200
12	0.25823	261.65	250
13	0.31248	316.62	300
14	0.37205	376.98	400
15	0.43568	441.45	
16	0.50168	508.33	500
17	0.56809	575.62	
18	0.63290	641.29	
19	0.69426	703.46	700
20	0.75076	760.71	
21	0.80142	812.04	
22	0.84579	857.00	850
23	0.88384	895.55	
24	0.91592	928.06	925
25	0.94255	955.04	
26	0.96437	977.15	
27	0.98208	995.09	
28	0.99500	1008.18	1000
	1.00000	1013.25	

A2.8 CFSR

CFSR uses a hybrid sigma–pressure (hybrid σ–p) vertical coordinates after *Simmons and Burridge* (1981). The pressure on each level is calculated as $p_k = A_k + B_k \times p_{srf}$, where p_{srf} is surface pressure. The following table provides example pressures at layer interfaces (*k*–1/2) and layer midpoints (*k*) for a surface pressure of 1013.25 hPa, from TOA to surface. Pressures at layer midpoints are defined as the average of pressures at layer interfaces.

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	p (hPa)
1	0.00000	0.000000	0.00	0.32	
2	0.64247	0.000000	0.64	1.01	1
3	1.37790	0.000000	1.38	1.80	2
4	2.21958	0.000000	2.22	2.70	3
5	3.18266	0.000000	3.18	3.73	
6	4.28434	0.000000	4.28	4.91	5
7	5.54424	0.000000	5.54	6.26	
8	6.98457	0.000000	6.98	7.81	7
9	8.63058	0.000000	8.63	9.57	10
10	10.51080	0.000000	10.51	11.58	
11	12.65752	0.000000	12.66	13.88	
12	15.10711	0.000000	15.11	16.50	
13	17.90051	0.000000	17.90	19.49	20
14	21.08366	0.000000	21.08	22.90	
15	24.70788	0.000000	24.71	26.77	
16	28.83038	0.000000	28.83	31.17	30
17	33.51460	0.000000	33.51	36.17	
18	38.83052	0.000000	38.83	41.84	
19	44.85493	0.000000	44.85	48.26	50
20	51.67146	0.000000	51.67	55.52	
21	59.37050	0.000000	59.37	63.71	
22	68.04874	0.000000	68.05	72.93	70
23	77.77150	0.000037	77.81	83.29	
24	88.32537	0.000431	88.76	94.89	100
25	99.36614	0.001636	101.02	107.87	
26	110.54853	0.004107	114.71	122.32	125
27	121.52937	0.008294	129.93	138.37	
28	131.97065	0.014637	146.80	156.11	150
29	141.54316	0.023556	165.41	175.63	175
30	149.93074	0.035442	185.84	197.00	200
31	156.83489	0.050647	208.15	220.26	225
32	161.97967	0.069475	232.37	245.44	250
33	165.11736	0.092167	258.51	272.50	
34	166.11603	0.118812	286.50	301.39	300
35	165.03144	0.149269	316.28	331.99	
36	161.97315	0.183296	347.70	364.14	350
37	157.08893	0.220570	380.58	397.64	400
38	150.56342	0.260685	414.70	432.25	
39	142.61435	0.303164	449.80	467.68	450
40	133.48671	0.347468	485.56	503.61	500
41	123.44490	0.393018	521.67	539.73	550
42	112.76348	0.439211	557.79	575.69	
43	101.71712	0.485443	593.59	611.17	600
44	90.57051	0.531135	628.74	645.84	650
45	79.56908	0.575747	662.94	679.44	
46	68.93117	0.618800	695.93	711.70	700
47	58.84206	0.659887	727.47	742.43	750
48	49.45029	0.698683	757.39	771.47	775
49	40.86614	0.734945	785.55	798.70	800
50	33.16217	0.768515	811.86	824.07	825
51	26.37553	0.799310	836.28	847.53	850
52	20.51150	0.827319	858.79	869.11	875
53	15.54789	0.852591	879.44	888.85	
54	11.43988	0.875224	898.26	906.80	900
55	8.12489	0.895355	915.34	923.06	925
56	5.52720	0.913151	930.78	937.72	
57	3.56223	0.928797	944.67	950.89	950
58	2.14015	0.942491	957.12	962.68	
59	1.16899	0.954434	968.25	973.21	975
60	0.55712	0.964828	978.17	982.58	
61	0.21516	0.973868	986.99	990.90	
62	0.05741	0.981742	994.81	998.27	1000
63	0.00575	0.988627	1001.73	1004.79	
64	0.00000	0.994671	1007.85	1010.55	
	0.00000	1.000000	1013.25		

A2.9 20CR

The 20CR uses a hybrid sigma–pressure (hybrid σ–p) vertical coordinates after *Simmons and Burridge*, (1981). The pressure on each level is calculated as $p_k = A_k + B_k \times p_{srf}$, where p_{srf} is surface pressure. The following table provides example pressures at layer interfaces (*k*–1/2) and layer midpoints (*k*) for a surface pressure of 1013.25 hPa, from TOA to surface. Pressures at layer midpoints are defined as the average of pressures at layer interfaces.

Model Levels					Pressure Levels
k	$A_{k-1/2}$ (hPa)	$B_{k-1/2}$	$p_{k-1/2}$ (hPa)	p_k (hPa)	p (hPa)
1	0.00000	0.000000	0.00	2.83	
2	5.66898	0.000000	5.67	9.29	10
3	12.90533	0.000000	12.91	17.51	20
4	22.10979	0.000000	22.11	27.94	30
5	33.76516	0.000000	33.77	41.10	
6	48.44036	0.000000	48.44	57.61	50
7	66.78608	0.000000	66.79	78.15	70
8	89.13767	0.000379	89.52	103.47	100
9	113.43654	0.003933	117.42	134.33	150
10	136.71427	0.014326	151.23	171.39	
11	156.13564	0.034950	191.55	215.13	200
12	169.12130	0.068675	238.71	265.66	250
13	173.64658	0.117418	292.62	322.64	300,350
14	169.59994	0.180667	352.66	385.13	400
15	158.12926	0.256084	417.61	451.65	450
16	140.89535	0.340293	485.70	520.25	500,550
17	119.91428	0.429195	554.80	588.72	600
18	97.31807	0.518457	622.64	654.89	650
19	75.08532	0.604055	687.14	716.87	700
20	54.81144	0.682747	746.60	773.25	750
21	37.57142	0.752347	799.89	823.16	800
22	23.89205	0.811785	846.43	866.32	850
23	13.81526	0.860975	886.20	902.86	900
24	7.01453	0.900581	919.53	933.27	
25	2.92577	0.931750	947.02	958.21	950
26	0.86457	0.955872	969.40	978.42	
27	0.11635	0.974402	987.43	994.63	1000
28	0.00009	0.988726	1001.83	1007.54	
	0.00000	1.000000	1013.25		

Major abbreviations and terms

20CR	20th Century Reanalysis (v2 for version 2, v2c for version 2c, and v3 for version 3)
2D-Var	2-dimensional variational assimilation scheme
3D-Var	3-dimensional variational assimilation scheme
3D-FGAT	3-dimensional variational assimilation scheme with FGAT
4D-Var	4-dimensional variational assimilation scheme
ABL	atmospheric boundary layer
ACARS	Aircraft Communications Addressing and Reporting System
ACRE	Atmospheric Circulation Reconstructions over the Earth
AER	Atmospheric and Environmental Research
AERONET	Aerosol Robotic Network
AGCM	atmospheric general circulation model
AHI	Advanced Himawari Imager
AIRS	Atmospheric Infrared Sounder
AMDAR	Aircraft Meteorological Data Relay
AMIP	Atmospheric Model Intercomparison Project
AMSR	Advanced Microwave Scanning Radiometer
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
AMSU	Advanced Microwave Sounding Unit
AMV	atmospheric motion vectors
ANA	“analyzed” state produced prior to IAU for MERRA and MERRA-2
AOD	aerosol optical depth
Aqua	a satellite in NASA’s Earth Observing System (EOS) A-Train constellation
ASCAT	Advanced Scatterometer
ASM	“assimilated” state produced by IAU for MERRA and MERRA-2
ATMS	Advanced Technology Microwave Sounder
ATOVS	Advanced TIROS Operational Vertical Sounder
Aura	a satellite in NASA’s Earth Observing System (EOS) A-Train constellation
AVHRR	Advanced Very High Resolution Radiometer
BAS	British Antarctic Survey
BOM	Bureau of Meteorology (Australia)
BUOY	Surface meteorological observation report from buoys
CAMSiRA	Copernicus Atmosphere Monitoring Service Interim Reanalysis
CCARDS	Comprehensive Aerological Reference Dataset, Core Subset
CCI	Climate Change Initiative (ESA)
CEDA	Centre for Environmental Data Analysis
CERA	a coupled atmosphere–ocean data assimilation system developed by ECMWF
CFC	chlorofluorocarbon
CFS	Climate Forecast System developed at NCEP
CFSR	Climate Forecast System Reanalysis
CFSv2	Climate Forecast System version 2
CHAMP	CHAllenging Minisatellite Payload
CIRES	Cooperative Institute for Research in Environmental Sciences (NOAA and University of Colorado Boulder)
CMA	China Meteorological Administration
CMAP	CPC Merged Analysis of Precipitation
CMIP5	Coupled Model Intercomparison Project Phase 5

CNSA	China National Space Administration
COBE	Centenial in-situ Observation-Based Estimates of variability of SST and marine meteorological variables
COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate
CPC	Climate Prediction Center (NOAA)
CRIEPI	Central Research Institute of Electric Power Industry
CrIS	Cross-track Infrared Sounder
CRTM	Community Radiative Transfer Model
CRUTEM	Climatic Research Unit Air Temperature Anomalies
CTM	chemical transport model
DAO	Data Assimilation Office (NASA); now GMAO
DAS	data assimilation system
DMSP	Defense Meteorological Satellite Program
DOE	Department of Energy
ECMWF	European Centre for Medium-Range Weather Forecasts
EDA	the 10-member "ensemble of data assimilations" produced for ERA5
EMC	Ensemble Modeling Center
EnKF	Ensemble Kalman Filter assimilation scheme
EOS	Earth Observing System of the NASA
ERA-15	ECMWF 15-year reanalysis
ERA-20C	ECMWF 20th century reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA-CLIM	European Reanalysis of Global Climate Observations
ERA-Interim	ECMWF interim reanalysis
ERA5	the fifth major global reanalysis produced by ECMWF
ERA5L	a land surface reanalysis with atmospheric forcing from ERA5
ERS	European Remote Sensing satellite
ESA	European Space Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FGAT	first guess at appropriate time
FGGE	First GARP (Global Atmospheric Research Program) Global Experiment
FORMOSAT	The name given to the Republic of China Satellite (ROCSat) following a public naming competition.
FY-3	FengYun-3 (a series of polar-orbiting satellites launched by the CMA and CNSA)
GAAS	Goddard Aerosol Assimilation System
GARP	Global Atmospheric Research Program
GATE	GARP (Global Atmospheric Research Program) Atlantic Tropical Experiment
GAW	Global Atmosphere Watch
GCM	general circulation model
GEO	geostationary satellites
GEOS	Goddard Earth Observing System Model of the NASA
GFDL	Geophysical Fluid Dynamics Laboratory of the NOAA
GFS	Global Forecast System of the NCEP
GISST	UKMO Global Ice and Sea Surface Temperature dataset
GLATOVS	Goddard Laboratory for Atmospheres TOVS (radiative transfer model)
GLCC	Global Land Cover Characteristics data base
GLDAS	Global Land Data Assimilation System
GMI	GPM Microwave Imager
GMS	Geostationary Meteorological Satellite
GNSS-RO	Global Navigation Satellite System Radio Occultation (see also GPS-RO)
GOCART	Goddard Chemistry, Aerosol, Radiation, and Transport model
GODAS	NCEP Global Ocean Data Assimilation System
GOES	Geostationary Operational Environmental Satellite

GOME	Global Ozone Monitoring Experiment
GPCP	Global Precipitation Climatology Project
GPM	Global Precipitation Measurement mission
GPS-RO	Global Positioning System Radio Occultation (see also GNSS-RO)
GRACE	Gravity Recovery and Climate Experiment
GRIB	General Regularly-distributed Information in Binary form (a file format)
GRIB2	GRIB, Version 2 (a file format)
GRUAN	Global Climate Observing System Reference Upper Air Network
GSI	Gridpoint Statistical Interpolation assimilation scheme
GSICS	Global Space-based Inter-calibration System
GSM	Global Spectral Model of the JMA
GTS	Global Telecommunication System
GWD	gravity wave drag
HadISST	UKMO Hadley Centre Sea Ice and SST dataset
HALOE	Halogen Occultation Experiment
HCFC	hydrochlorofluorocarbon
HDF	Hierarchical Data Format (a file format)
HIRS	High-resolution Infrared Radiation Sounder
HRES	the high-resolution analysis produced for ERA5
IASI	Infrared Atmospheric Sounding Interferometer
IAU	Incremental Analysis Update procedure (or products resulting from that procedure)
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
IFS	Integrated Forecast System of the ECMWF
IGY	International Geophysical Year (July 1957–December 1958)
IPCC	Intergovernmental Panel on Climate Change
IR	infrared
ISPD	International Surface Pressure Databank
JAXA	Japan Aerospace Exploration Agency
JCDAS	JMA Climate Data Assimilation System
JCSDA	Joint Center for Satellite Data Assimilation
JMA	Japan Meteorological Agency
JRA-25	Japanese 25-year Reanalysis
JRA-55	Japanese 55-year Reanalysis
JRA-55AMIP	Japanese 55-year Reanalysis based on AMIP-type simulations
JRA-55C	Japanese 55-year Reanalysis assimilating Conventional observations only
LAI	leaf area index
LCL	lifting condensation level
LEO/GEO	Low Earth Orbit / Geostationary
LIE	Line Islands Experiment
LSM	land surface model
MARS	Meteorological Archival and Retrieval System of the ECMWF
McICA	Monte Carlo Independent Column Approximation
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2
Met Office	see UKMO
METEOSAT	geostationary meteorological satellites operated by EUMETSAT
MetOp	A series of three polar orbiting meteorological satellites operated by the EUMETSAT
MHS	Microwave Humidity Sounder
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MISR	Multiangle Imaging Spectroradiometer
MIT	Massachusetts Institute of Technology

MLS	Microwave Limb Sounder
MODIS	MODerate resolution Imaging Spectroradiometer
MOM	Modular Ocean Model
MRF	Medium Range Forecast Version of the NCEP Global Forecast System
MRI-CCM1	Meteorological Research Institute (JMA) Chemistry Climate Model, version 1
MSU	Microwave Sounding Unit
MTSAT	Multi-functional Transport Satellite
MW	microwave (sounders)
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCDC	National Climatic Data Center of the NOAA
NCEI	National Centers for Environmental Information (NOAA)
NCEP	National Centers for Environmental Prediction (NOAA)
NCEP-DOE R-2	Reanalysis 2 of the NCEP and DOE
NCEP-NCAR R-1	Reanalysis 1 of the NCEP and NCAR
NESDIS	National Environmental Satellite, Data, and Information Service
NetCDF	Network Common Data Form (a file format)
NH	Northern Hemisphere
NIST	National Institute of Standards and Technology
NMC	National Meteorological Center
NOAA	National Oceanic and Atmospheric Administration
NOAA-CIRES 20C	20th Century Reanalysis of the NOAA and CIRES (see also 20CR)
NSIDC	National Snow and Ice Data Center
OI	optimal interpolation
OISST	NOAA Optimum Interpolation Sea Surface Temperature (v2 for version 2)
OMI	Ozone Monitoring Instrument
OSTIA	Operational Sea Surface Temperature and Sea-Ice Analysis
OSU LSM	Oregon State University LSM
PAOBS	Bogus surface pressure data for the Southern Hemisphere produced by the Australian Bureau of Meteorology
PCMDI	Program of Climate Model Diagnosis and Intercomparison
PDF	probability distribution function
PIBAL	Pilot Balloon
QBO	Quasi-Biennial Oscillation
QC	quality control
QuikSCAT	Quick Scatterometer
R1	see NCEP–NCAR R1
R2	see NCEP–DOE R2
RAOBCORE	Radiosonde Observation Correction using Reanalyses
RCP	representative concentration pathway (IPCC)
RDA	Research Data Archive (NCAR)
RH	relative humidity
RICH	Radiosonde Innovation Composite Homogenization
RO	radio occultation
RRTM	Rapid Radiative Transfer Model developed by AER
RRTM-G	Rapid Radiative Transfer Model for application to GCMs developed by AER
RTG	NCEP Real-Time Global sea surface temperature
RTTOV	Radiative Transfer for TOVS
S-RIP	SPARC Reanslysis Intercomparison Project
SBUV	Solar Backscatter Ultraviolet Radiometer
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY

SEVIRI	Spinning Enhanced Visible and Infrared Imager (EUMETSAT)
SH	Southern Hemisphere
SHIP	Surface meteorological observation report from ships
SiB	Simple Biosphere model
SIC	sea ice concentration
SMMR	Scanning Multichannel Microwave Radiometer
SNDR	Sounder (for radiance measurements by the GOES 8 to 12)
SNO	Simultaneous Nadir Overpass method
SOLARIS-HEPPA	Solar Influences for SPARC–High Energy Partical Precipitation in the Atmosphere
SPARC	Stratosphere-troposphere Processes And their Role in Climate
SSI	Spectral Statistical Interpolation (an assimilation scheme)
SSM/I or SSMI	Special Sensor Microwave Imager
SSMIS	Special Sensor Microwave Imager Sounder
SST	sea surface temperature
SSU	Stratospheric Sounding Unit
SYNOP	Surface meteorological observation report from manned and automated weather stations
TCWV	total column water vapour
TD	tape deck ("TD" is a name of a rawinsonde dataset. For example, TD54 is a dataset of mandatory level data from rawinsondes during 1946-1972 prepared by the USAF. See **http://rda.ucar.edu/docs/papers-scanned/pdf/rj0187.pdf** (accessed 29 May 2015).
Terra	a satellite in NASA's Earth Observing System (EOS).
TerraSAR-X	a German satellite with a phased array Synthetic Aperture Radar (SAR) antenna at the X-band wavelength
TIM	Total Irradiance Monitor
TIROS	Television Infrared Observation Satellite
TIROS-N	Television InfraRed Operational Satellite - Next-generation
TMI	TRMM Microwave Imager
TOA	top of atmosphere
TOMS	Total Ozone Mapping Spectrometer
TOVS	TIROS Operational Vertical Sounder
TRMM	Tropical Rainfall Measuring Mission
TSI	total solar irradiance
UARS	Upper Atmosphere Research Satellite
UKMO	United Kingdom Meteorological Office (or Met Office)
USAF	U.S. Air Force
USCNTRL	U.S. controlled oceanweather stations
USGS	U.S. Geological Survey
UTC	Universal Coordinated Time
VTPR	Vertical Temperature Profile Radiometer
WMO	World Meteorological Organization

Chapter 3: Overview of Temperature and Winds

Chapter lead authors

Craig Long	National Oceanic and Atmospheric Administration *(retired)*	USA
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan

Co-authors

Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Daniel Mitchell	University of Bristol	UK
Corwin Wright	University of Bath	UK

Abstract. Two of the most basic parameters generated from a reanalysis are temperature and winds. Temperatures in the reanalyses are derived from conventional (surface and balloon), aircraft, and satellite observations. Winds are observed by conventional systems, cloud tracked, and derived from height fields, which are in turn derived from the vertical temperature structure. In this chapter we evaluate as part of the SPARC Reanalysis Intercomparison Project (S-RIP) the temperature and wind structure of all the recent and past reanalyses. This evaluation is mainly among the reanalyses themselves, but comparisons against independent observations, such as HIRDLS, MLS, COSMIC, ozonesonde, and rocketsonde temperatures are also presented. This evaluation uses monthly mean and 2.5° zonal mean data sets and spans the satellite era from 1979 – 2014. There is very good agreement in temperature seasonally and latitudinally among the more recent reanalyses (CFSR/CFSv2, MERRA, ERA-Interim, JRA-55, and MERRA-2) between the surface and 10 hPa. At lower pressures there is increased variance among these reanalyses that changes with season and latitude. This variance also changes during the time span of these reanalyses with greater variance during the TOVS period (1979 – 1998) and less variance afterward in the ATOVS period (1999 – 2014). There is a distinct change in the temperature structure in the middle and upper stratosphere during this transition from TOVS to ATOVS systems. Consult *Chapter 2 Section 2.4.3.2* (Observational Data/Satellite Data) for more information about the TOVS and ATOVS suite of instruments and usage by the various reanalyses. Zonal winds for the entire period have lower variance among the reanalyses than the temperatures and this lower variance extends to lower pressures than the temperatures. The temperatures and winds throughout the stratosphere in the older reanayses (NCEP-NCAR R1, NCEP-DOE R2, ERA-40, and JRA-25) show significant differences from a Reanalysis Ensemble Mean for the same time periods than in the more recent reanalyses, both during the TOVS and ATOVS periods. The transition of temperatures from the TOVS to ATOVS periods continues to be an issue even for the more recent reanalyses. All reanalyses to date have issues analysing the quasi-biennial oscillation (QBO) winds. Comparisons with Singapore QBO winds show disagreement in the amplitude of the westerly and easterly anomalies. The disagreement with Singapore winds improves with the transition from TOVS to ATOVS observations. Aura HIRDLS and Aura MLS temperatures have similar bias characteristics when compared with a reanalysis ensemble mean (MERRA, ERA-Interim, and JRA-55). There is good agreement among the NOAA TLS, SSU1, and SSU2 Climate Data Records and layer mean temperatures from the more recent reanalyses. Caution is advised when using reanalysis temperatures for trend detection and anomalies from a long climatology period as the quality and character of reanalyses have changed over time.

Long *et al.* (2017) have published a shortened version of this chapter.

Contents

3.1 Introduction 83
3.2 Improvements from older reanalyses to newer versions....84
3.3 Reanalysis global mean temperature anomaly variability 84
3.4 Reanalysis ensemble mean (REM) 86
3.4.1 Methodology 86
3.4.2 Climatology of the REM 87
3.4.2.1 Temperature....87
3.4.2.2 Zonal wind....88
3.4.2.3 Meridional and vertical winds....89
3.4.3 Agreement among the REM members....92
3.4.3.1 Temperature....92
3.4.3.2 Zonal winds 93
3.5 Intercomparisons of the reanalyses....95
3.5.1 Temperature....96
3.5.1.1 SH polar latitudes....96
3.5.1.2 NH polar latitudes....97
3.5.1.3 Equatorial latitudes....97
3.5.1.4 NH and SH mid-latitudes....99
3.5.2 Zonal wind....99
3.5.2.1 SH polar latitudes....99
3.5.2.2 NH polar latitudes....99
3.5.2.3 Equatorial latitudes....99
3.5.2.4 NH and SH mid-latitudes....101
3.5.2.5 Comparisons with Singapore QBO winds 101
3.6 Amplitude of polar annual temperature cycle....103
3.7 Comparisons with satellite temperature observations....104
3.7.1 HIRDLS and MLS temperatures 105
3.7.2 Comparisons with COSMIC temperature observations 106
3.7.3 Atmospheric layer temperature anomalies 107
3.7.4 Ringing of SH polar latitude temperature differences 111
3.8 Comparisons against other observations 111
3.8.1 Ozonesonde temperatures 111
3.8.2 Long-duration balloon observations 112
3.8.3 Rocketsonde observations 114
3.9 Effects of volcanic eruptions and other natural variabilities 115
3.10 Summary and conclusions....114
3.11 Key findings and recommendations....116
References 117
Major abbreviations and terms 120

3.1 Introduction

Reanalyses are used in many ways, including as initial conditions for historical model runs, developing climatologies, comparison with experimental models, and the examination of atmospheric features or conditions over long periods of time. This chapter mainly evaluates eight reanalysis data sets: NCEP-NCAR Reanalysis 1 (*Kistler et al.*, 2001; referred to hereafter as "*R1*"; *Kalnay et al.*, 1996), ERA-40 (*Uppala et al.*, 2005), JRA-25 (*Onogi et al.*, 2007), NCEP/CFSR (*Saha et al.*, 2010), ERA-Interim (*Dee et al.*, 2011; referred to hereafter as ERA-I), MERRA (*Rienecker et al.*, 2011), JRA-55 (*Kobayashi et al.*, 2015), and MERRA-2 (*Gelaro et al.*, 2017; *GMAO*, 2015), with some notes on NCEP-DOE Reanalysis 2 (*Kanamitsu et al.*, 2002) (referred to hereafter as *R2*) and 20CR (*Compo et al.*, 2011). See *Chapter 2* for more information about these reanalyses. The ERA-15 (*Gibson et al.*, 1997) is not included in this intercomparison due to its short period and subsequent replacement by ERA-40. When a reanalysis product is chosen for use in a study or comparison, the choice is made based upon several factors such as newness of the reanalysis systems, span of time evaluated, horizontal and vertical resolution, top layer, and observational data assimilated. In this chapter, we present an intercomparison of these 10 reanalyses focusing mainly upon their temperature and zonal wind fields. The five more recent reanalyses (CFSR/CFSv2, MERRA, ERA-I, JRA-55, and MERRA-2) are the primary focus and we concentrate on how these reanalyses intercompare in the upper troposphere and entire stratosphere.

Intercomparisons of middle atmosphere winds and temperatures using reanalyses have been performed since the very first reanalyses were generated in the late 1990s. *Pawson and Fiorino* (1998a, b; 1999) were the first to evaluate reanalyses winds and temperatures by comparing *R1* and ERA-15 analysis of the tropics before and after satellite data were used in the reanalyses. *Randel et al.* (2004) intercompared wind and temperature climatologies from *R1*, ERA-15, and ERA-40 along with meteorological centre analyses. *R1* and the ERA-40 have been used by thousands of researchers for tropospheric studies. Notable middle atmosphere studies evaluating *R1* and ERA-40 winds and temperatures include the following. *Manney et al.* (2005) used these two reanalyses along with other analyses to evaluate their ability to capture the unique 2002 Antarctic winter, while *Charlton and Polvani* (2007) intercompared the two for detecting Northern Hemispheric sudden stratospheric warmings (SSWs). *Martineau and Son* (2010) used temperature and wind fields from *R1*, *R2*, JRA-25, ERA-I, and MERRA to compare their depiction of stratospheric vortex weakening and intensification events against GPSRO temperature data. *Simmons et al.* (2014) intercompared the ERA-I, MERRA, and JRA-55 stratospheric temperature analyses over the 1979 – 2012 period showing where and when they agreed and disagreed and the reasons why they did so. They also pointed out the difficulties of the transition from the TOVS to ATOVS observations, most notably in the upper stratosphere and lower mesosphere. *Lawrence et al.* (2015) used polar processing diagnostics to compare the ERA-I and MERRA. They noted good agreement in the diagnostics after 2002, but cautioned that the choice of one over the other could influence the results of polar processing studies. *Miyazaki et al.* (2016) intercompared six reanalyses (*R1*, ERA-40, JRA-25, CFSR/CFSv2, ERA-I, JRA-55) to study the mean meridional circulation in the stratosphere and eddy mixing and their implications upon the strength of the Brewer–Dobson circulation. *Fujiwara et al.* (2015) used nine reanalyses (JRA-55, MERRA, ERA-I, CFSR/CFSv2, JRA-25, ERA-40, *R1*, *R2*, and 20CR) to examine their stratospheric temperature response to the eruptions of Mount Agung (1963), El Chichón (1982), and Mount Pinatubo (1991). *Mitchell et al.* (2015) performed a multiple linear regression analysis on the same nine reanalyses to test the robustness of their variability. *Martineau et al.* (2016) intercompare eight reanalyses (ERA-40, ERA-I, *R1*, *R2*, CFSR/CFSv2, JRA-25, JRA-55, and MERRA) for dynamical consistency of wintertime stratospheric polar vortex variability. *Kawatani et al.* (2016) compare the representation of the monthly mean zonal wind in the equatorial stratosphere with a focus on the quasi-biennial oscillation (QBO; *Baldwin et al.*, 2001) among nine reanalyses (*R1*, *R2*, CFSR/CFSv2, ERA-40, ERA-I, JRA-25, JRA-55, MERRA, and MERRA-2).

The report by the SPARC Reference Climatology Group (*SPARC*, 2002) and the subsequent journal article by *Randel et al.* (2004) were in response to the need to compare and evaluate the then-existing middle atmosphere climatologies that were housed and made readily available to the research community at the SPARC Data Centre. Both reports provide an intercomparison of eight middle atmosphere climatologies: UK Met Office data assimilation, NOAA Climate Prediction Center objective analysis, UK Met Office objective analysis using TOVS data, the Free University of Berlin Northern Hemisphere subjective analysis, CIRA86 (COSPAR International Reference Atmosphere, 1986), *R1*, ERA-15, and ERA-40. This intercomparison was mostly based upon analyses rather than reanalyses, as only the *R1*, ERA-15, and ERA-40 reanalyses were available at that time. Notable differences were found among analyses for temperatures near the tropical tropopause and polar lower stratosphere and zonal winds throughout the tropics. Comparisons of historical reference atmosphere and rocketsonde temperature observations with the more recent global analyses showed the influence of decadal-scale cooling of the stratosphere. Detailed comparisons of the tropical semi-annual oscillation (SAO) and QBO showed large differences in amplitude among analyses; the more recent data assimilation schemes showed better agreement with equatorial radiosonde, rocket, and satellite data (*e.g. Baldwin and Gray*, 2005).

This chapter is an extended version of the paper by *Long et al.* (2017). In this chapter we show the results from the eight "full-input" (*Chapter 2*) reanalyses, which are systems that assimilate surface and upper-air conventional and satellite data (*i.e.* MERRA-2, MERRA, ERA-I, JRA-55, CFSR/CFSv2, JRA-25, ERA-40, *R1*), though we will show one figure for 20CR, which is one of the "surface-input" reanalyses. We will concentrate only on the satellite era period of 1979 to 2014.

Table 3.1: *Information about NCEP, JMA, ECMWF, and GMAO earlier and later reanalyses pertinent to the stratosphere. Information includes the model version, horizontal resolution, number of model levels, model top pressure, and radiative transfer model (RTM) used for assimilating satellite radiances. See Chapter 2 for more details.*

	Model Version		Horizontal Resolution	Model Levels	Model Top Level	RTM
R1/R2	NCEP MRF	(1995/98)	T62: 1.875°	28 (σ)	3 hPa	Temperature retrievals
CFSR CFSv2	NCEP CFS	(2007) (2011)	T382: 0.3125° T574: 0.2045°	64 (hybrid σ-p)	~0.266 hPa	CRTM
ERA-40	IFS Cycle 23r4	(2001)	TL159: ~125 km	60 (hybrid σ-p)	0.1 hPa	RTTOVS-5
ERA-I	IFSCycle 31r2	(2007)	TL255: ~79 km	60 (hybrid σ-p)	0.1 hPa	RTTOVS-7
ERA5	IFSCycle 41r2	(2016)	TL639: ~31 km	137 (hybrid σ-p)	0.01 hPa	RTTOVS-11
JRA-25	JMA GSM	(2004)	T106: 1.125°	40 (hybrid σ-p)	0.4 hPa	RTTOVS-6 TOVS period RTTOVS-7 ATOVS period
JRA-55	JMA GSM	(2008)	TL 319: ~55 km	60 (hybrid σ-p)	0.1 hPa	RTTOVS-9
MERRA	GEOS 5.0.2	(2009)	0.5° lat x 0.667° lon	72 (hybrid σ-p)	0.01 hPa	GLATOVS for SSU; CRTM for others
MERRA-2	GEOS 5.12.4	(2015)	C180: ~50 km	72 (hybrid σ-p)	0.01 hPa	CRTM

ERA5 results are shown only in **Figure 3.1**. Several of the reanalyses do not cover the entire span of the later period (*e.g.* ERA-40 ends in August 2002, 20CR ends in December 2012 (for its version 2), and JRA-25 ends in January 2014). The *R2* is an updated version of the *R1*. Almost all of the changes and enhancements incorporated into the *R2* were surface or boundary layer oriented. The only possible change to the stratosphere would be due to a change to a newer ozone climatology (*Fujiwara et al.*, 2017). As a result, preliminary comparisons of *R1* and *R2* show very minor differences in temperatures and winds above the boundary layer. Therefore, we will not show *R2* comparisons, but one can expect all *R2* qualities to be nearly exactly the same as *R1*. All of the reanalyses except for the CFSR/CFSv2 used the same forecast model and assimilation scheme throughout their time span. In 2010 the CFSR/CFSv2 had an undocumented update to its GSI assimilation scheme. Another change to the CFSR/CFSv2 occurred in 2011 with the implementation of the version 2 Climate Forecast System (CFSv2; *Saha et al.*, 2014), in which the resolution, forecast model, and assimilation scheme were all upgraded. *Chapter 2* and *Fujiwara et al.* (2017) distinguish this latter analysis as CFSv2 or CDAS-T574.

The rest of this chapter will be organized as follows: *Section 3.2* presents a summary of changes and improvements from each reanalysis centre's earlier to later versions. *Section 3.3* presents and discusses temperature variability with time of the reanalyses. *Section 3.4* presents the methodology used to compare the various reanalyses, the creation of a reanalysis ensemble mean (REM), and the ensemble mean attributes and variability with time. *Section 3.5* presents the differences in the temperatures and winds in individual reanalyses from the REM. *Section 3.6* examines the seasonal temperature amplitude of the reanalyses in the polar latitudes. *Section 3.7* discusses the results of comparisons with satellite observations that are not assimilated in the reanalyses by showing specific data analyses. *Section 3.8* shows comparisons against other types of non-satellite observations. *Section 3.9* discusses the effects of volcanic eruptions on reanalysis temperatures and winds. *Section 3.10* provides summaries and main conclusions.

We characterize the stratosphere into altitude ranges using the following generalizations: "upper" for 1 hPa to 5 hPa, "middle" for 7 hPa to 30 hPa, and "lower" for 50 hPa to 100 hPa.

3.2 Improvements from older reanalyses to newer versions

Chapter 2 of this report and *Fujiwara et al.* (2017) provide many details of each reanalysis, such as model characteristics, physical parameterizations used, observations assimilated, execution stream characteristics, and assimilation strategies. The most recent reanalyses are later generations of earlier versions (MERRA and MERRA-2, ERA40 and ERA-Interim, JRA-55 and JRA-25, CFSR/CFSv2, and *R1* and *R2*). Using information contained in *Chapter 2*, we will highlight what we consider the major improvements and changes from the earlier version to the more recent version. Pertinent to the stratosphere, we present in **Table 3.1** a summary of the earlier and later reanalysis models used, model resolution, top pressure level, and radiative transfer model (RTM) used for assimilating satellite radiances. Several reanalyses improved their model horizontal and vertical resolution between versions. All of the later versions used a more recent version of the RTM. Explanations for the various labelling of horizontal resolution can be found in *Chapter 2*.

3.3 Reanalysis global mean temperature anomaly variability

The 1979 - 2014 period includes the assimilation of satellite observations in addition to the assimilation of conventional (surface, aircraft, and balloon) observations (see *Chapter 2* for details). During this period, there are multiple transitions, additions, and removals of satellites and instruments observing the atmosphere. The calibration and quality control of the observations from these satellite instruments in many instances have improved over time from the earlier reanalysis systems to the more current reanalysis systems.

The radiative transfer models used in the forecast models have also improved over time. Reanalysis centres devote major efforts to minimizing the transition from one satellite or observing system to the next (*e.g.* TOVS to ATOVS in 1998; see *Chapter 2*). However, the forecast models used by the reanalysis centres have their own biases throughout the atmosphere. If and how well the bias correction is performed will also dictate how the reanalysis uses these observations. Additionally, most reanalyses are not run as one stream, but rather it is more efficient timewise and computationally for the reanalysis to be broken up into multiple streams with overlap periods of at least 1 or more years (see *Section 2.5* of *Chapter 2*). These overlap periods are intended to allow the new stream to spin up sufficiently to ensure minimal discontinuity when the older stream ends. Because of these factors, it will be shown that the more recent reanalyses have fewer discontinuities at different times throughout this data record than older reanalyses.

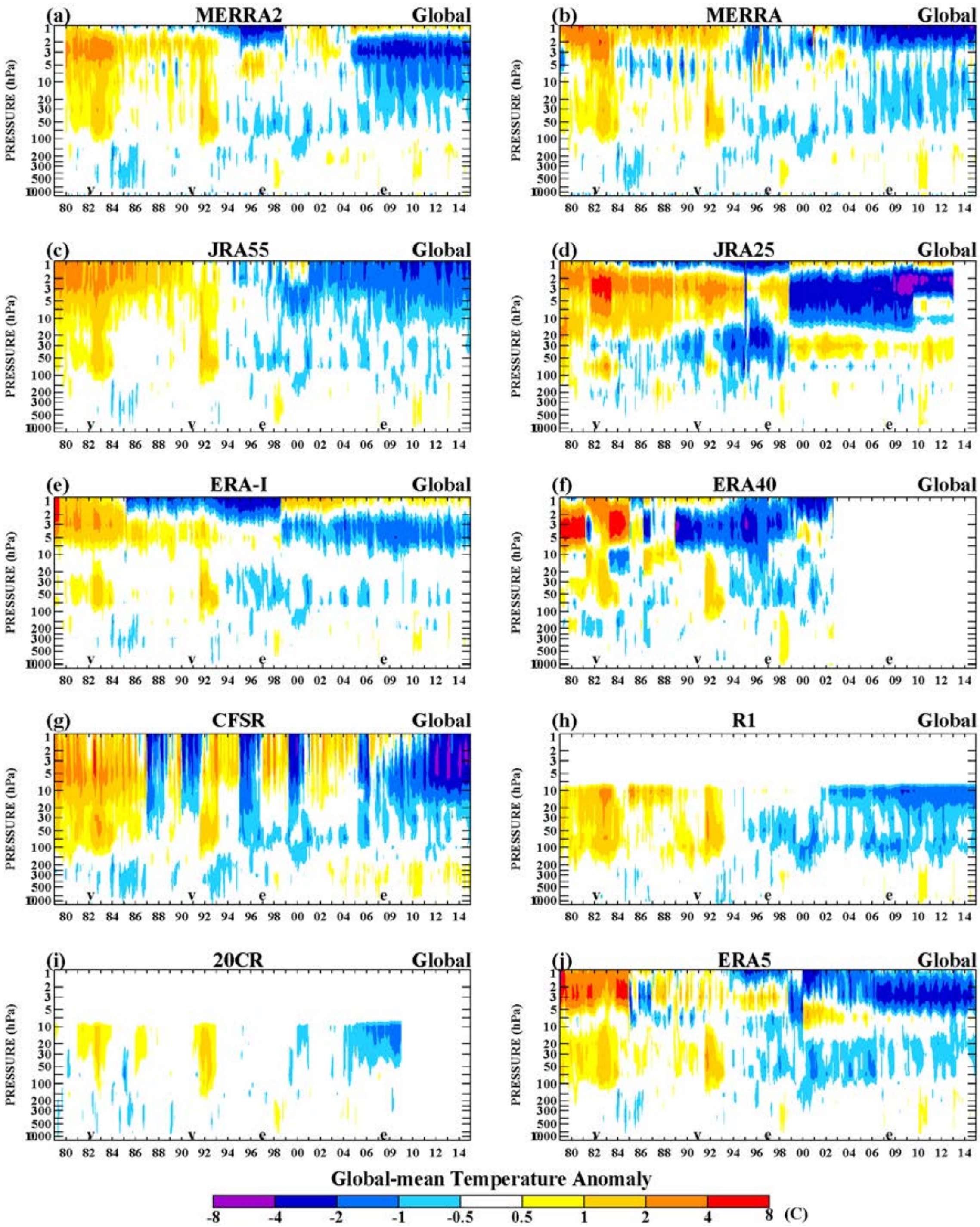


Figure 3.1: *Pressure versus time plots of the global mean temperature anomalies (K) of reanalyses. The anomalies are from the monthly climatology of each reanalysis. The reanalyses shown are (a) MERRA-2, (b) MERRA, (c) JRA-55, (d) JRA-25, (e) ERA-Interim, (f) ERA-40, (g) CFSR/CFSv2, (h) R1, (i) 20CR, and (j) ERA5. Note that R1 and 20CR do not provide analyses above 10 hPa; "v" and "e" denote the occurrence of volcanos and El Niños. Updated from Long et al. (2017).*

To illustrate how well the various reanalyses were able to transition between satellites and other data sources, **Figure 3.1** presents time series for each reanalysis of the global mean temperature anomalies from their own long-term (1979–2014) monthly means. In all of the time series plots, several climatic features are evident: the tropospheric warming during the 1998 and 2010 El Niño events (located on the time axis with an "e") and the lower and middle stratospheric warming associated with the El Chichón (1982) and Mount Pinatubo (1991) volcanic eruptions (located on the time axis with a "v"). However, the older reanalyses (ERA-40 and JRA-25) show several distinct discontinuities in the stratosphere. The ERA-40, which was the first reanalysis to assimilate SSU radiances, shows discontinuities during several changes in the NOAA polar satellites with the SSU instrument in the early 1980s. The ERA-40 assimilated both SSU and AMSU-A radiances from the end of 1998 through 2002 (*Uppala et al.*, 2005). The JRA-25 shows smaller discontinuities in the 1980s but has an abrupt change in 1998 coincident with the immediate transition from TOVS (SSU, MSU) to the ATOVS (AMSU) observing systems. The bias correction schemes for the TOVS and ATOVS radiances were also different. Consult *Chapter 2 Section 2.4.2* (Quality Control Procedures) to learn more about the various reanalyses' bias correction practices. The combination of both resulted in large discontinuity in the stratosphere (*Onogi et al.*, 2007). Of the five more recent reanalyses, the CFSR/CFSv2 shows multiple discontinuities in the upper and middle stratosphere. This is because the CFSR is made up of six streams (end years: 1986, 1989, 1994, 1999, 2005, 2009) and also because it corrects the biases in the SSU channel 3 observations with a forecast model that has a noted warm bias in the upper stratosphere. After 1998 the CFSR/CFSv2 only used the AMSU-A radiances (it did not assimilate channel 14) and just monitored the SSU channels (*Saha et al.*, 2010). The ERA-I shows two distinct discontinuities: in 1985 from the transition from NOAA-7 SSU to NOAA-9 SSU and in August 1998 when ATOVS observing systems began to be assimilated. ERA-I assimilated both SSU and AMSU-A radiances until 2005. Channel 3 of the SSU prior to August 1998 and AMSU-A channel 14 were not bias corrected. After August 1998 the SSU channel 3 radiances were bias corrected (*Simmons et al.*, 2014). MERRA merged the SSU and AMSU observations over a period of time. The version of the CRTM (*Han et al.*, 2006) that MERRA used for other satellite radiances was not able to work with the SSU radiances, and as an alternative the GLATOVS (*Susskind et al.*, 1983) was used. The latter was not updated with the necessary adjustments to the channels due to pressure cell leaks and changes in the stratospheric CO_2 concentration (*Gelaro et al.*, 2017). MERRA immediately stopped using the SSU channel 3 in October 1998 but continued to assimilate channels 1 and 2 through 2005. JRA-55 also merged the SSU and AMSU observations, but for a shorter overlap period of 1 year, and bias corrected all the SSU and AMSU-A channels (*Kobayashi et al.*, 2015). MERRA-2 shows a discontinuity in 1995 from the transition from NOAA-11 to NOAA-14 SSU channel 3 radiances. A second discontinuity occurs when MERRA-2 immediately transitions from SSU and MSU to the AMSU in October 1998. A third discontinuity occurs when it begins using Aura MLS observations in August 2004. Just as with MERRA, MERRA-2 did not bias correct SSU channel 3 and AMSU-A channel 14. MLS temperatures were used to remove a bias in the upper stratosphere and to sharpen the stratopause (*Gelaro et al.*, 2017). *R1*, *R2*, and the 20CR reanalyses only extend up to 10 hPa due to their fewer model layers, so the upper stratosphere is not analysed. *R1* and *R2* use NESDIS-derived temperature retrievals, which minimized satellite transitions. The 20CR is shown as an example that assimilated only surface-based observations. Therefore, it shows no discontinuities, but its forecast model included the volcanic aerosols and the historical changes in carbon dioxide to produce inter-annual variations in the stratosphere (see *Chapter 2* and *Fujiwara et al.*, 2017, for more details).

The timing and degree of these discontinuities will play a role in how well the various reanalyses compare with each other over time. Difficulties associated with assimilating the SSU observations due to their CO_2 pressure-modulated cells slowly leaking, and the changing of atmospheric CO_2 impaired the earlier reanalyses (ERA-40, JRA-25, MERRA). The more recent reanalyses should agree more closely with each other after 1998 because there are fewer issues assimilating the ATOVS, AIRS, and GPSRO observations (MERRA did not assimilate GPSRO data.)

The latest reanalysis (ERA5) exhibits several notable discontinuities, primarily in the middle and upper stratosphere. Similar to the ERA-I, ERA5 has a temperature discontinuity in 1985 from the transition from NOAA-7 SSU to NOAA-9 SSU and in August 1998 resulting from the transition from TOVS to ATOVS. *Simmons et al.* (2020) discusses these discontinuities and the apparent cold bias in the lower stratosphere beginning in 2000. ECMWF decided to rerun the period between 2000 and 2006 making several corrections. This corrected sub-reanalysis is called ERA5.1, which is now available.

Because of these discontinuities and transitions discussed above, reanalyses should be viewed very carefully for use in trend analysis and trend detection, especially in the middle and upper stratosphere.

3.4 Reanalysis ensemble mean (REM)

3.4.1 Methodology

No one reanalysis is the de facto standard for all variables and processes. Consequently a reanalysis ensemble mean (REM) of three of the more recent reanalyses (MERRA, ERA-I, and JRA-55) will be used as the reference from which differences and anomalies will be determined. The CFSR/CFSv2 is excluded from the REM primarily because of the stream-change impacts upon the temperature structure in the middle and upper stratosphere. MERRA-2 is not included in the REM because it had just become available at the time of the preparation of this chapter and does not include 1979.

The data sets used to perform the intercomparisons are monthly mean zonal means at a 2.5° resolution. Standard post-processed pressure levels are used (1000, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, 1 hPa). The focus time period of this intercomparison is 1979 through 2014. The current WMO 30-year climatology period (1981 - 2010) will be the base period of the climatology used. It should be noted that most reanalyses, with the exception of MERRA and MERRA-2, provide data below the surface for some regions (*e.g.,* at 1000 hPa under Antarctica and the Tibetan Plateau). These data are calculated via vertical extrapolation. When the REM is created, re-gridded zonal means are first calculated for each reanalysis, and then the three data sets are averaged where valid data exist. Since most of the latitude zones poleward of 60°S are part of the Antarctic land mass with surface elevations reaching 3 km, pressures higher than 700 hPa have invalid data and hence are not analysed.

3.4.2 Climatology of the REM

3.4.2.1 Temperature

The seasonal variation in the REM temperature monthly means and their inter-annual variability as standard deviation (SD) in three different zonal regions (60° - 90°N, 10°S - 10°N, and 90° - 60°S) are shown in **Figure 3.2**. It is of note that at polar latitudes the lowest temperatures occur in the upper stratosphere in November (for the Northern Hemisphere, NH) and May (for the Southern Hemisphere, SH) and descend with time such that the lowest temperatures in the lower stratosphere do not occur until January in the NH and September in the SH. Thus, when lower stratospheric temperatures are reaching a minimum, upper stratospheric temperatures are already increasing.

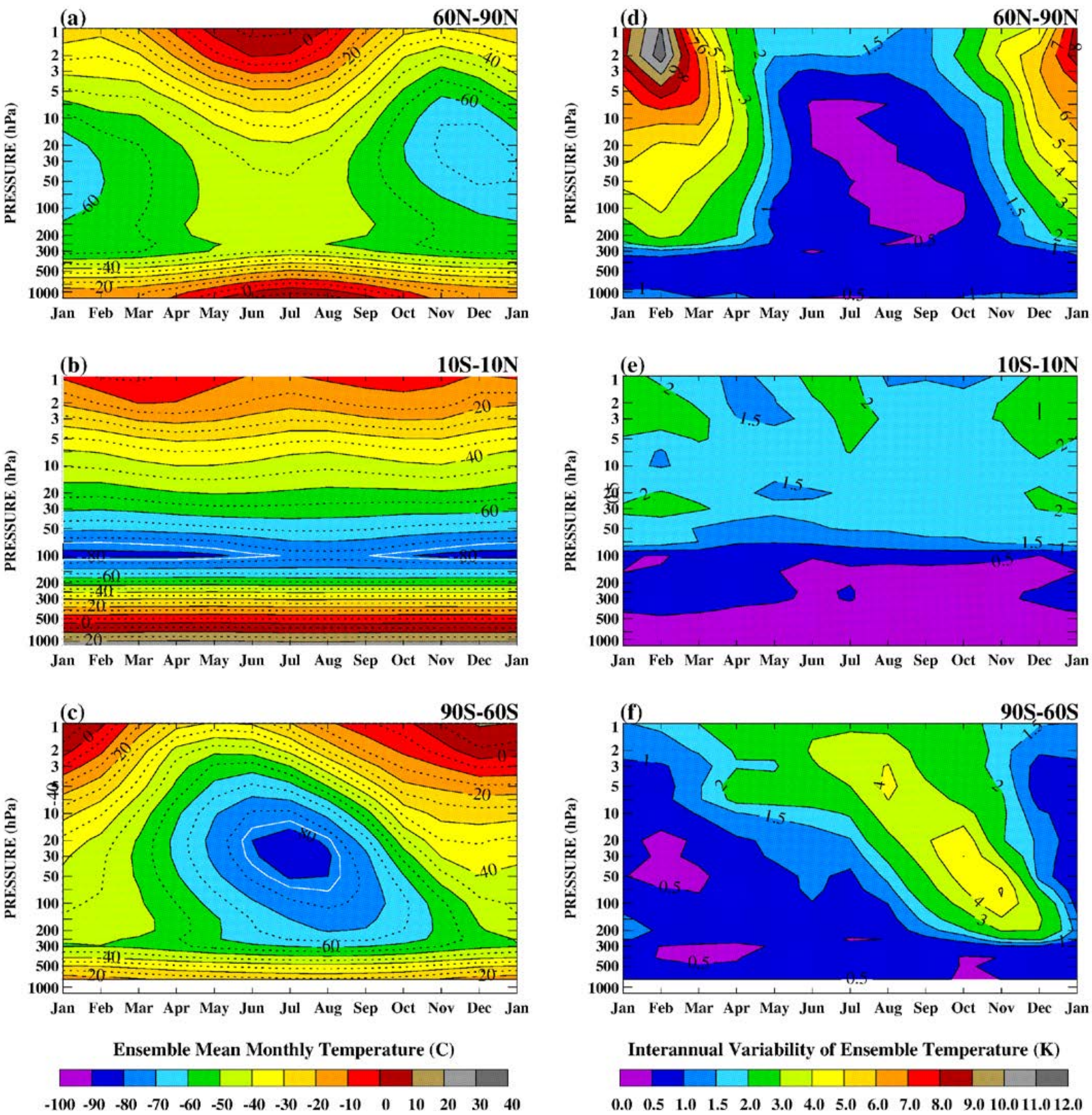


Figure 3.2: *Annual variation in the REM temperature monthly means (°C) and their Standard Deviation (K) in three different zonal regions: 60° – 90° N (a, b), 10° S–10° N (c, d), and 90° – 60° S (e, f). Note in c and f contours cease at 700 mb due to Antarctic surface. Reproduced from Long et al. (2017).*

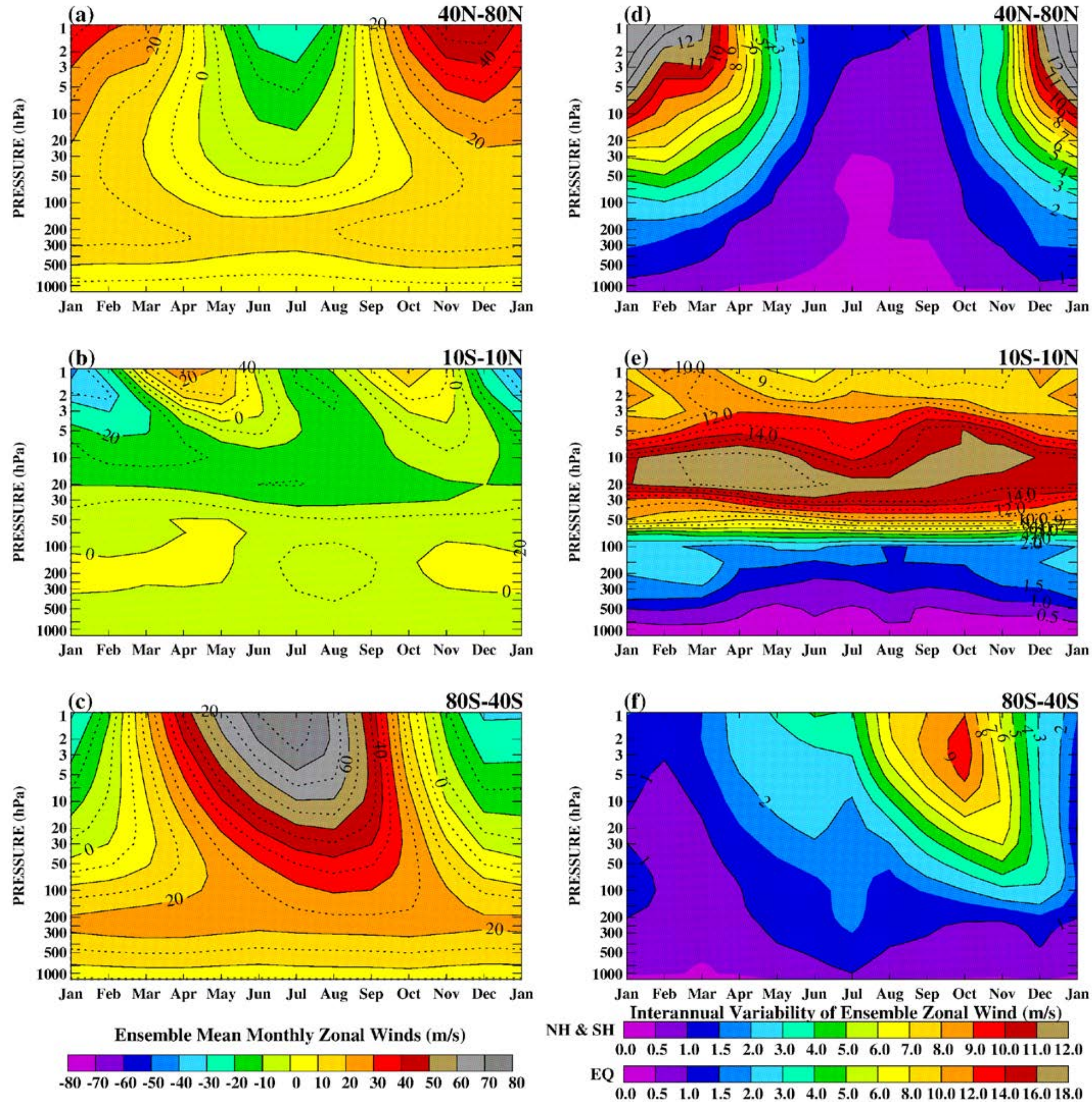


Figure 3.3: *Annual variation in the REM zonal wind monthly means (m s^{-1}) and their SD (m s^{-1}) in three different zonal regions: 40° – 80° N (a, b), 10° S – 10° N (c, d), and 80° – 40° S (e, f). Reproduced from Long et al. (2017).*

The upper stratosphere polar circulation is well defined prior to solstice shutting down any meridional advection of heat into the polar region. Consequently, radiative cooling drives the temperatures to their lowest values prior to solstice. The lowest temperatures occur at about 30 hPa in both polar regions. However, the lowest SH polar temperatures are more than 15 K colder than the lowest NH polar temperature. The interannual variability graphs show that the greatest variability in the NH temperatures is in the upper stratosphere in February when wave activity is most pronounced. In the SH the greatest variability occurs in October and November, associated with the winter to spring transition from low to high temperatures when wave activity becomes significant in that hemisphere. This variability is associated with how quickly that transition occurs. In some years the circulation over Antarctica is very zonal and stable, which prolongs the period of low temperatures in the polar latitudes. In other years there may be greater wave activity transporting heat from the extra-tropics into the polar latitudes, thus shortening the period of low temperatures. In the tropics, the variability is much smaller than in the polar regions but is associated with the phase of the SAO and the QBO in the upper and middle stratosphere, respectively.

3.4.2.2 Zonal wind

The seasonal variation in the REM zonal wind monthly means and their inter-annual variability in three different zonal regions (40° – 80° N, 10° S – 10° N, and 80° – 40° S) are shown in **Figure 3.3**. In the NH polar jet region (40° - 80° N) the maximum winds occur in the upper stratosphere in November and December, and the greatest variability occurs from December through March.

In the SH polar jet region (80°–40°S) wintertime westerlies are about 30 ms^{-1} stronger than the wintertime NH westerlies. These stronger westerlies are due to the much weaker disruption of the polar vortex by the vertically propagating planetary-scale waves and the stronger temperature gradients. Similar to the temperature variability, the variability in the SH polar night jet between May and August is not as great as in the NH polar jet. The SH zonal wind variability increases during the final warming and transition from westerlies to easterlies as wave activity increases from August through November.

In the tropical upper stratosphere, there is a strong semi-annual oscillation (SAO; *Ray et al.*, 1998; *Smith et al.*, 2017, 2020) with maximum westerlies of up to 20 ms^{-1} at equinox and intervening easterlies during the solstice periods. There is a marked asymmetry in the amplitude of the easterly SAO phase, with amplitudes of –40 ms^{-1} to –50 ms^{-1} in the easterlies in December to February but only –20 ms^{-1} to –30 ms^{-1} in July–September.

The easterly SAO phase is believed to result from the advection of easterlies from the summer hemisphere by the Brewer–Dobson circulation (*Gray and Pyle*, 1987), and this asymmetry is consistent with the much stronger circulation in December to February associated with greater wave activity in the NH winter. In the equatorial mid-stratosphere where the QBO dominates, the climatological winds in the tropical middle stratosphere have mean easterlies of –5 ms^{-1} to –10 ms^{-1}. Because of the quasi-biennial nature of the winds, the inter-annual variability is very large, peaking between 10 and 20 hPa. The SAO wind transition in the upper stratosphere also shows a high amount of inter-annual variability.

Please consult *Chapter 11*: Upper Stratosphere Lower Mesosphere for a detailed analysis of the SAO.

3.4.2.3 Meridional and vertical winds

The zonal-mean meridional and vertical winds will be discussed into two layers of the atmosphere: the troposphere and the stratosphere. In the troposphere, the zonal mean vertical and meridional winds combine to form the Hadley, Ferrel, and Polar circulation cells. In the stratosphere, the vertical winds are much smaller compared to those in the troposphere due to the stable nature of the stratosphere. In **Figure 3.4** we present the Eulerian mean meridional and vertical winds for January and July 30-year climatological conditions in the stratosphere and troposphere. Seasonally, the dominant Hadley circulation exists in the winter hemisphere, although the centre of upwelling is in the summer hemisphere. **Figure 3.4b** shows that in January the greatest upwelling occurring at 10°S and greatest downwelling at 20°N. Completing this circulation are strong northerly meridional winds at 10°N and 200 hPa and southerly winds at 10°N between the surface and 850 hPa. A weaker SH Hadley circulation extends from 10°S to 40°S.

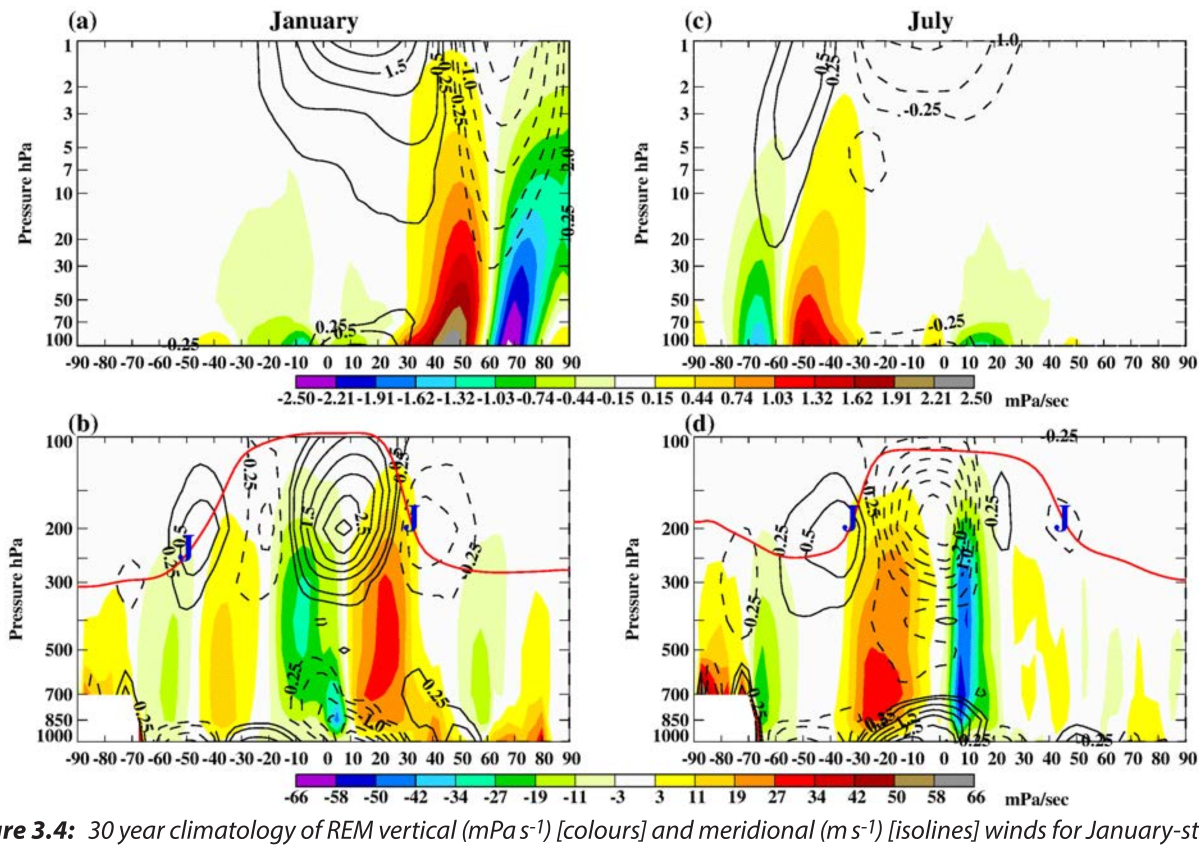


Figure 3.4: *30 year climatology of REM vertical (mPa s^{-1}) [colours] and meridional (m s^{-1}) [isolines] winds for January-stratosphere (a), July-stratosphere (b), January-troposphere(c), and July troposphere (d). REM tropopause (solid red line) is shown in (b) and (d). Subtropical jet central location is denoted with a 'J'. Note: diagnostics presented are Eulerian means.*

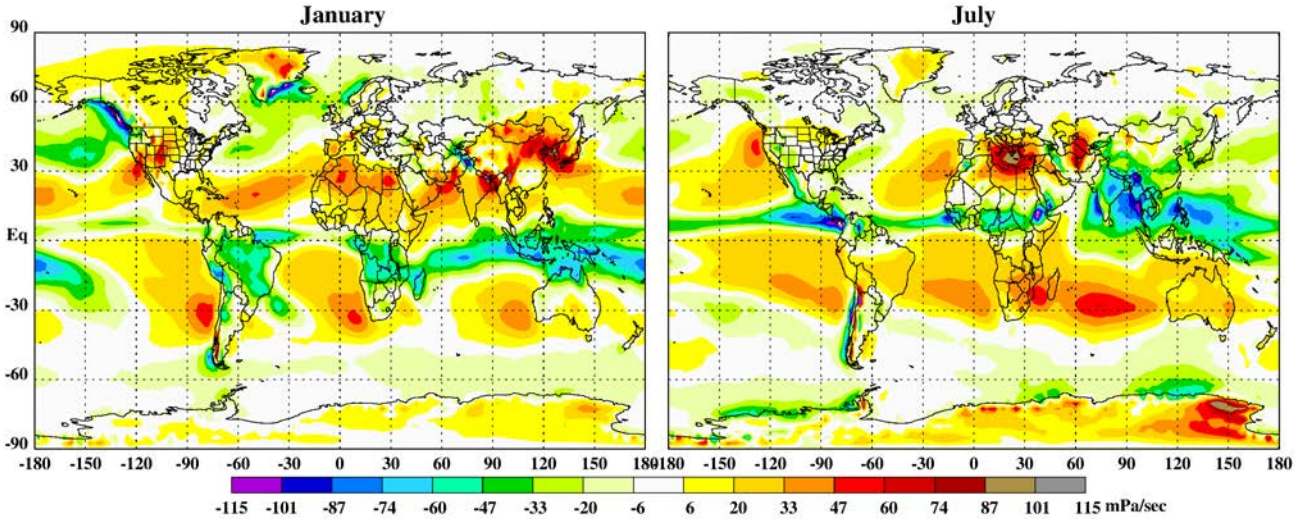


Figure 3.5: *30 year climatology of the mean REM vertical winds (mPa/sec) between 700-200 hPa for January (a) and July (b).*

Ferrel circulations extend from 20°N to 60°N and 40°S to 60°S. Polar circulations extend from 60°N to 80°N and 60°S to 80°S. There is slight upwelling over both poles. The ensemble mean tropopause is overlaid upon these wind fields and shows that in January it descends rapidly between 20°N and 30°N where the convergence of the NH Ferrel and Hadley circulations occur. The NH subtropical jet is located at this point of convergence and steep tropopause gradient. The tropopause descent in the SH is more gradual and the SH subtropical jet is not as strong. In July the circulation patterns are nearly opposite that of January: the SH Hadley circulation is dominant with the centre of upwelling moving northward to 10°N and the descending branch centred at 20°S. The NH Hadley circulation is much weaker; the locations of the Ferrel and Polar circulations are about in the same locations. The tropopause descends most rapidly in July between 20°S and 40°S, the SH subtropical jet is intensified, while the NH subtropical jet is weaker.

In the stratosphere, the vertical winds are much lighter with Eulerian ascending winds occurring poleward of the polar jet axis and descending winds occurring equator-ward of the polar jet axis. Very slight upward motion occurs in the summer hemisphere tropics. Eulerian meridional winds increase with altitude in the winter hemisphere with equator-ward winds between the regions of ascending and descending vertical winds and poleward winds occurring on the winter hemisphere side of the SAO easterlies. The convergence near the stratopause and associated descending winds and ascending polar winds are stronger in the NH winter than in the SH winter. A more in-depth analysis of these winds which includes the Transformed Eulerian Mean residual circulation (*Andrews et al.*, 1987) is presented in *Chapter 11* (Upper Stratosphere and Lower Mesosphere). *Chapter 5* (Brewer-Dobson Circulation) provides greater detail about the stratospheric mean meridional circulation.

A global analysis of the ensemble mean tropospheric vertical velocities for January and July between 700 hPa and 200 hPa are shown in **Figure 3.5**. The analyses illustrate that the zonal mean Hadley circulation depiction is dominated by the convection and upward vertical winds over the Western Pacific and Indian Ocean. The longitudinal band of upward vertical velocities in the tropics represents the Intertropical Convergence Zone (ITCZ). The analysis for the January climatology shows that over Africa, South America, the Indian Ocean and Western Pacific Ocean the maximum upward velocities are south of the Equator. The Atlantic and Eastern Pacific ITCZ bands are north of the Equator. In July, the centres of convection and resulting upward vertical velocities over the Indian Ocean and Western Pacific have moved north of the equator, while the Atlantic and Eastern Pacific bands have moved slightly further to the north. These differences in latitudinal location of upward velocities will impact the longitudinal make-up of the Hadley Circulation and the height and location of the tropopause. Greater details about the tropical circulation patterns, cross tropopause flow, the tropical cold point are provided in *Chapter 8* (Tropical Tropopause Layer).

The longitudinal location of greatest convection and upward velocities that make up the Walker circulations varies with the state of the El Niño Southern Oscillation (ENSO). Under neutral conditions upward velocities are located over Africa, Western Pacific and Western Atlantic. Descending centres are located in the Arabian Sea, and over the Eastern Pacific. Under La Niña conditions, the convection and upward velocities over the Western Pacific and Western Atlantic are enhanced, while the descending motion over the Eastern Pacific is enhanced. The weaker ascending and descending circulations over Africa and the Arabian Sea, respectively, are replaced by a single strong descending circulation. Under El Niño conditions, the convection over the Pacific moves toward the middle Pacific. Enhanced convection and ascending motion occur over Africa. Descending motion occurs over the Western Pacific and Western Atlantic. **Figure 3.6a** shows a time sequence of the 10°N-10°S average ensemble mean vertical velocities at all longitudes between 700-250 hPa. The characteristic longitudinal locations of ascending and descending velocities agree with the above statements. **Figure 3.6b** shows the deseasonalised anomalies, which more clearly shows the major El Niño occurrences of 1984, 1988, 1993, 1998, 2003, 2010, and 2016. La Niña events are harder to pick out since they are close in characterization to neutral conditions. The La Niña event following the 1998 El Niño does show the descending winds over Africa and enhanced ascending velocities over the Western Pacific. There are some time dependent changes from the TOVS to ATOVS periods that show up in the anomalies. Between 10° E and 30° E there is a change from negative anomalies during the 1980's and 1990's to positive anomalies after 2000.

3.4.3 Agreement among the REM members

3.4.3.1 Temperature

The previous section dealt with the mean of three of the more recent reanalyses (MERRA, ERA-I, and JRA-55). Now we examine their variability or "degree of disagreement" over time. We define the degree of disagreement as the SD of the three reanalyses for each month, for each latitude zone, and for each pressure level for the 1979 - 2014 period. Latitude zones (*e.g.* 60° - 90°) are the cosine-weighted summations of the 2.5° zonal SDs. We must note that agreement of the three reanalyses does not imply correctness because the three reanalyses could possibly have similar erroneous analyses. For some months in the upper stratosphere, the temperature disagreement can be greater than 5 K. **Figure 3.7** presents pressure versus time series plots of the temperature SD (K) of the three members of the ensemble. **Figure 3.7** shows how the monthly temperature disagreement varies in three latitude zones (60° N - 90° N, 10° S - 10° N and 90° S - 60° S) in a time versus pressure plot. The mid-latitude plots are not shown but evaluations will be presented below. In all three latitude bands the disagreements are greatest at pressures lower than 10 hPa at which there are fewer conventional observations available for assimilation and the satellite observations generally have very broad weighting functions in the vertical. See *Chapter 2*, **Figure 2.16** to see the vertical extent of satellite channels. The 60° N - 90° N plot shows that at pressures greater than the 20 hPa level, all three reanalyses agree with each other very well, with an SD smaller than 0.5 K. Generally, from 1979 to 2001 the pressure at which the 0.5 K difference contour occurs stays constant between 20 hPa and 10 hPa. Interrupting this period during the 1990s, the NH polar activity was unusually quiet and cold (*Charlton and Polvani*, 2007; *Pawson and Naujokat*, 1999). Then from 2001 to 2014 the pressure at which the 0.5 K contour occurs moves upward to between 7 hPa and 5 hPa. The increased agreement between 20 hPa and 7 hPa is most likely due to the assimilation of AMSU and AIRS observations. The disagreement among the three reanalyses is greater in June–August than in other months due to the ERA-I having warmer temperatures at this level than MERRA and JRA-55.

In the tropics, the disagreement maximizes in two separate layers: between 150 hPa and 70 hPa during the TOVS period (1979 – 1998) and above 20 hPa throughout the entire 1979 – 2014 period. The former disagreement is at the vertical location of the cold point temperature. Apparently, there is greater disagreement among the three reanalyses in determining this temperature during

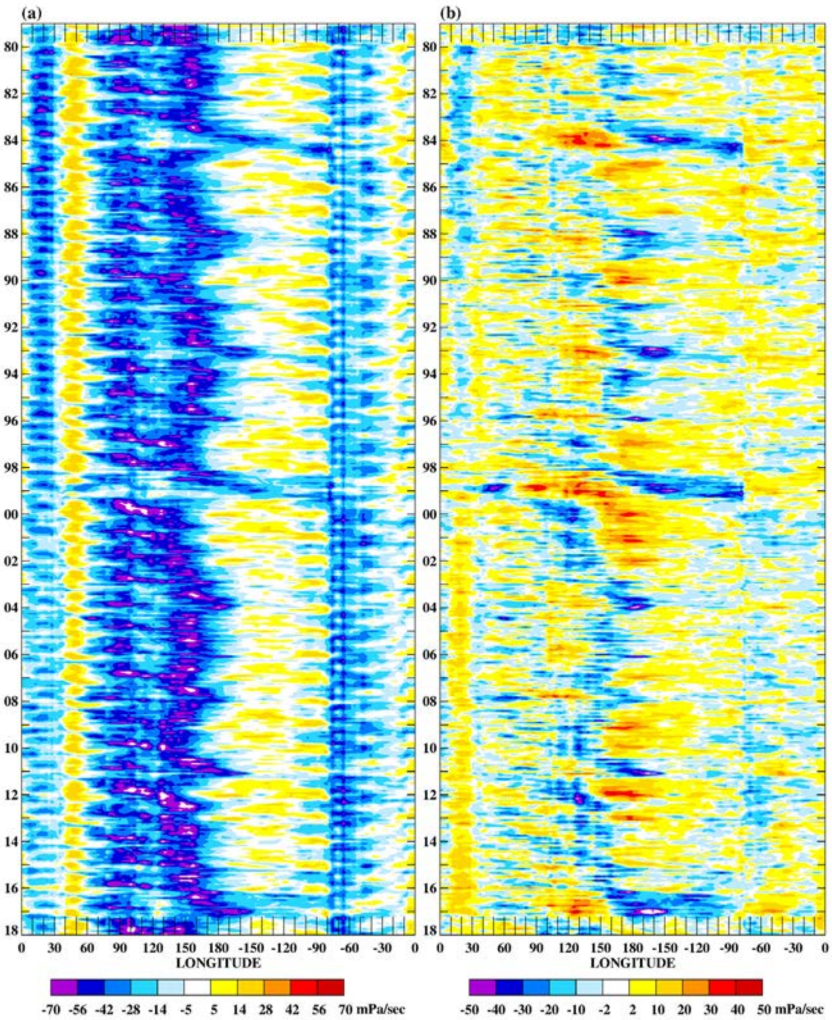


***Figure 3.6:** Hovmöller diagram of (a) the 10° N - 10° S average REM vertical velocities (mPa/sec) at all longitudes from 1979 (top) through 2017 (bottom) between 700 - 250 hPa. (b) Hovmöller diagram of the deseasonalized 10° N - 10° S average REM vertical velocity anomalies (mPa sec⁻¹).*

the TOVS period than during the ATOVS period. During the TOVS period there are only four MSU channels sounding the troposphere and lower stratosphere. The AMSU-A instrument has five channels (5 through 9; 1 through 4 are water vapour channels) sounding the same layer. These additional channels provide information about the temperature structure near the tropopause, thus allowing the reanalyses to better analyse and agree upon the temperature structure there. The pressure at which the greatest differences (3 - 4 K) occur is 2 hPa and has a seasonally varying pattern.

In the 90°S - 60°S zone, the disagreement among the three reanalyses extends lower into the stratosphere than the NH polar zone. This region encompasses all of Antarctica and the ocean surrounding it. There are very few observation sites in this latitude zone. *Manney et al.* (2005) and *Lawrence et al.* (2015) have shown that reanalyses of temperatures in the polar stratosphere can differ significantly depending on what observations are available. Differences greater than 0.5 K during the TOVS period extend to 70 hPa. There are two layers of greatest disagreement in the TOVS period: between 7 hPa and 5 hPa and above 3 hPa. The disagreement between 7 hPa and 5 hPa terminates after 2001, which may be due to the assimilation of AIRS radiances.

The northern mid-latitude (30°N - 60°N) disagreement (not shown) does not change significantly throughout the entire 1979 - 2014 period. Values larger than 0.5 K begin at pressures lower than 7 hPa and have summertime peaks of 2 K at pressures lower than 3 hPa.

The southern mid-latitude (60°S - 30°S) disagreement (not shown) is similar to the SH polar disagreement in that the disagreement during the TOVS period extends to

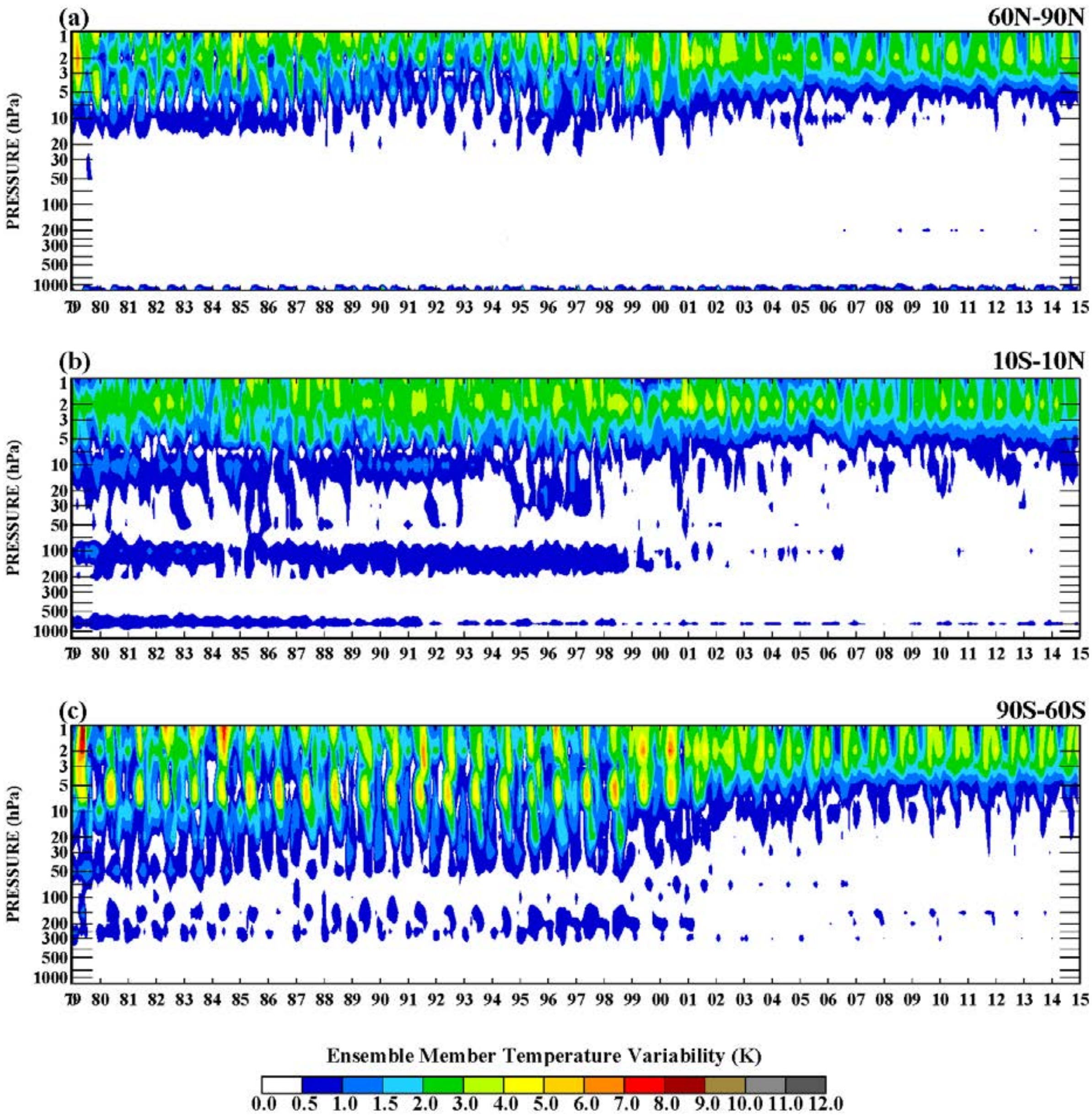


Figure 3.7: *Pressure versus time plots of the temperature SD (K) for each month of the three reanalyses making up the REM for three zonal regions: 60°N - 90° N (a), 10°S - 10°N (b), and 90°S - 60°S (c). Reproduced from Long et al. (2017).*

higher pressures (between 20 hPa and 30 hPa) than during the ATOVS period (between 7 hPa and 10 hPa). Also similar to the SH polar region, there are two layers of greatest disagreement in the TOVS period between 7 hPa and 5 hPa and above 3 hPa. The 7 hPa to 5 hPa layer disagreements also terminate after 2001, just as in the SH polar region.

3.4.3.2 Zonal winds

Figure 3.8 presents pressure versus time series plots of the zonal wind SD (m s^{-1}) of the three members of the ensemble mean. **Figure 3.8** shows the disagreement of the monthly ensemble members' zonal wind in the polar jet regions (40° N – 80° N and 80° S – 40° S) and in the tropics (10° S – 10° N). As with the temperatures, the zonal wind disagreement in the mid-latitudes are not shown but are described in the text below. There is very good agreement of the zonal winds among the three reanalyses in the NH and SH polar jet regions with SDs smaller than 0.5 m s^{-1}. In the NH polar jet region, significant disagreement (> 0.5 m s^{-1}) among the three reanalyses is consistently confined to pressures lower than 5 hPa. Disagreements greater than 0.5 ms^{-1} are nearly eliminated after the transition to ATOVS observations occurs at the end of 1998.

The altitude range of disagreement greater than 0.5 m s^{-1} in the SH polar jet region extends from the upper stratosphere down into the middle stratosphere (10 - 20 hPa) during the TOVS time period, but improves considerably in the ATOVS time period.

The tropical zonal wind disagreement shows much larger values of the order of 10 m s^{-1} in the upper stratosphere than the polar jet values, resulting from disagreement in SAO and QBO winds and winds near the surface at 850 hPa. There is

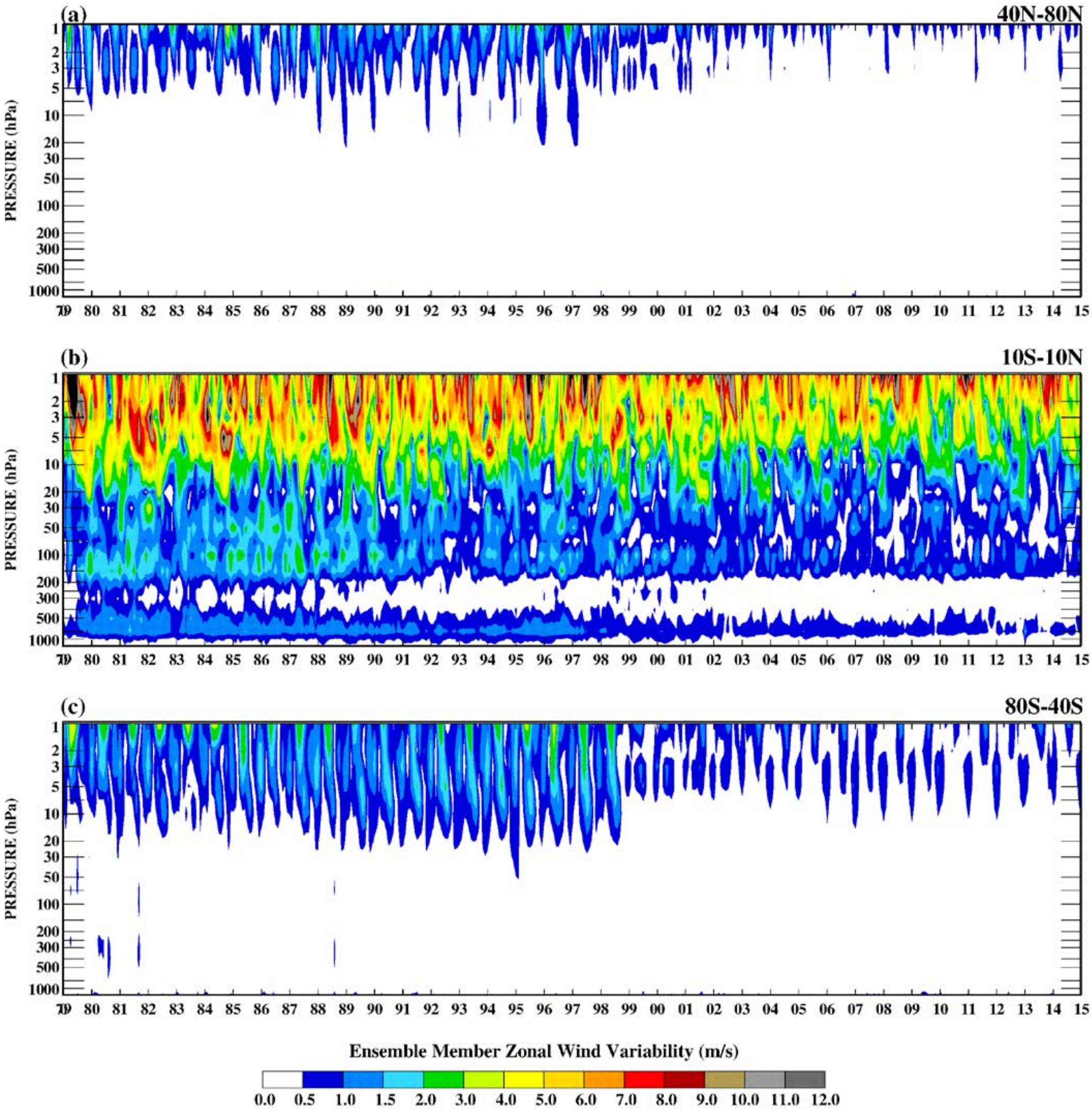


Figure 3.8: *Pressure versus time plots of the zonal wind SD (m s^{-1}) for each month of the three reanalyses making up the REM for three zonal regions: 40° N - 80° N (a), 10° S - 10° N (b), and 80° S - 40° S (c). Reproduced from Long et al. (2017).*

improvement with time in the agreement of the QBO winds and 850 hPa winds, but this improvement does not extend to the SAO height region. The greater improvement in the NH and SH polar jet winds after 1998 versus minor improvement in the equatorial winds illustrates the differences between the mechanisms controlling these winds. The polar jet winds are largely dictated by the latitudinal thermal gradient and resulting thermal wind. However, in the tropics the thermal wind relation breaks down and the wind fields are not well constrained by the assimilated satellite radiances. Alternatively, geostrophic winds using the geopotential can be used to derive the equatorial wind. (*Smith et al.*, 2017). In addition, the tropical winds are primarily determined by the transfer of momentum from upward-propagating waves with spatial scales that are too small to be adequately resolved by the forecast models used in these reanalyses (*Baldwin et al.*, 2001). The tropical winds are therefore highly dependent upon radiosonde observations for speed and direction (and these only extend to 10 hPa). In general the amplitude of the reanalysis tropical winds are smaller than observations. Following the change to ATOVS data, the differences among the reanalyses decrease slightly. No single forecast model included in the REM is capable of generating a QBO on its own. To date, only the forecast model used in MERRA-2 is capable of doing so, and *Coy et al.* (2016) show that after 2000 the MERRA-2 QBO winds are greatly improved versus those in MERRA. It should also be noted that the SAO and polar night jets extend well into the mesosphere (*Smith et al.*, 2020) and reanalyses with higher model tops may produce better results in the upper stratosphere than the lower top reanalysis models.

The characteristics of the NH and SH mid-latitude regions (20°N – 40°N and 40°S – 20°S, respectively; not shown) are very similar to their respective polar jet regions. The NH mid-latitude disagreements during the TOVS period occur at pressures lower than 7 hPa and do not exceed 1.5 m s^{-1}. During the ATOVS period the disagreements are more sporadic and occur at pressures lower than 3 hPa.

The SH mid-latitude disagreement (not shown) occurs at pressures lower than 20 hPa during the TOVS period with values not exceeding 4 m s^{-1}. During the ATOVS period the disagreements become more sporadic, smaller in value, and occur at pressures lower than 7 hPa.

3.5 Intercomparisons of the reanalyses

In this section we extend our evaluation to the individual reanalyses and examine how each of eight reanalyses (CFSR/CFSv2, MERRA, ERA-I, JRA-55, MERRA-2, JRA-25, ERA-40, and *R1*) differs from the REM for both temperatures and winds. We do not show comparisons of *R2*, but one can expect all *R2* qualities to be nearly the same as those of *R1*. We also do not show comparisons with the 20CR as that reanalysis assimilated no upper-air observations. As a result, the 20CR does not show any QBO features in the tropical winds or temperatures, does not observe the occurrences of sudden stratospheric warmings making NH winters 5 K colder and polar zonal winds stronger than they should, and is 3 K – 4 K warmer at 100 hPa in the tropics, which may be due to its coarse model vertical resolution.

3.5.1 Temperature

Figures 3.9 – 3.11 present the time-mean zonal mean temperature difference from the REM (reanalysis – REM) for each month (left columns). The right columns show the time series of the zonal-mean monthly mean differences from 1979 through 2014. The left columns show the average monthly mean differences, while the right columns show the monthly differences over time. Both are useful to illustrate where in the vertical and when in the annual cycle the differences occur and whether these improve over time. Differences in the right column typically do not extend throughout the entire 1979 – 2014 period. Rather, much like the other differences discussed earlier, large improvements are seen going from the TOVS to ATOVS time periods, with the TOVS time period having the larger differences extending down further from the upper stratosphere into the middle stratosphere. Except where specifically mentioned, temperature differences between the individual reanalyses and the REM are within 0.5 K. In general the earlier reanalyses (JRA-25, ERA-40, and *R1*) show greater differences from the REM than the more recent reanalyses (MERRA-2, MERRA, ERA-I, JRA-55, and CFSR/CFSv2). Also, the NH and SH polar latitudes generally show similar difference patterns, with much greater differences in the SH. Thus, in the following, we start with the description on the SH polar latitudes, then mention the NH polar latitudes relatively briefly, and finally describe the equatorial latitudes where the patterns are quite different from those at higher latitudes.

3.5.1.1 SH polar latitudes

MERRA-2 has a year-round cold bias of -1 K to -2 K compared to the REM from 1 hPa to 2 hPa, a year-round the warm bias from 3 hPa to 5 hPa, and a cold bias at 10 hPa from March through June. The time series shows that these biases are largest during the TOVS period, with much smaller differences during the ATOVS period, and that any bias is greatly reduced after August 2004 when Aura MLS temperatures at pressures less than 5 hPa are assimilated.

MERRA shows a warm bias of 1 K to 2 K in the time-mean plot compared to the REM between 2 hPa and 3 hPa from July through February. Below this, between 5 hPa and 20 hPa, there is a cold bias of -1 K to -2 K from April through August. The time series plot shows that this cold bias only exists during the TOVS period, while the warm bias at higher altitudes persists throughout the entire period.

The ERA-I has a mixture of cold (-1 K, March through August) and warm (2 K, November through February) biases compared to the REM between 1 hPa and 3 hPa. An

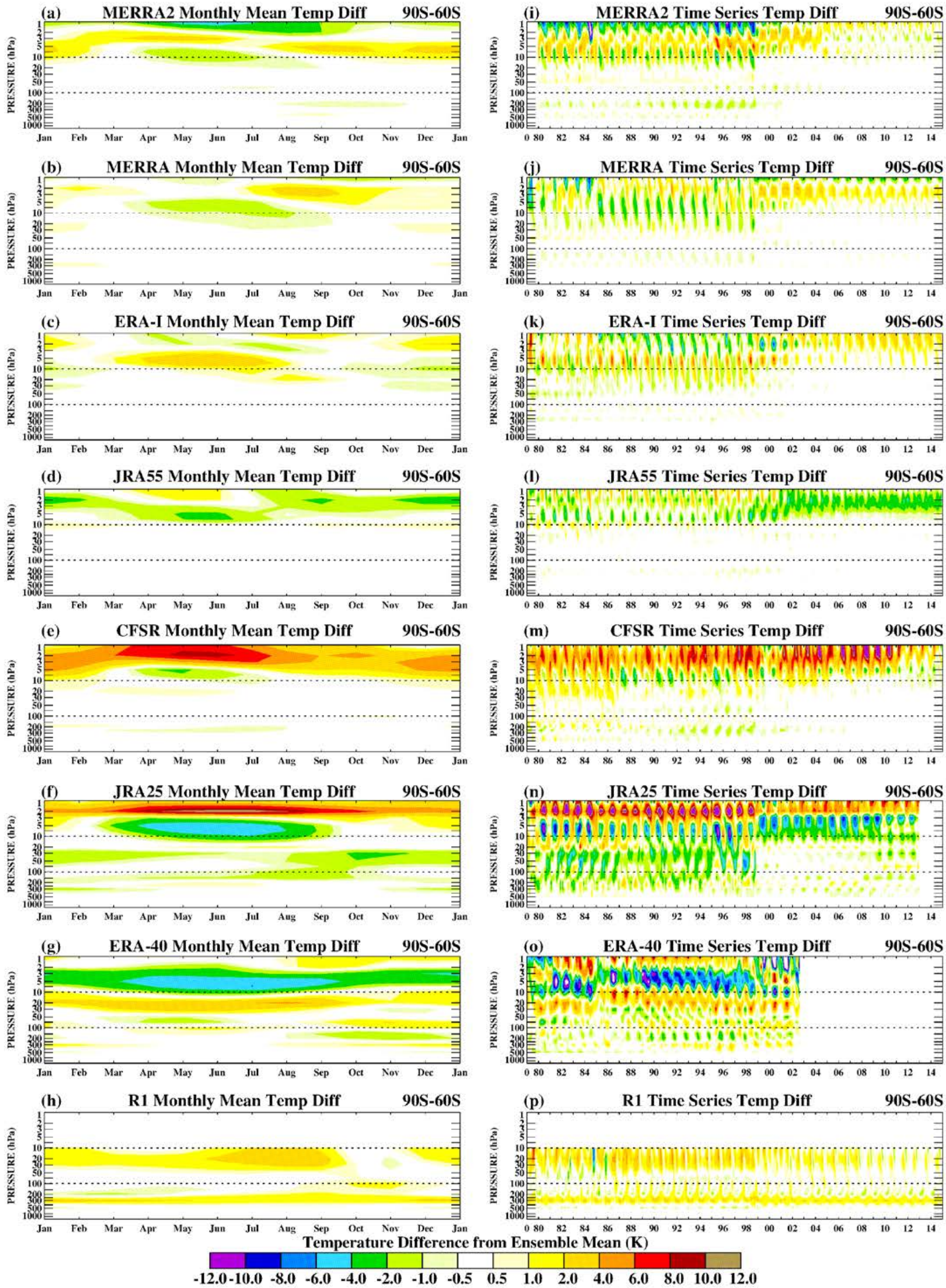


Figure 3.9: *Pressure versus month plots (a - h) and pressure versus time plots (i - p) of the temperature difference (K) in individual reanalyses from the REM for the zonal region 90° S - 60° S. The reanalyses are (a, i) MERRA-2, (b, j) MERRA, (c, k) ERA-I, (d, l) JRA-55, (e, m) CFSR/CFSv2, (f,n) JRA-25, (g, o) ERA-40, and (h, p) R1. The left column plots are the monthly mean differences for the entire 1979 - 2014 period. The right column plots are each month's difference from the REM for that same month. Reproduced from Long et al. (2017).*

opposite set of biases exist slightly below, between 5 hPa and 10 hPa, during roughly the same time periods. The time series plot shows that the upper stratosphere cold bias exists during the 1990s. The upper stratosphere warm bias occurs after 1998, while the warm bias between 10 hPa and 5 hPa persists throughout the entire TOVS period.

The JRA-55 shows a cold bias (-2 K to -4 K) compared to the REM between 1 hPa and 5 hPa from July through March, which then descends to 7 hPa as a warm bias forms between 1 hPa and 2 hPa from March through June. The time series plot shows that temperature differences transitioned from the TOVS to ATOVS period with the cold bias of -4 K to -6 K becoming the dominant feature during this later period.

The CFSR/CFSv2 temperatures are 6 K – 8 K warmer than the REM in the upper stratosphere, peaking during the period of minimum temperatures in that region between March and July. Just below this warm region, there is a small altitude region with colder temperatures than the REM of -1 K and -2 K. The time series plot shows that the CFSR/CFSv2 upper stratospheric warm bias occurs throughout the entire 1979 - 2014 time span with similar seasonal variability.

The JRA-25 time-mean plot shows greater differences from the REM than the above five reanalyses, with a year-long warm bias (8 K to 10 K) compared to the REM from 1 hPa to 3 hPa and a very cold bias (-4 K to -6 K) during the SH winter period between 5 hPa and 10 hPa. In the middle stratosphere there are periods of persistent cool bias with a maximum (-2 K to -4 K) occurring in the August–November months. The time series plot shows that the upper stratosphere warm bias (8 K to 12 K) persists throughout the entire time period, with greater values (> 12 K) in the TOVS period. The cold bias (ranging between -2 K and -10 K) just below the warm bias occurs mostly during the ATOVS time period. The middle stratosphere cold bias (-2 K to -6 K) occurs during the TOVS period (see *Section 5.2* of *Fujiwara et al.*, 2017, for the reason).

The ERA-40 time mean plot shows a strong cold bias (-2 K to -6 K) compared to the REM persisting year-long between 2 hPa and 10 hPa. Just below this is a warm bias (2 K - 4 K) between 10 hPa and 30 hPa. The annual cycle of both the cold bias and warm bias show a slight rising in summer and a lowering in winter months. In the lower stratosphere and upper troposphere, there are layers and monthly periods of slight cold (> -2 K) and slight warm (< 2 K) bias. The time series plot shows that these biases occur throughout most of the ERA-40 time period, which ends in 2002.

R1 does not analyse at pressures lower than 10 hPa, so there is no evaluation in the upper stratosphere. However, there is a nearly year-round warm bias (1 K to 2 K) compared to the REM between 10 hPa and 50 hPa peaking between June and September. Another shallow layer of warm bias (1 K to 2 K) exists between 100 hPa and 400 hPa. The time series plot shows that the middle stratospheric warm bias is most pronounced in the TOVS period.

3.5.1.2 NH polar latitudes

Many features in the upper stratosphere are common for their respective seasons between the NH and SH polar latitudes (**Figure 3.10**). However, differences with the REM in the middle and lower stratosphere in the SH are reduced or eliminated in the NH. The cold bias that occurred between 10 hPa and 5 hPa in the MERRA-2 differences during the SH winter season is not present in the NH winter differences. MERRA differences from the REM in the NH are much smaller in the monthly means, with just a thin warm bias layer between 3 hPa and 5 hPa. The time series shows only slight differences in the middle and lower stratosphere during the TOVS period compared to the same altitude region in the SH. The ERA-I and JRA-55 have very similar seasonal biases as those that occurred in the SH. Similar to MERRA, the time series of differences for the ERA-I during the TOVS period in the middle and lower stratosphere are nearly eliminated. The JRA-55 time series does not have noticeable differences from what was observed in the SH. The CFSR/CFSv2 wintertime warm bias that occurs at pressures lower than 7 hPa extends from October through March. There is no evidence of a cold bias underneath this warm bias in the monthly means as occurs in the SH. The time series of differences shows that the differences that occur in the middle and lower stratosphere in the SH do not exist in the NH. The JRA-25, ERA-40, and *R1* all show similar seasonal biases from the REM in the upper stratosphere. Their time series show reduced differences in the middle and lower stratosphere.

3.5.1.3 Equatorial latitudes

Differences in reanalysis temperatures from the REM in the equatorial regions (10° S – 10° N) vary more on a semi-annual basis. **Figure 3.11** shows that such is the case for the CFSR/CFSv2 upper stratosphere warm bias of 2 K to 4 K and for the JRA-55 upper stratosphere cold bias of -2 to -4 K. MERRA-2 shows relatively small differences (< 1 K) at all altitudes compared to the REM and the near elimination of any bias after August 2004 when MLS temperatures at pressures less than 5 hPa were assimilated. The MERRA and ERA-I exhibit a slight warm bias at pressures lower than 5 hPa. The time series plots for the CFSR/CFSv2 show the jumps associated with the different streams and the gradually increasing warm bias in the upper stratosphere during each of these streams. A warm bias centred at 100 hPa and a cold bias below persist though the TOVS period. The MERRA and ERA-I have temperature biases that are greater during the TOVS period than the ATOVS period. In the ATOVS period the bias in both reanalyses is confined to the upper stratosphere at pressures less than 3 hPa with a warm bias of 0.5 K to 2 K. The JRA-55 reanalyses show

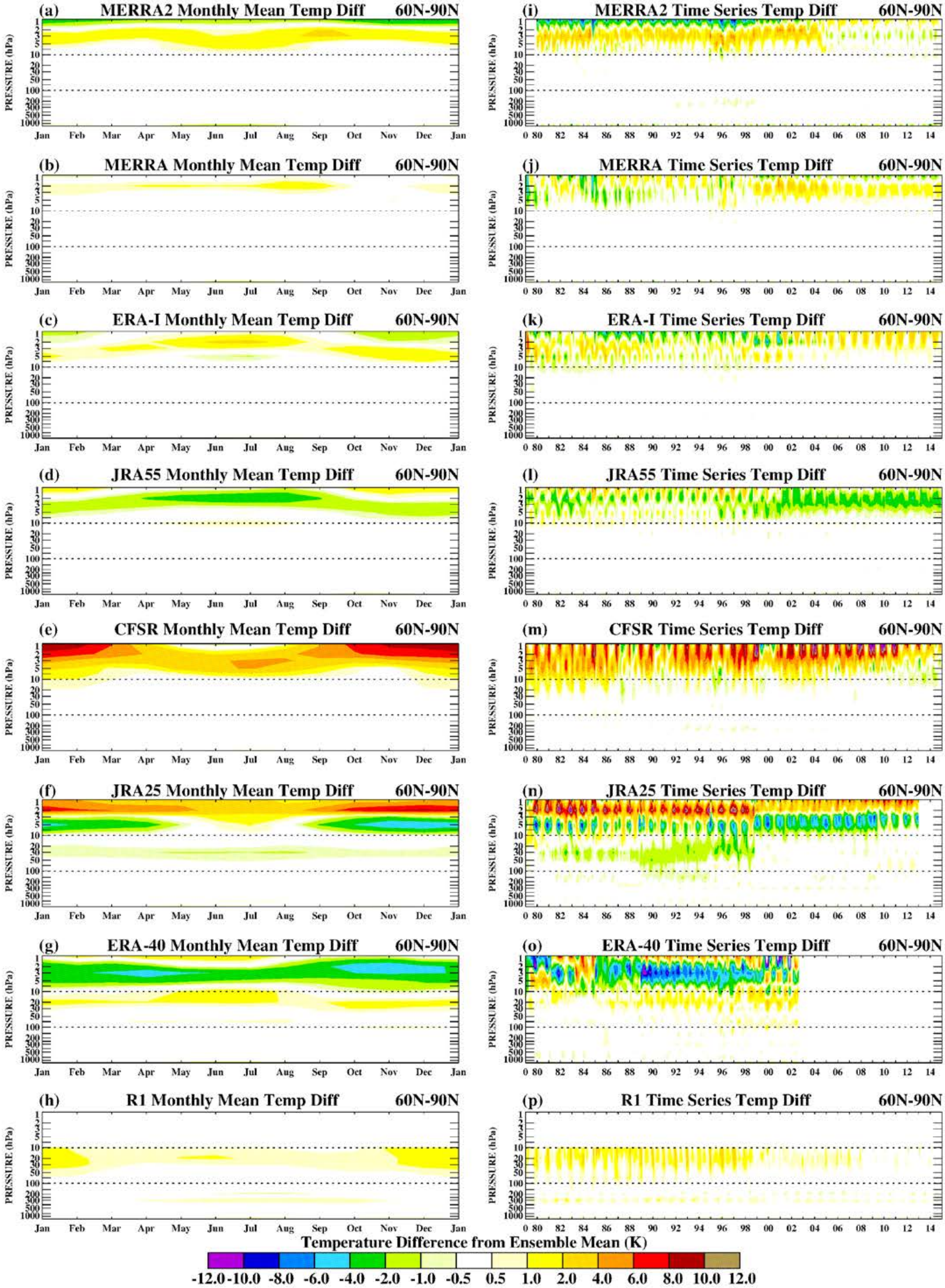


Figure 3.10: *Same as* ***Figure 3.9*** *but for the 60° N – 90° N latitude zone. Reproduced from Long et al. (2017).*

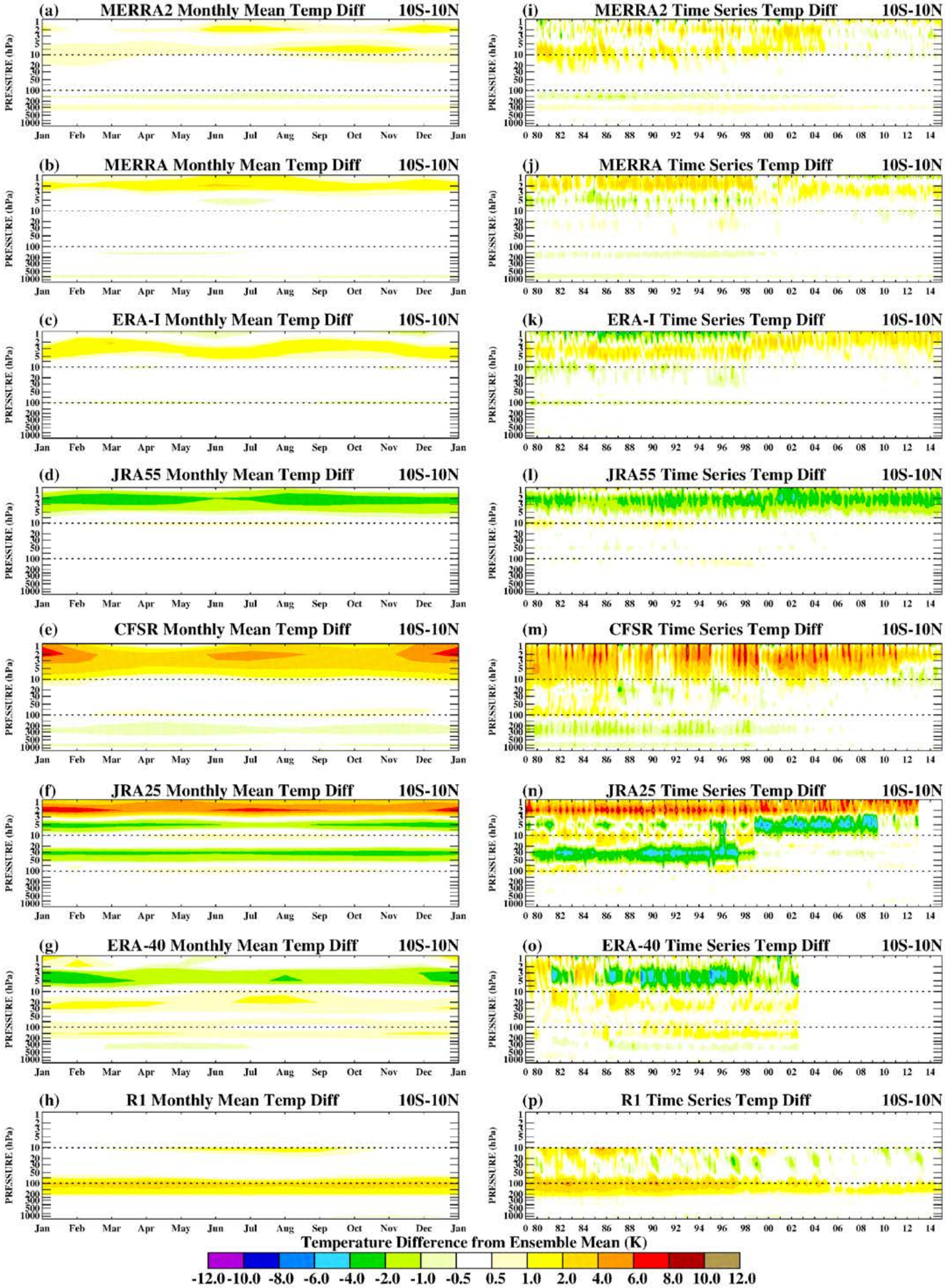


Figure 3.11: *Same as **Figure 3.9** but for the 10° S – 10° N latitude zone. Reproduced from Long et al. (2017).*

that the cold biases are nearly constant throughout the entire time series. The JRA-25 has a consistent warm bias of 4 K to 6 K in the upper stratosphere at pressures less than 3 hPa. Immediately below this at 5 hPa is a cold bias of -2 K to -8 K that is largest during the ATOVS period. Between 30 hPa and 50 hPa, there is another layer of cold bias of -2 K to -6 K that is present only during the TOVS period. ERA-40 has a persistent cold bias of -2 K to -6 K in the upper stratosphere between 2 hPa and 7 hPa and two layers of warm bias of 0.5 K to 1 K in the middle stratosphere and tropopause regions. *R1* in the middle stratosphere has slight warm and cold biases associated with the QBO (seen in the time series plot). There is also a persistent warm bias of 2 K to 4 K in the upper troposphere to tropopause layer between 70 hPa and 200 hPa. This warm bias persists from the TOVS period to the ATOVS period when its magnitude decreases to a warm bias of 1 K to 2 K. *Randel et al.* (2004) pointed this out in their comparison of analyses and attributed the inability to capture lower tropopause temperatures to the coarse vertical resolution and the assimilation of retrieved temperatures (as opposed to radiances).

As discussed in *Section 3.4.3* the three members of the ensemble mean have their greatest disagreement in the upper stratosphere. From the above differences compared to the REM temperatures, the upper stratospheric warm bias of MERRA and ERA-I at all latitudes is nearly counterbalanced by the cold bias of the JRA-55. The ERA-I warm bias between 5 hPa and 7 hPa in the SH polar latitudes is counterbalanced somewhat equally by the MERRA and JRA-55 reanalyses.

3.5.1.4 NH and SH mid-latitudes

The NH and SH mid-latitude zone (30°N – 60°N and 60°S – 30°S, respectively) monthly mean temperature differences and time series temperature differences (not shown) are nearly exactly the same in character, altitude, and value as the respective polar region differences.

3.5.2 Zonal wind

3.5.2.1 SH polar latitudes

The time-mean SH polar jet differences (see the supplement of *Long et al.* (2017); not shown) of the individual reanalyses from the REM are relatively small, ranging from -2 m s^{-1} to 1 m s^{-1}, with most differences smaller in magnitude than that. As presented in *Section 3.4.3.2*, the REM members agree quite well in the polar jet region in both hemispheres. Some notable features are as follows. For all reanalyses except *R1*, the upper stratosphere is the region where the greatest differences from the REM are seen, but shows much improvement from the TOVS to ATOVS periods. MERRA-2 shows further improvements after 2004 when the MLS temperatures started to be assimilated at pressures less than 5 hPa. JRA-25 and ERA-40 show greater differences compared to more recent reanalyses. Finally, *R1* shows an easterly bias to the westerlies during the transition months from westerlies to easterlies in the middle and lower stratosphere for most of the entire time series.

3.5.2.2 NH polar latitudes

Just as with the NH temperature differences in *Section 3.4.1.2*, the NH polar jet wind differences from the REM (see the supplement of Long et al. (2017); not shown) are smaller in magnitude than the SH differences and are restricted mainly to the upper stratosphere.

3.5.2.3 Equatorial latitudes

In **Figure 3.12**, differences in the stratosphere at pressures less than 7 hPa show how the reanalyses differ from each other in the strength of the westerly and the easterly phases in the SAO region. CFSR/CFSv2 and JRA-55 have weaker westerlies and thus have negative biases of greater than -5 m s^{-1} during the March – April and September – November westerly periods. They also have positive biases greater than 3 m s^{-1} during the December – February easterly period. MERRA and ERA-I have stronger westerlies and show positive biases of greater than 3 m s^{-1} during the March – April and September – November westerly periods. They also have stronger easterlies during the December – February period but differ slightly during the July – August easterly period. This results in the MERRA and ERA-I having negative biases of less than -3 m s^{-1} during the former period. The SAO westerlies in MERRA-2 are more than 10 m s^{-1} stronger than those in the REM. The time series shows that the stronger westerlies occur primarily during the TOVS period. *Kawatani et al.* (2016) and *Molod et al.* (2015) note that the downward-propagating westerly phase of the SAO is enhanced during the 1980s and could be caused by strong gravity wave forcing.

MERRA-2 also transitions from QBO westerlies to easterlies more rapidly than the REM during the TOVS period. The time series plots also show where each reanalysis has a slight easterly or westerly bias associated with the phase of the descending QBO winds. The JRA-25 and *R1* show greater differences from the REM than the other reanalyses. *R1* shows a westerly bias of > 4 m s^{-1} during the easterly phase of the QBO from 10 hPa down to 100 hPa. This was also discussed by *Pawson and Fiorino* (1998b). The JRA-25 has an easterly bias of > 4 m s^{-1} during the easterly phase of the QBO from 10 hPa down to 30 hPa. It should be noted that the CFSR/CFSv2 used ERA-40 zonal winds as substitute observations between 30° S and 30° N and from 1 hPa to 30 hPa from 1 July 1981 to 31 December 1998 (*Saha et al.*, 2010); hence their differences from the REM during that time period and in that pressure range are very similar.

Interestingly, in **Figure 3.12** there are also sizable differences in the troposphere. The CFSR zonal winds in the tropical

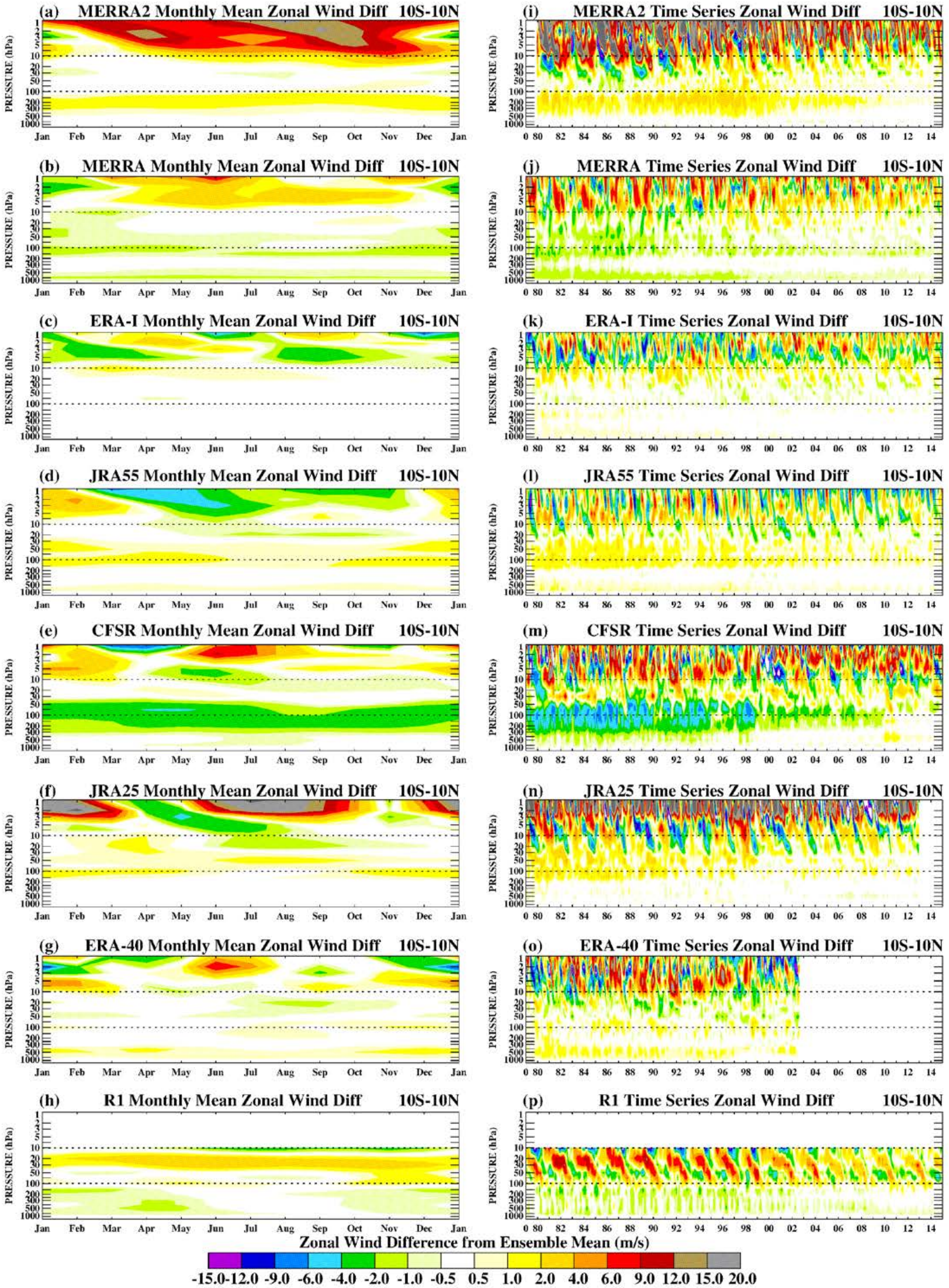


Figure 3.12: *Pressure versus month plots (a - h) and pressure versus time plots (i - p) of the zonal wind difference (m s^{-1}) in individual reanalyses from the REM for the zonal region 10 °S–10 °N. The reanalyses are (a, i) MERRA-2, (b, j) MERRA, (c, k) ERA-I, (d, l) JRA-55, (e, m) CFSR/CFSv2, (f, n) JRA-25, (g, o) ERA-40, and (h, p) R1. Reproduced from Long et al. (2017).*

upper troposphere during the TOVS years have an easterly bias. This may be associated with the CFSR having a cold bias of about 1 K in the upper troposphere during this time period. The JRA-55 zonal winds have a westerly bias during this time period. The MERRA and ERA-I zonal wind differences in the upper troposphere are no larger than 0.5 m s^{-1}. Hence, the differences from the REM show that the CFSR has a consistent layer of negative biases of -1 m s^{-1} to -2.5 m s^{-1} from 50 hPa to 300 hPa. The JRA-55 shows the other extreme of a consistent positive bias of 1 m s^{-1} to 2 m s^{-1} from 30 hPa to 200 hPa. The time series plots confirm that these upper troposphere zonal wind biases are persistent during the TOVS time period and are reduced in the ATOVS period. MERRA-2 shows large positive differences of > 6 m s^{-1} from the REM in the upper stratosphere (SAO region). The time series show that these large differences occur mostly during the 1980s and periodically extend to 20 hPa. These large differences continue throughout the time series but are confined to the upper stratosphere after the 1990s.

3.5.2.4 NH and SH mid-latitudes

Characteristically, the zonal winds in the NH and SH mid-latitudes (20° N - 40° N and 40° S - 20° S, respectively; not shown) are different depending upon the altitude. In the troposphere there is the subtropical jet with maximum winds near 200 hPa. In the lower stratosphere there is a lull between the equatorial winds and the polar jet. The upper stratosphere is seasonally transitioning from the SAO to the winter polar jet. The differences from the REM show that all the reanalyses are in very good agreement with the tropospheric subtropical jet. In the lower stratosphere *R1* has a westerly bias of 0.5 m s^{-1} to 1 m s^{-1}, which is greatest in the early 1980s and diminishes to nil by the 2000s. The CFSR/CFSv2, interestingly, has an easterly bias of -0.5 m s^{-1} to -1 m s^{-1} during the TOVS period and is eliminated in the ATOVS period. All the other reanalyses are in good agreement (differences within ±0.5 m s^{-1}) with the REM in the lower stratosphere. In the middle stratosphere the JRA-25 has differences between -0.5 m s^{-1} and -1 m s^{-1} from the REM in both the NH and SH mid-latitudes. In the upper stratosphere the more recent reanalyses have differences between -1 m s^{-1} and 1 m s^{-1} from the REM, which diminish further during the ATOVS period. The JRA-25 and ERA-40 have slightly larger differences, which also diminish appreciably in the ATOVS period.

3.5.2.5 Comparisons with Singapore QBO winds

Kawatani et al. (2016) provides a thorough evaluation of the RMS differences in QBO (70 hPa – 10 hPa) zonal winds among the more recent reanalyses and observations from all the radiosonde sites in the equatorial-latitude zone. *Kawatani et al.* (2016) also show that of the nearly 220 radiosonde stations in the 20° S - 20° N zone, Singapore (1° N, 104° E) is the only station that reports 10 hPa observations 80 – 100 % of the time between 1979 and 2001. For this reason, we will focus just upon comparisons between the reanalyses and zonal winds at Singapore. This is not to imply that Singapore is representative of the entire tropical zone. There is longitudinal variability in the zonal-mean zonal winds (*Kawatani et al.*, 2016). **Figure 2** of *Kawatani et al.*,2016 shows that the RMS differences between the various reanalyses and the Singapore wind observations declines over the 1979 – 2011 time period. Correlations among the monthly mean MERRA-2,

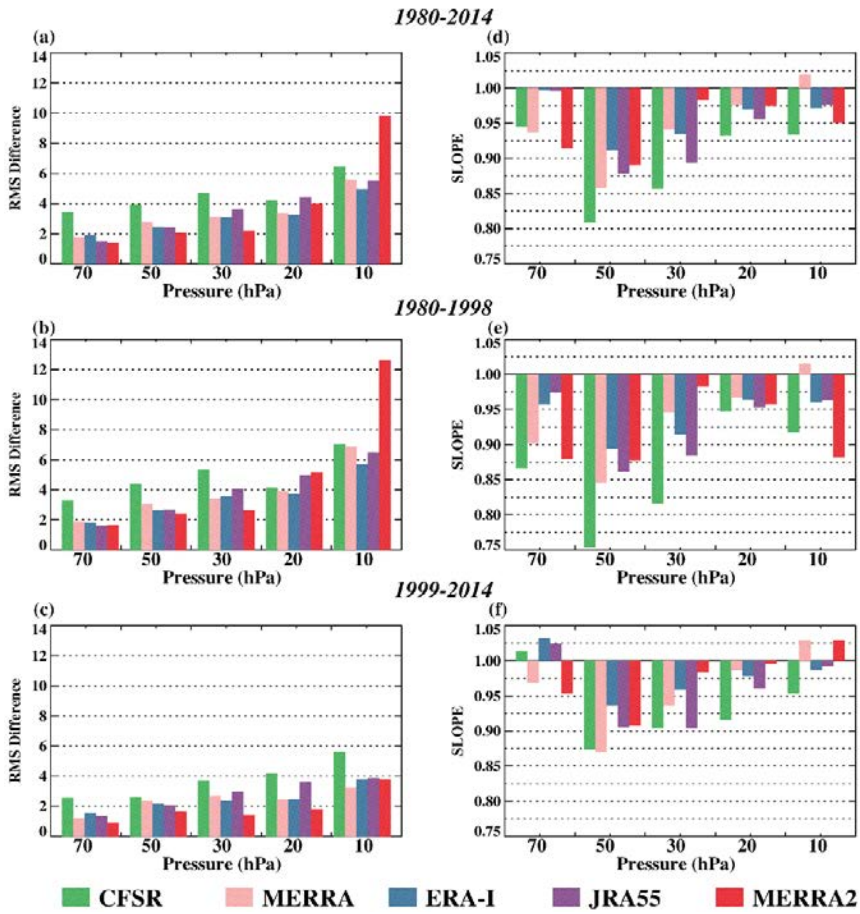


Figure 3.13: *RMS differences (m s^{-1}) (a - c) and linear slopes (d - f) of the matched QBO zonal wind anomalies at 70, 50, 30, 20, and 10 hPa for the CFSR/CFSv2, MERRA, ERA-I, JRA-55, and MERRA-2 reanalyses interpolated to Singapore (1° N, 104° E) versus the observed Singapore monthly mean zonal winds from the FUB. RMS differences and slopes are computed for the 1980 - 2014 time period (a, d), the 1980 - 1998 period (b, e), and the 1999 - 2014 period (c, f). Slopes less than 1.0 indicate that the reanalysis zonal winds are weaker than the Singapore zonal winds. Reproduced from Long et al. (2017).*

MERRA, ERA-I, JRA-55, and CFSR/CFSv2 QBO zonal winds (interpolated to Singapore) and the monthly mean radiosonde wind observations at Singapore (obtained from the Free University of Berlin) are mostly above 0.9. More information about how the reanalyses differ from the Singapore winds can be obtained by evaluating the linear regression line between the observed and analysed QBO winds and their scatter. **Figure 3.13a–c** shows the RMS differences in the reanalyses QBO winds and those at Singapore. Comparisons are shown for the entire 1980-2014 period and then divided into the TOVS (1980-1998) and ATOVS (1999-2014) periods. All of the reanalyses RMS differences are smaller during the ATOVS period. All of the RMS differences increase from 70 hPa to 10 hPa as does the amplitude of the winds at these levels. The RMS differences decrease by one-half to one-third from the TOVS to the ATOVS period. Of these five reanalyses, the CFSR/CFSv2 performs the poorest with higher RMS differences at nearly all pressure levels during all periods. MERRA-2 has the largest RMS differences at 10 hPa during the TOVS period, but improves during the ATOVS period. As seen in **Figure 3.12**, MERRA-2 has large irregularities in the 1980s and in 1993. As mentioned earlier, these irregularities are a result of overly strong SAO westerlies that propagate down to the middle stratosphere. *Coy et al.* (2016) explain that during the 1980s and early 1990s MERRA-2 overemphasized the annual signal. **Figures 3.13d–f** show the slope of the regression line between the individual reanalysis QBO winds and the Singapore QBO winds. The maximum underestimation (slope smaller than 1) at 50 hPa is present in all of the reanalyses. The reanalysis winds and Singapore winds become more similar in strength at lower pressure levels and are closer in strength during the ATOVS period than the TOVS period. The CFSR/CFSv2 has consistently weaker winds at all

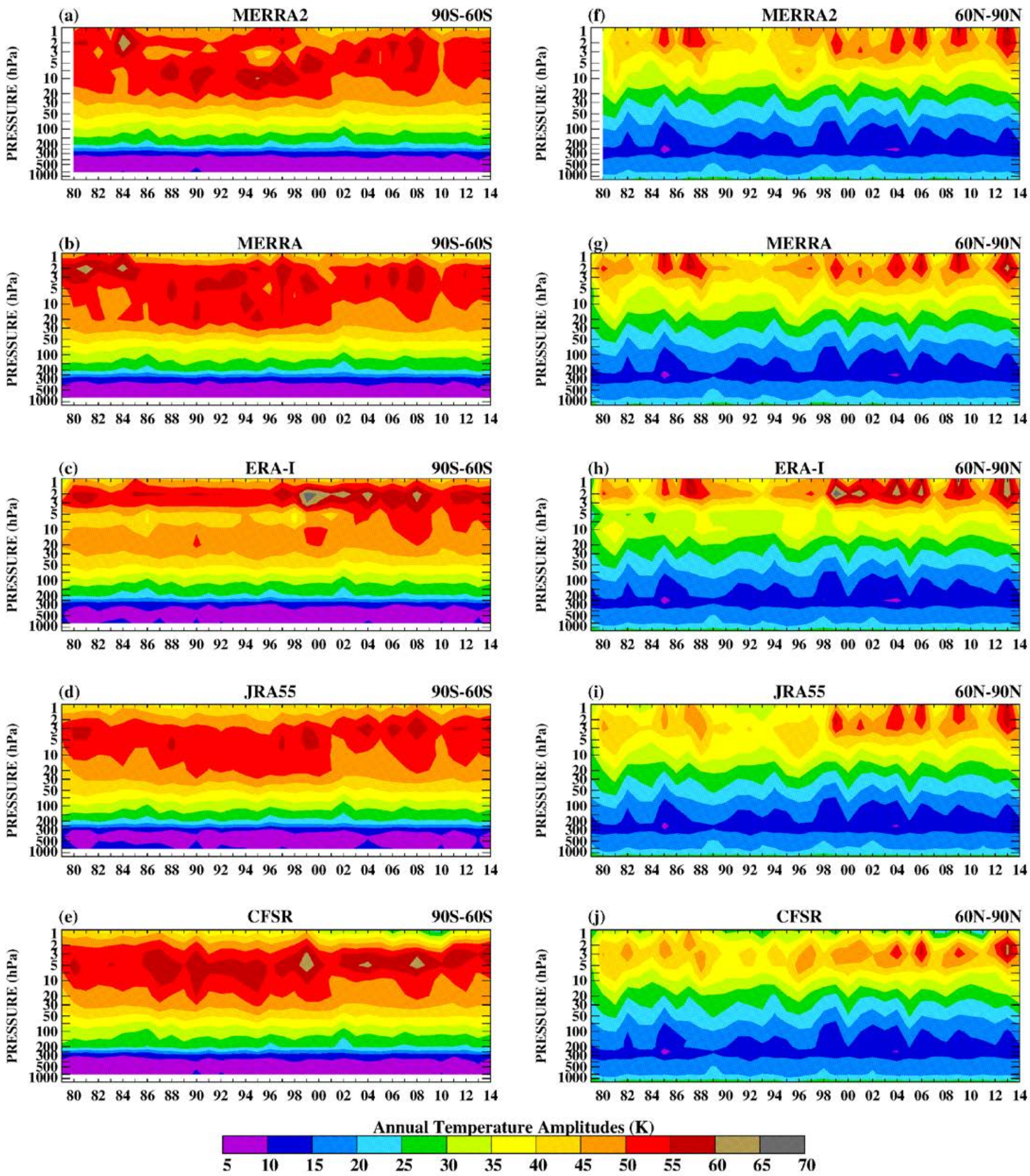


Figure 3.14: *Yearly annual temperature amplitude (K) for 90°S-60°S (a-e) and 60°N-90°N (f-j) from the (a, f) MERRA-2, (b, g) MERRA, (c, h) ERA-I, (d, i) JRA55, and (e, j) CFSR/CFSv2 reanalyses. Note that the SH annual amplitude is much larger than the NH amplitude. No analysis is performed between 1000 and 700 hPa for the SH plots as this is below the Antarctic surface. Reproduced from Long et al. (2017).*

pressure levels during both the TOVS and ATOVS periods. No one reanalysis is better than the others at all QBO levels in either the TOVS or ATOVS period. Greater detail about the QBO is provided in *Chapter 9* (Quasi-Biennial Oscillation).

3.6 Amplitude of polar annual temperature cycle

Another way to examine the differences among the reanalyses is to compare their annual temperature amplitude (warmest summer month minus coldest winter month) in the polar latitudes. If a reanalysis has a wintertime warm bias or a summertime cold bias, then its annual temperature amplitude will be smaller compared to the other reanalyses. Generally, as **Figure 3.2** shows, the summertime temperatures do not vary much from year to year, while the wintertime temperatures have greater inter-annual variability. The mean polar temperatures in **Figure 3.2** indicate which months would likely be used as the warmest and coldest at the various pressure levels. For these differences we use the coldest (warmest) month from November through March and the warmest (coldest) month from May through September for the Northern Hemisphere (Southern Hemisphere). The lower variability in the SH temperatures ensures that the same months are used for the 1979 to 2014 period. However, in the NH the coldest month at a particular pressure level depends upon whether an SSW occurs. In the upper stratosphere, after an SSW the low temperatures following the warming are usually the lowest of the year. Without a warming the lowest temperatures may well have occurred in November or December. In the middle stratosphere the lowest temperatures will usually occur in December. In the lower stratosphere the lowest temperatures will usually occur in December or January. In **Figure 3.14** a time series of the SH and NH polar zone annual temperature amplitudes is presented. In general, the SH annual amplitudes in the middle and upper stratosphere are up to 25 K larger than at the same level in the NH, largely because of the persistent and colder SH winters. At pressures greater than 300 hPa, temperature amplitudes in the SH are smaller than those in the NH. SH temperature amplitudes increase from 5 K – 15 K in the troposphere to 45 K – 60 K in the middle stratosphere. Maximum amplitudes (60 K – 70 K) in the SH occur above 10 hPa. In the NH polar latitudes, the minimal amplitude of 5 K – 15 K occurs at the polar tropopause. Between the surface and the tropopause, the temperature amplitude is larger at 15 K – 25 K. Above the tropopause the temperature amplitude increases up to about 2 hPa – 3 hPa where the temperature amplitude lies in the 55 K – 60 K range, although the depth of this layer is not nearly as extensive as in the SH polar regions. The depth of this variability below 10 hPa ends prior to 2002. The assimilation of GPSRO data beginning in 2002 could be the cause for this decrease in SH high latitude variability. There is good agreement among these five more recent reanalyses on the years of peak amplitude in the NH polar region upper stratosphere.The peak SH amplitudes of the five reanalyses are in lesser agreement in year and pressure range.

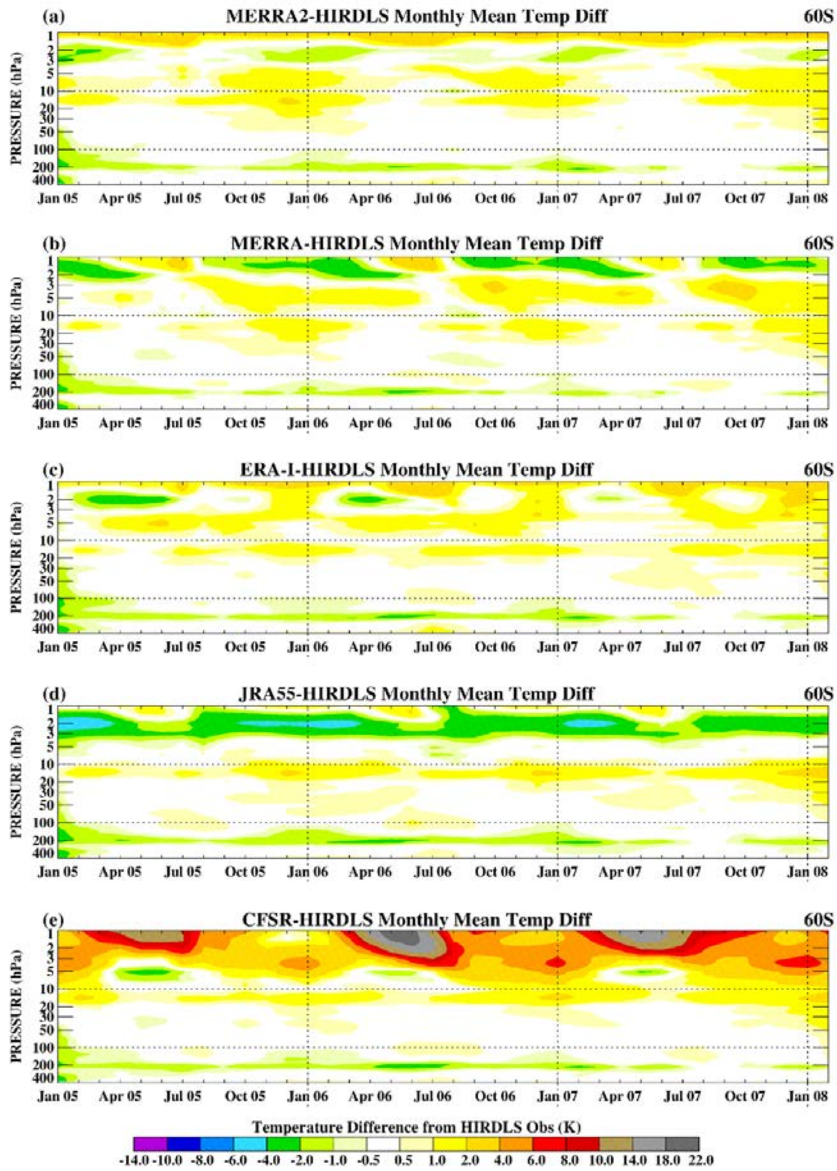


Figure 3.15: *Pressure versus time plots of differences in reanalyses minus HIRDLS temperatures (K) from January 2005 through January 2008 for the Southern Hemisphere high-latitude zone (60° S). The reanalyses are (a) MERRA-2, (b) MERRA, (c) ERA-I, (d) JRA-55, and (e) CFSR/CFSv2. Reproduced from Long et al. (2017).*

Individually, the five more recent reanalyses agree well with each other from the surface through the lower stratosphere in both hemispheres. However, the ERA-I shows an annual temperature amplitude in the middle stratosphere that is 5K – 15K smaller than the other four reanalyses in the SH and about 5K smaller in the NH polar regions from 1979 – 2002. The JRA-55 has smaller maximum amplitudes in the SH than the other four reanalyses, which is associated with its seasonally low temperature bias in the upper stratosphere, whereas the CFSR/CFSv2 tends to have consistently large maximum amplitudes which are associated with its seasonally warm bias. However, the CFSR/CFSv2 temperature amplitudes peak at greater pressures in the upper stratosphere and then decrease rapidly between 3hPa and 1hPa in both hemispheres, particularly in the ATOVS period. This is most likely due to the fact that the CFSR/CFSv2 did not bias correct the SSU channel 3 observations and did not assimilate the top AMSU-A channel 14.

As a group the NH plots show that the greatest amplitudes occur at 2hPa. The years with this large amplitude are years in which an SSW occurred. This is a result of the very cold air that immediately follows the warming in the upper stratosphere. The years in which an SSW did not occur (*e.g.* the 1990s) have smaller temperature amplitudes in the upper stratosphere. The SH years in which there was a great amount of wave activity during the winter months had warmer winters and consequently smaller annual amplitudes. This is particularly noticeably in 2002 and 2010. These two years exhibited a very early transition from winter circulation to summer circulation, similar to a final warming in the NH. Final warmings are not followed by very cold air in the upper stratosphere. The ERA-I stands out as having smaller annual amplitudes in the SH middle stratosphere compared to the other four reanalyses during the TOVS period.

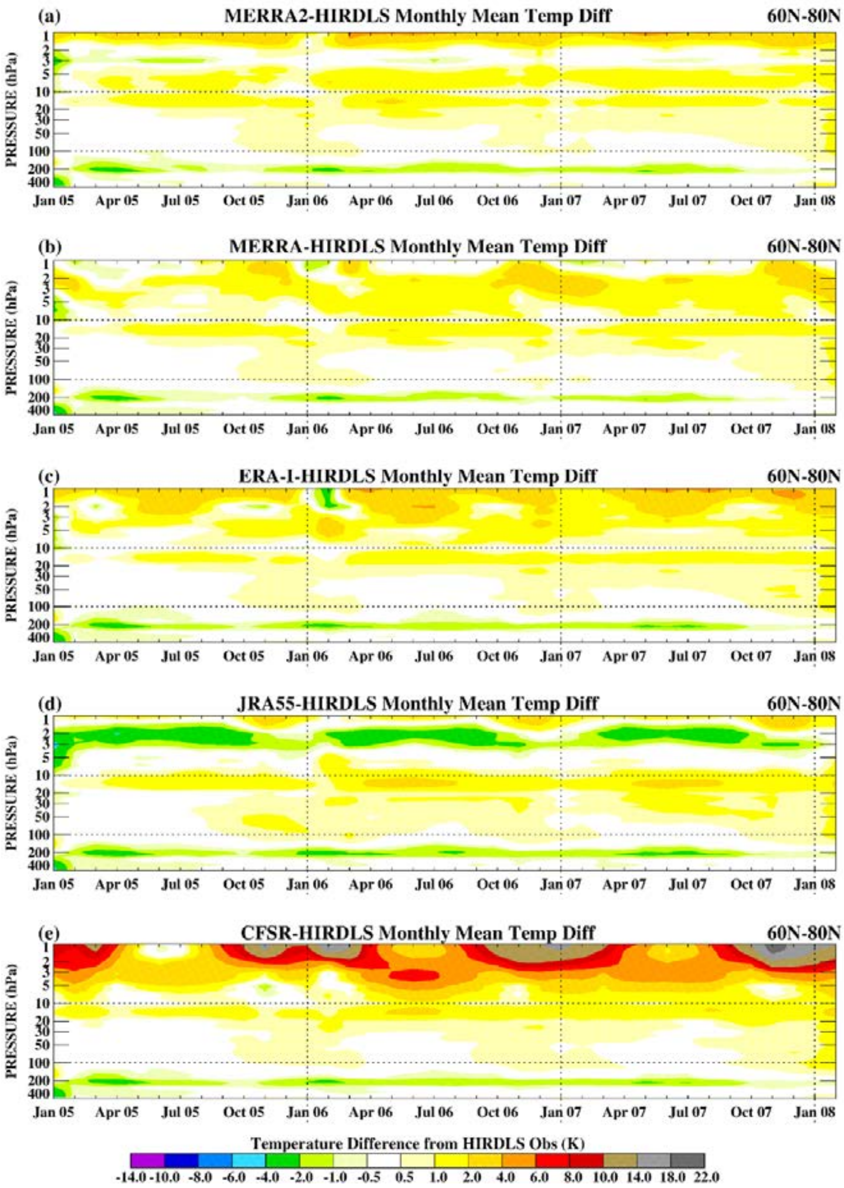


***Figure 3.16:** Same as **Figure 3.15** except for the Northern Hemisphere high-latitude zone (60° N - 80° N). Reproduced from Long et al. (2017).*

3.7 Comparisons with satellite temperature observations

3.7.1 HIRDLS and MLS temperatures

The NASA Earth Observing System (EOS) Aura spacecraft was launched in July 2004 and has several on-board instruments that measure multiple atmospheric constituents. The High Resolution Dynamics Limb Sounder (HIRDLS; *Gille et al.*, 2008) instrument on the Aura spacecraft made measurements from the upper troposphere through the mesosphere until it prematurely ceased functioning in mid-2008. Quality temperature measurements extend from January 2005 through March 2008. The HIRDLS measurements were not assimilated by any of the reanalyses and thus are independent measurements. Monthly mean temperature differences in reanalyses from the HIRDLS (reanalysis – HIRDLS) temperatures at NH high latitudes (60°N - 80°N), the tropics (10°S - 10°N), and SH high latitudes (60°S) were generated for the 2005 through 2008 period. **Figures 3.15–3.17** present the differences in MERRA-2, MERRA, ERA-I, JRA-55, and CFSR/CFSv2 from the HIRDLS monthly means for these latitude zones, respectively. The time, location, and amplitude of the SH differences are generally similar to those of the reanalyses from the REM (**Figure 3.6**).

MERRA-2 has a warm bias all year long at 1 hPa and a 1 - 2 K cold bias from November through March. MERRA has a cold bias of 2 - 4 K from August through April at 1 - 3 hPa and a 2 K warm bias from May through July. ERA-I has a 2 K cold bias at 2 hPa from February through May. JRA-55 has a 4 - 6 K cold bias from July through April between 2 hPa and 3 hPa that becomes thinner in altitude from April to July as a warm bias occurs from 1 hPa to 2 hPa. The CFSR/CFSv2 has a very warm bias of over 14 K in the April to July period at pressures lower than 5 hPa with a cold bias at 7 hPa during this same time period. All of the reanalyses show a slight (< 1 K) warm bias in the middle stratosphere during the November through March period.

In the NH, the cold bias of MERRA-2 in the summer period is smaller in the NH, while the year-long warm bias exists at 1 hPa. The cold bias that MERRA has in the SH does not exist in the NH. The mid-winter warm bias that was in the SH is about 1 K warmer in the NH. Similarly, the ERA-I does not have a cold bias in the late winter–spring period, but there is a warm bias in midsummer in the upper stratosphere. The CFSR/CFSv2 and JRA-55 differences with HIRDLS occur in the same seasons as in the SH with little change in amplitude. Of interest is that all the reanalyses show a similar warm bias as in the SH during the November through March period.

In the tropics, MERRA-2 continues to have a year-long warm bias at 1 hPa and a slight warm bias near 10 hPa. In 2006 – 2007 MERRA has a warm bias between 2 hPa and 3 hPa during January and February and moves lower to 5 hPa to 10 hPa during the other months of the year. ERA-I seems to have a year-long 0.5 K to 1 K warm bias at pressures lower than 10 hPa. JRA-55 has a year-long 1 - 2 K cold bias between 5 hPa and 2 hPa. The CFSR has a warm bias, similar to that at high latitudes, on a semi-annual basis in the upper stratosphere.

The Microwave Limb Sounder (MLS) is also on the EOS Aura spacecraft. Monthly zonal means of temperatures from the version 4 retrievals were provided by the MLS team for comparisons with reanalyses for the 2005–2014 period. The characteristics of the MLS temperatures are described by *Schwartz et al.* (2008) and *Livesey et al.* (2015). Note again that among the reanalyses, MERRA-2 is the only one that assimilated MLS temperatures but only at pressures less than 5 hPa. HIRDLS temperatures have been noted to be colder than the Aura MLS temperatures (*Gille et al.*, 2008) in the upper stratosphere. Evidently, differences in MERRA-2, MERRA, ERA-I, JRA-55, and CFSR/CFSv2 temperatures from the MLS temperatures (not shown) are very similar to those with the HIRDLS but less positive. Differences greater than ± 2 K only occur above 10 hPa. Bands of differences of the order of 1 K are present below 10 hPa; however, the MLS documentation notes that there are known oscillations of this magnitude in comparison with other satellite temperature sensors, so these latter differences are not considered significant. Overall differences from the MLS observations are in agreement with the characteristics already described for each of these reanalyses.

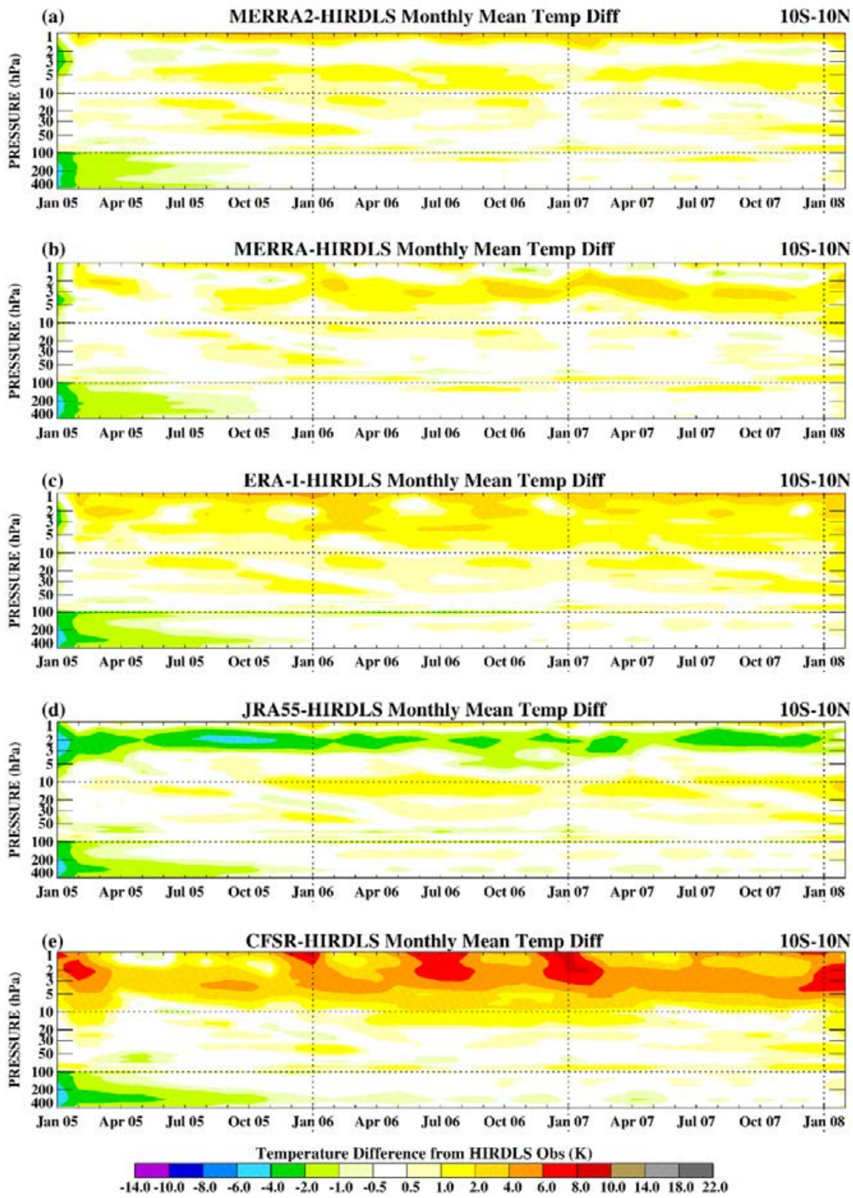


Figure 3.17: *Same as* ***Figure 3.15*** *except for the equatorial-latitude zone (10° S - 10° N). Reproduced from Long et al. (2017).*

3.7.2 Comparisons with COSMIC temperature observations

COSMIC GPSRO monthly zonal mean dry temperatures from January 2007 through December 2014 (level 3, version 1.3) were obtained from the JPL GENESIS data portal. *Leroy et al.* (2012) explain the technique through which the RO observations were turned into temperatures and transposed from altitude to pressure surfaces. We use these data to compare against the MERRA-2, MERRA, ERA-I, JRA-55, and CFSR/CFSv2 monthly zonal mean temperature for the same period. The COSMIC data set provides temperature from 400 hPa to 10 hPa. We will not perform comparisons with data at pressures higher than 200 hPa as atmospheric water vapour causes deviations in the actual temperatures from the dry temperatures. **Figure 3.18** shows the 8-year time series of differences (reanalysis – COSMIC) between the reanalysis temperatures and the COSMIC temperature in the SH polar latitudes (90°S – 60°S). Most obvious is a recurring 1 K difference between the reanalyses and COSMIC from January through July from 10 hPa down to 100 hPa. This is during the transition from SH summer to winter. During the transition from SH winter to summer, there is a 0.5 K to 1 K difference also extending from 10 hPa to 100 hPa. The source of these two biases could be in how the COSMIC zonal mean temperatures are generated as there is a 3 – 5-day time averaging in which temporal transitions may be smoothed out. Also, *Steiner et al.* (2020) indicates that all GPSRO processing centres show greater uncertainties in their temperature product above 25 km especially in the high latitudes. All of the reanalyses differed (except MERRA) in assimilating either the GPSRO bending angle or refractivity (*Curcurull et al.*, 2007; *Poli et al.*, 2010).

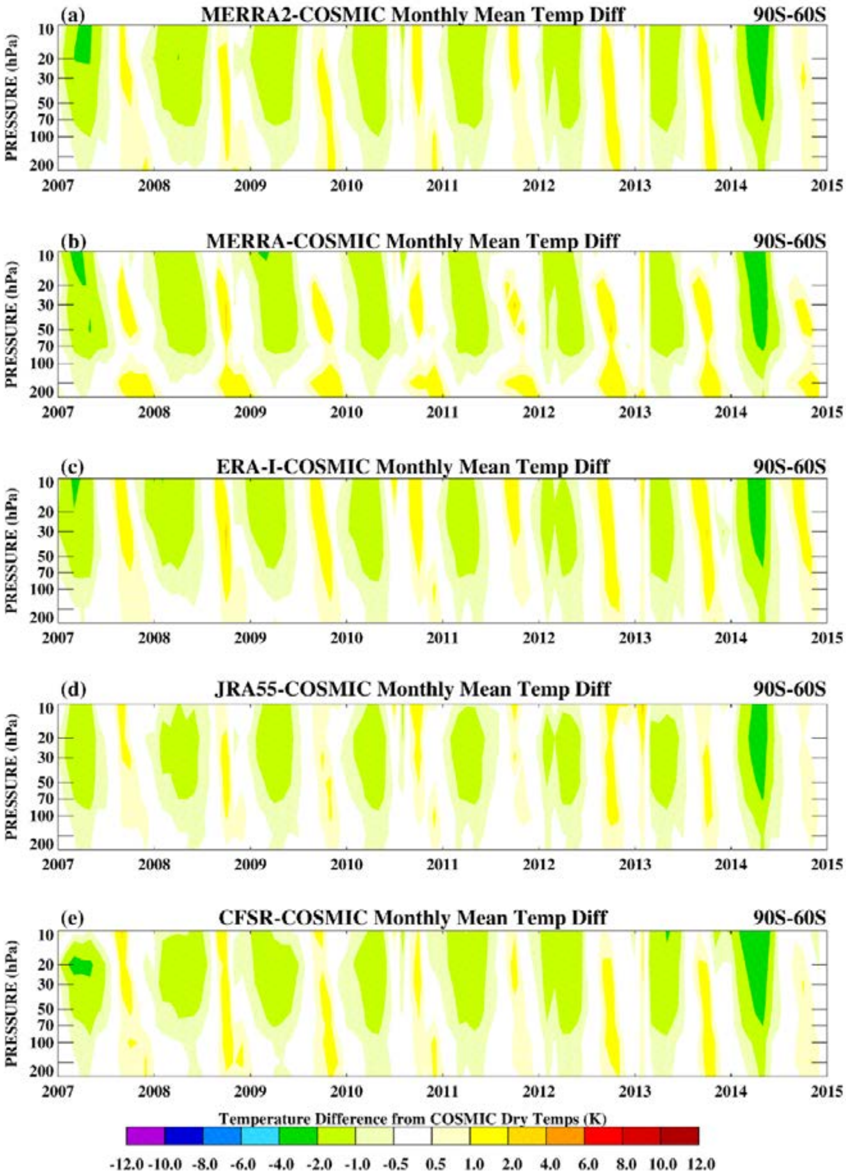


Figure 3.18: *Pressure versus time plots of differences in reanalyses minus COSMIC GPSRO dry temperatures (K) from 2007 to 2014 for the Southern Hemisphere high-latitude zone (90° S - 60° S). The reanalyses are (a) MERRA-2, (b) MERRA, (c) ERA-I, (d) JRA-55, and (e) CFSR/CFSv2. Reproduced from Long et al. (2017).*

Figure 3.19 shows the reanalysis minus COSMIC differences for the NH polar region (60° N - 90° N). Similar negative differences occur during the transition from NH summer to winter. The depth and time length of the -1 K differences are smaller than the SH differences. There are also short-term negative differences that extend from 10 hPa to 100 hPa during the years in which an SSW occurred (2009, 2010, and 2013). In 2009 this is preceded by a short-term (1-month) positive difference also extending from 10 hPa to 100 hPa. The positive differences occur during the months when the SSW produced very warm temperatures in the NH polar region. The negative spikes occurred in the month(s) following the warming when very cold temperatures followed the warming in the upper and middle stratosphere. These differences imply that the dry temperature data set does not capture the maximum warming during the SSW or the cooling which follows. This may be due to the fewer COSMIC observations in the polar region versus the number of observations peaking between 50° N and 60° N in both hemispheres.

Differences between the reanalyses and COSMIC dry temperatures in the tropics (10° S - 10° N) (**Figure 3.17**) show much smaller negative differences. MERRA-2, JRA-55, and especially ERA-I show very few occurrences of differences exceeding -0.5 K.

The few differences with the JRA-55 have a seasonal occurrence from December through February. MERRA, which did not assimilate the GPSRO data, has negative differences fairly consistent between 10 hPa and 30 hPa. CFSR/CFSv2, which did assimilate GPSRO observations, has more occurrences of negative differences than MERRA-2, JRA-55, and ERA-I.

The NH and SH mid-latitudes (not shown) have seasonal differences similar to their respective polar regions but to a smaller time extent and shallower from 10 hPa down into the middle atmosphere. We conclude that between 60° S and 60° N, the lower stratosphere temperatures in the more recent reanalyses and the COSMIC dry temperatures are within ±0.5 K of each other consistently throughout the year.

3.7.3 Atmospheric layer temperature anomalies

Long-term satellite observations from NOAA polar orbiting satellites of temperatures in the lower stratosphere (TLS) are available from the MSU-4 and AMSU-A9 microwave channels, while the Stratospheric Sounding Unit channel 1 (SSU1) and channel 2 (SSU2) provide temperature observations of the middle and upper stratosphere, respectively. *Zou and Qian* (2016) explain the process of merging and extending the infrared-based SSU observations with the microwave-based AMSU-A and ATMS observations. The satellite weighting functions for these three channels can be found in *Chapter 2* (**Figure 2.16**) and *Seidel et al.* (2016, their Figure 1) and on the NOAA STAR SSU website (**http://www.star.nesdis.noaa.gov/smcd/emb/mscat/index.php**). These satellite-observation climate data records have been used to compare with climate model runs to determine whether the model accurately captures the atmospheric vertical temperature changes since 1979 (*Zhao et al.*, 2016). Other studies use these temperature data records to monitor changes in the Brewer–Dobson circulation (*Young et al.*, 2011, 2012). *Randel et al.* (2016) compared global and latitudinal trends from SSU with Aura MLS and SABER temperatures. *Simmons et al.* (2014) discuss the impacts of the MSU, SSU, AMSU-A, HIRS, and AIRS channels assimilated in the ERA-I. *Seidel et al.* (2016) intercompared the TLS trends from three satellite centres for the entire (1979 – 2015) period and separate trends for pre-1997 and post-1997. *Mitchell et al.* (2015) generated TLS and SSU channel-weighted temperatures from reanalyses to see how well they compare with the satellite observations. We perform a similar exercise by applying the TLS, SSU1, and SSU2 weights to the reanalyses temperatures at their standard pressure-level temperatures. **Table 3.2** provides weighting function information about each of the SSU and MSU-4 channels. *Chapter 2* (**Figure 2.16**) shows the vertical extent of the satellite channels. SSU3 layer temperatures were not generated because there were insufficient pressure levels from the majority of the reanalyses to adequately represent this layer in the lower mesosphere. Global mean TLS, SSU1, and SSU2 temperatures are generated for each month from 1979 through 2014. Anomalies from the 30-year period (1981 - 2010) for the TLS, SSU1, and SSU2 are generated. These anomalies are compared against the NOAA STAR SSU v2.0 data set (*Zou et al.*, 2014) and MSU/AMSU mean layer atmospheric temperature v3.0 (*Zou and Wang*, 2012).

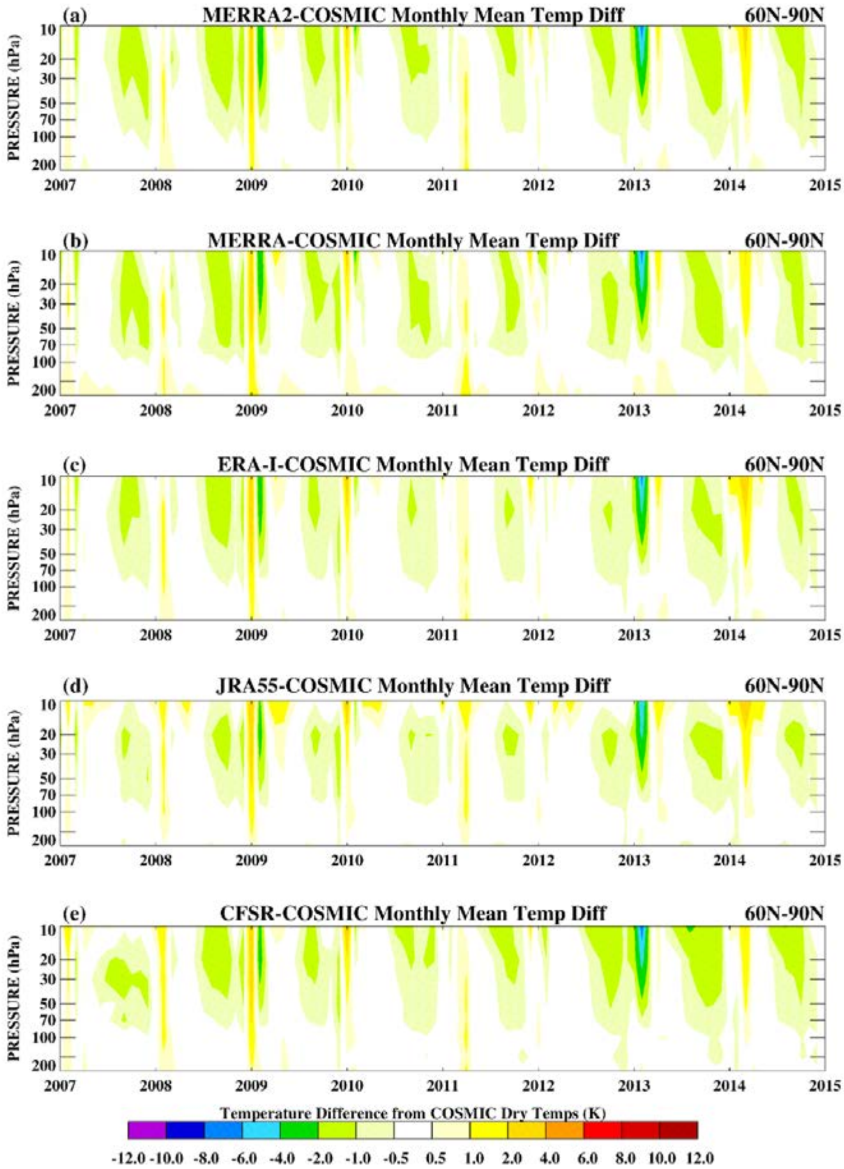


Figure 3.19: *Same as* ***Figure 3.18*** *except for the Northern Hemisphere high-latitude zone (60° N - 90° N). Reproduced from Long et al. (2017).*

The left column of **Figure 3.21** shows the monthly TLS, SSU1, and SSU2 temperature anomalies from the CFSR/CFSv2, ERA-I, JRA-55, MERRA, and MERRA-2 from 1979 through 2014 with the NOAA STAR anomalies overplotted in black. In general, the anomalies show that the layer temperatures were higher in the 1980s than at present. The El Chichón and Mt. Pinatubo volcanic eruptions increased the layer mean temperature by over 1 K from 1982 - 1984 and 1991 - 1993, respectively. Smaller impacts occurred in the SSU1 and SSU2 layer temperatures, as the volcanic influence was mostly in the lower stratosphere. The TLS temperature anomalies show a flat trend between the two volcanoes and after Mt. Pinatubo. The SSU1 and SSU2 temperature anomalies have a persistent cooling trend from 1979 to 2010 and have become flatter since then.

To better assess how each reanalysis differs from the NOAA STAR anomalies, the right column shows the differences in the anomalies of each reanalysis from the NOAA STAR anomalies. The reanalyses TLS anomalies differ from the NOAA STAR anomalies by less than ±0.5 K for most of the time series. Most noticeable is that the ERA-I has smaller anomalies than NOAA STAR in the early 1980s and then has larger anomalies after 2006. Aside from the ERA-I, the other reanalyses seem to agree with the NOAA STAR anomalies during the El Chichón volcanic period (1982 - 1984), with the exception of MERRA and MERRA-2, which have smaller anomalies during the Mt. Pinatubo volcanic period (1991 - 1993). There is a noticeable decrease in the reanalyses anomalies with respect to the NOAA STAR anomalies in 1999 followed by a gradual increase in time until 2006, after which the reanalyses begin to disagree more with each other. GPSRO observations from the COSMIC constellation became available for assimilation in 2006.

Table 3.2: *Pressure (hPa) of SSU channels 1, 2, and 3 and MSU channel 4 weighting function peaks, 50 % of peak weight above, 50 % of peak weight below, 10 % of peak weight above, and 10 % of peak weight below the peak.*

	Peak	50 % above	50 % below	10 % above	10 % below
SSU Ch 3	1.5	0.5	5	0.15	45
SSU Ch 2	3.5	1.5	20	0.30	100
SSU Ch 1	15.0	4.5	60	1.10	150
MSU Ch 4	85.0	35.0	150	15.0	175

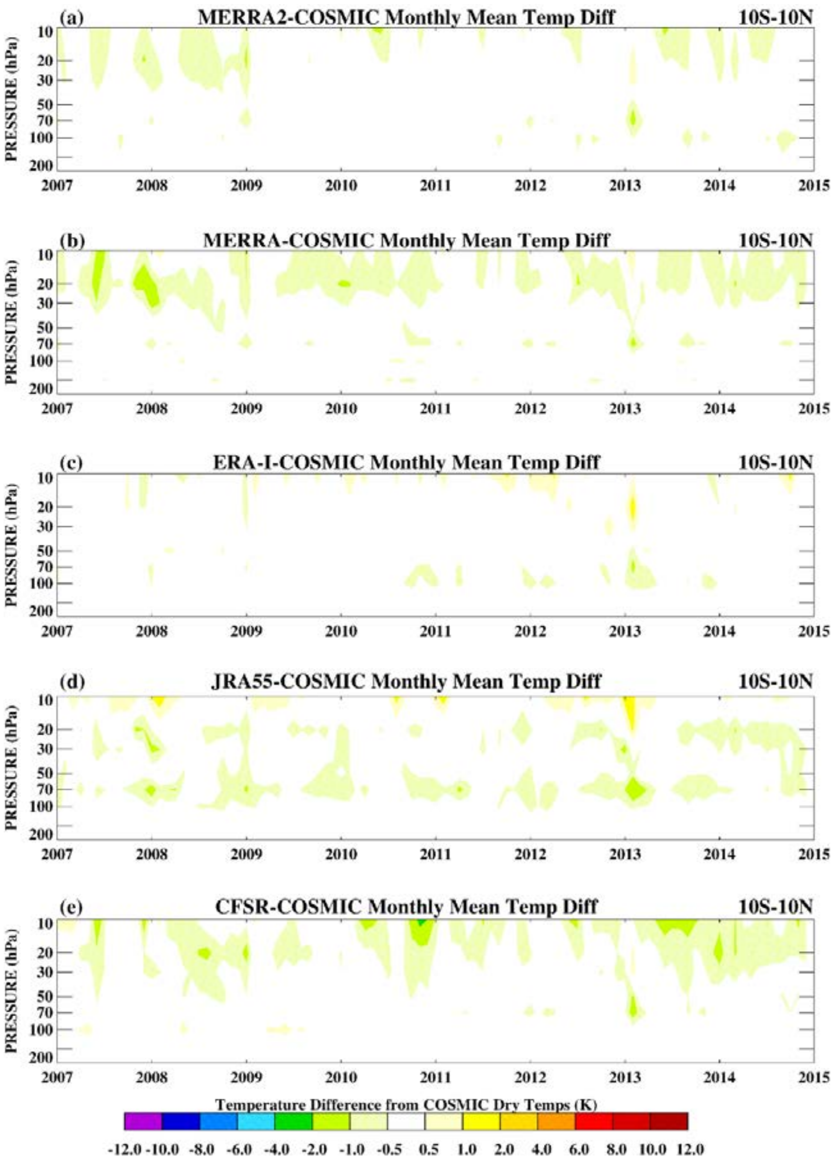


Figure 3.20: *Same as* ***Figure 3.18*** *except for the equatorial-latitude zone (10° S - 10° N). Reproduced from Long et al. (2017).*

The SSU1 temperature anomalies from the CFSR/CFSv2 show large temperature jumps associated with the six streams, preventing any useful evaluation. The other four reanalyses differ from the NOAA STAR by less than ±0.5 K for most of the time series. The ERA-I, MERRA, MERRA-2, and JRA-55 all show smaller anomalies than the NOAA STAR in the early 1980s. There is minor disagreement among the four reanalyses with the NOAA STAR between the late 1980s and the early 2000s. MERRA exhibits two spikes in the SSU1 and SSU2 differences from NOAA STAR. The first spike is a result of missing SSU data from 8 April - 21 May 1996. The second is from a lack of AMSU-A channel 14 data on NOAA-15 from 30 October – 31 December 2000 (*W. McCarthy, personal communication*, 2017). When there are no observations to constrain the model in the upper stratosphere, analyses migrate to the model climatology, which is warmer than the observations. MERRA-2 found the missing SSU observations in 1996 and began using

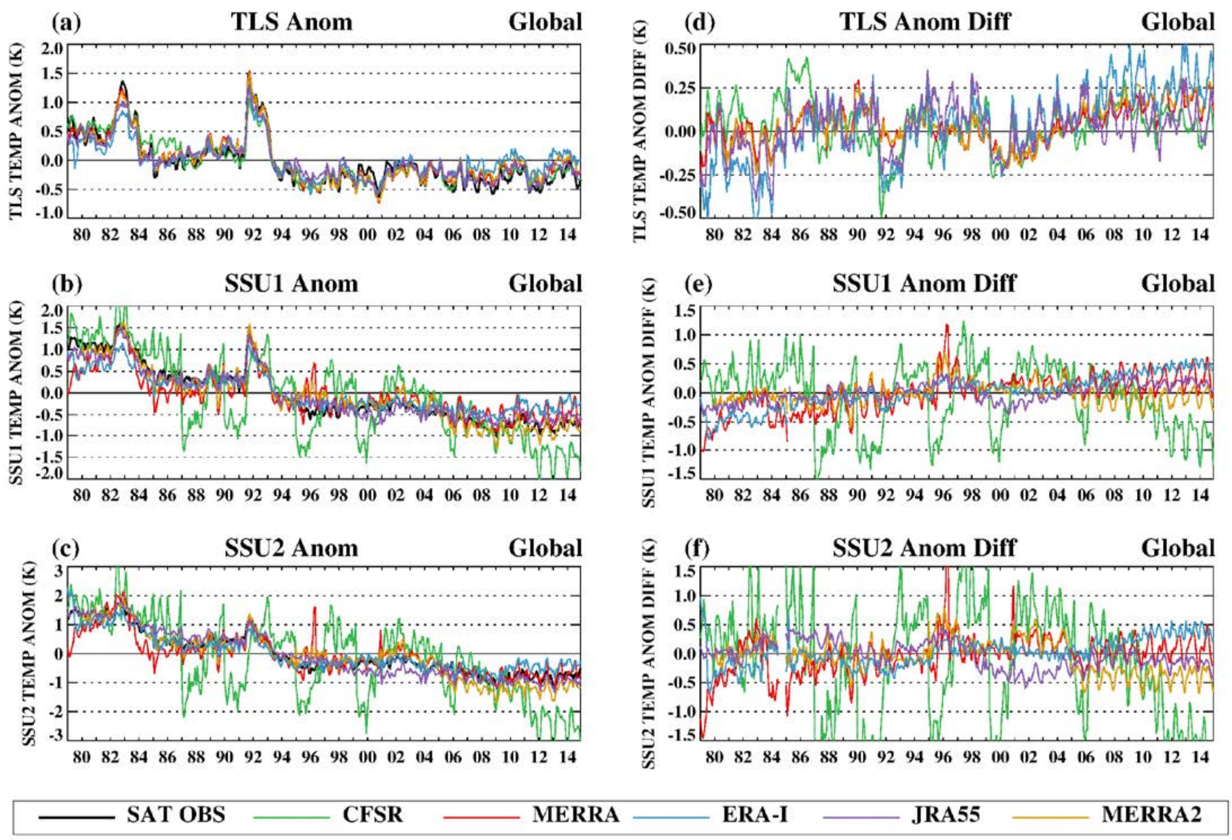


Figure 3.21: *Time series plots of the global layer mean temperature anomalies (K) from the 1981 - 2010 climatology (a - c) and reanalyses anomaly differences from the NOAA STAR anomalies (d - f) for (a, d) the lower stratosphere (TLS) equivalent to the MSU 4 observations, (b, e) the middle stratosphere (SSU1) equivalent to the SSU channel 1 observations, and (c, f) the upper stratosphere (SSU2) equivalent to the SSU channel 2 observations. TLS, SSU1, and SSU2 weights are applied to the MERRA-2, MERRA, ERA-I, JRA-55, and CFSR/CFSv2 pressure-level data to produce layer mean temperatures and anomalies. NOAA STAR TLS, SSU1, and SSU2 anomalies are plotted along with the reanalyses in the left column. Reproduced from Long et al. (2017).*

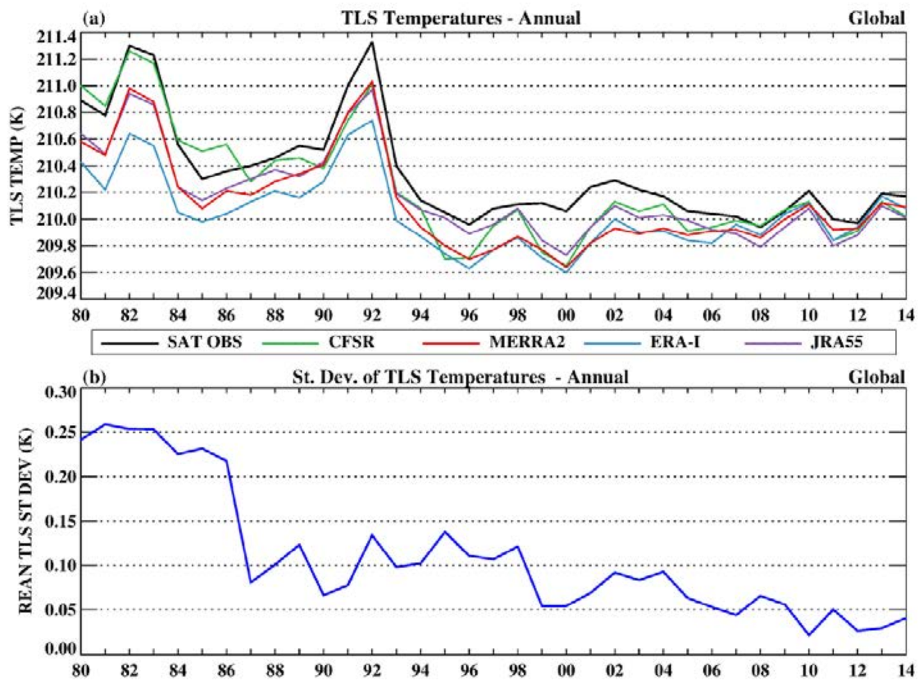


Figure 3.22: *Time series plot of the (a) global annual average of the lower stratospheric temperature layer (TLS) temperatures (°C) for MERRA-2, ERA-Interim, JRA-55, CFSR/CFSv2, and the NOAA STAR TLS CDR. (b) The TLS temperature SD (K) of the four reanalyses for each year. The climatological period spanned from 1981–2010. COSMIC GPSRO observations began to be assimilated in 2006. Reproduced from Long et al. (2017).*

NOAA-16 AMSU-A observations earlier than in MERRA to shrink the gap to just several days. Beginning in 2006, just as with the TLS anomalies, the disagreement among the four reanalyses increases.

Just as with the SSU1 anomalies, the large temperature jumps associated with the CFSR stream transitions prevent a proper evaluation of its SSU2 time series. Aside from the CFSR, the other four reanalyses are within ±0.5 K of the NOAA STAR anomalies. The JRA-55 matches the NOAA STAR SSU2 observations very well throughout the entire time series with the exception of a period in the late 1990s and early 2000s when its anomalies are smaller than the NOAA STAR anomalies. The ERA-I matches the NOAA STAR SSU2 observations very well except after 2006 when it exhibits a positive trend. *Simmons et al.* (2014) state that the use of radiosonde data that are not bias adjusted is the likely cause of this trend. MERRA initially begins with lower SSU2 anomalies than NOAA STAR, whereas MERRA-2 anomalies are much closer to the NOAA STAR anomalies. MERRA-2 separates from MERRA after 2005 with more negative anomalies. This is most likely due to the assimilation of MLS temperatures at pressures less than 5 hPa, which have been shown to produce lower temperatures than before 2005.

The CFSR/CFSv2, JRA-55, ERA-I, and MERRA-2 all use GPSRO observations after 2006, yet the later years in **Figure 3.21** show that their anomalies increasingly disagree with each other after 2006. This apparent larger disagreement is because in **Figure 3.21**, the anomalies are calculated from the climatology of each reanalysis and because the climatology differs for different reanalyses quantitatively; thus, **Figure 3.21** may give us wrong impression in terms of the actual differences among the reanalyses. **Figure 3.22a** presents the actual TLS temperatures for these

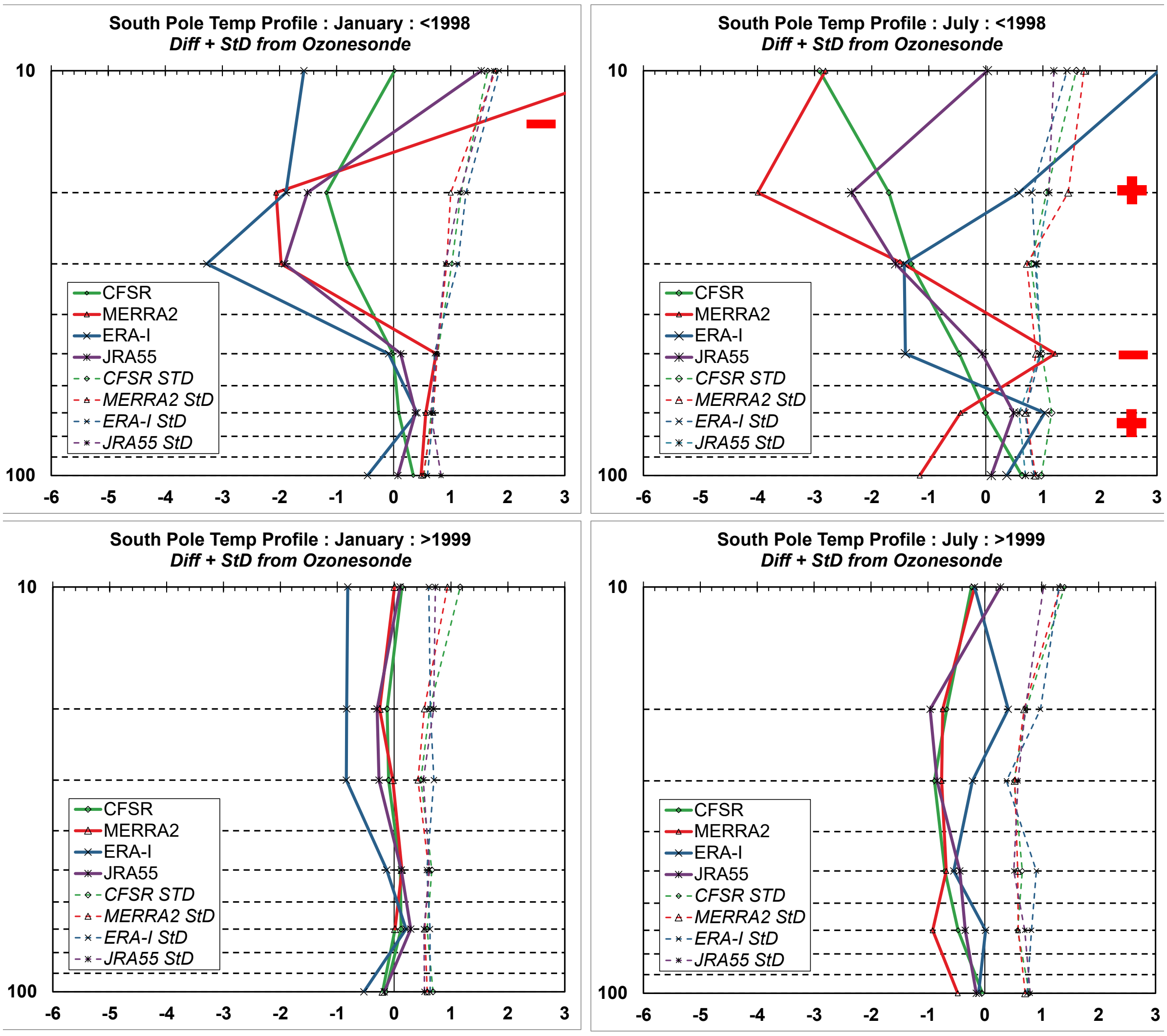


Figure 3.23: *Mean vertical profile temperature differences (solid) and their variability (dashed) of four reanalyses from South Pole ozonesonde temperatures for the months of January (left) and July (right) separated into pre-1998 (top) and post-1999 (bottom) periods. The reanalyses are CFSR/CFSv2, MERRA-2, ERA-I, and JRA55. Note that in July during the pre-1998 period that ERA-I and MERRA-2 have opposite sign differences from the ozonesonde temperatures (see the plus and minus symbols for [ERA-I minus MERRA-2]on the panel). Note also that in the post-1999 period all of the reanalysis mean differences and their variability are much smaller than during the pre-1998 period.*

four reanalyses over time from 1980–2014. There is a large spread in the TLS temperatures of 0.8K between the coldest TLS temperature (ERA-I) and the warmest TLS temperature (CFSR/CFSv2). Over time this large spread decreases until the difference is less than 0.1K. This illustrates how the various reanalyses actually approach agreement of the TLS values as more observations are assimilated. **Figure 3.22b** presents the SD of the four reanalyses TLS temperatures over time. There is a large decrease from 1986 to 1987, which is attributed to the CFSR/CFSv2 TLS values cooling during the transition from its initial stream to its second. Another drop in 1999 follows the availability of ATOVS in **Figure 3.21**; the quality and character of the temperature values between 1981 and 2010 changed. This makes generating long-term climatology and anomalies misleading.

Similar comparisons of the SSU1 and SSU2 temperatures are not presented as the temperature biases of each reanalysis above 10hPa prevents agreement in the layer mean temperature. This shows the value of the GPSRO data to anchor the temperatures in the middle and lower stratosphere, which is where most of the TLS weighting function occurs.

3.7.4 Ringing of SH polar latitude temperature differences

MERRA-2 and ERA-I SH polar temperatures at 50hPa agree well (less than ±0.25K) during September in the ATOVS period but have about a 1K difference in the TOVS period. JRA-55 and CFSR/CFSv2 also show this 1K colder temperature difference from MERRA-2 during the TOVS period. An examination of other months reveals that this difference exists during the SH winter months. Examining the [ERA-I – MERRA-2] differences during the TOVS period at other levels reveals that an oscillating (ringing) set of differences in the SH polar latitudes starts with positive differences at 100hPa; negative differences at 50hPa; positive differences at 10-20hPa; negative differences at 3hPa; and positive differences at 1hPa. This 'ringing' of temperature differences is much smaller during the SH summer months and just partially occurs in the NH polar latitudes with positive differences at 5hPa and negative differences at 2hPa.

Figure 3.23 shows the differences of MERRA-2 and ERA-I (as well as JRA-55 and CFSR/CFSv2) from the South Pole ozonesonde temperatures (obtained from the NDACC) in July and January separating the TOVS period (1979-1998) and ATOVS period (1999-2014). When one has positive differences from the ozonesonde temperatures the other has negative differences. The maxima of these differences occur from 100hPa-70hPa, at 50hPa, and 20hPa-10hPa. CFSR/CFSv2 and JRA-55 do not show an oscillating pattern in their differences from the ozonesonde temperature, but rather an increasing negative difference with altitude from 100hPa to 10hPa. During January all four reanalyses agree well with the South Pole ozonesonde temperatures from 100hPa up to 50 hPa. At higher altitudes, all four show negative differences at 30hPa and 20hPa. During the ATOVS period, differences from the South Pole ozonesonde temperatures are much smaller (-1K to 0K) during July and even smaller (except for ERA-I which shows a consistent cold bias) during January.

The likely candidate for this 'ringing' during the TOVS period is how the reanalyses assimilate the three broader SSU channels vs. the narrower five AMSU channels (10-14) during the ATOVS period and how the assimilation systems handle the propagation of errors in the vertical.

3.8 Comparisons against other observations

3.8.1 Ozonesonde temperatures

Most radiosonde temperature observations are assimilated by the reanalyses. This makes comparisons of reanalyses with radiosondes problematic since the reanalyses should differ least at these observation points. However, temperature profiles accompanying ozonesondes are not assimilated by the reanalyses and thus are a viable source for comparison in the lower and middle stratosphere. Four ozonesonde locations (Ny Ålesund, 79°N, 12°E, beginning 1992; Hohenpeissenberg, 48°N, 11°E, beginning 1987; Lauder, 45°S, 170°E, beginning 1986; and Neumayer, 70°S, 8°W, beginning 1992) obtained from the NDACC were chosen to compare their temperature profiles with that from reanalyses. These locations had the greatest longevity and consistency in ozonesonde measurements. **Figure 3.24** shows that there is nearly a 30% reduction of ozonesonde flights from 10hPa to 7hPa. An additional 50% do not reach 5hPa, and only a few reach 3hPa. This makes analysis of differences with these radiosonde temperatures above 10hPa impractical. **Figure 3.25** presents the annual mean differences and standard deviations of CFSR/CFSv2, MERRA, and ERA-I from the radiosonde observations upward from 100hPa. Of the four locations, only Lauder shows substantial differences from the observations from 100hPa to 10hPa. The variability of the differences is consistently about 2K

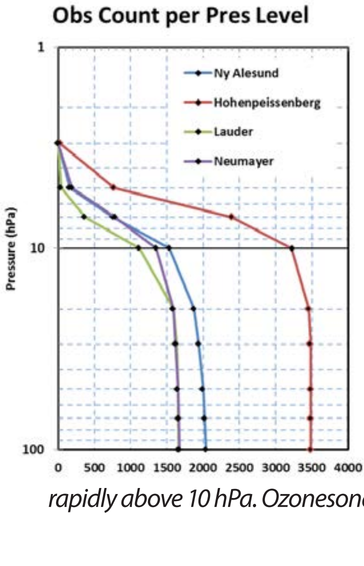


Figure 3.24: *Total number of temperature observations made at each pressure level in the lower and mid-stratosphere at Ny Ålesund, Hohenpeissenberg, Lauder, and Neumayer since observations started at each site. Hohenpeissenberg has a higher frequency of ozonesonde launches per month than the other three sites. Note that the number declines rapidly above 10 hPa. Ozonesonde data is from the NDACC.*

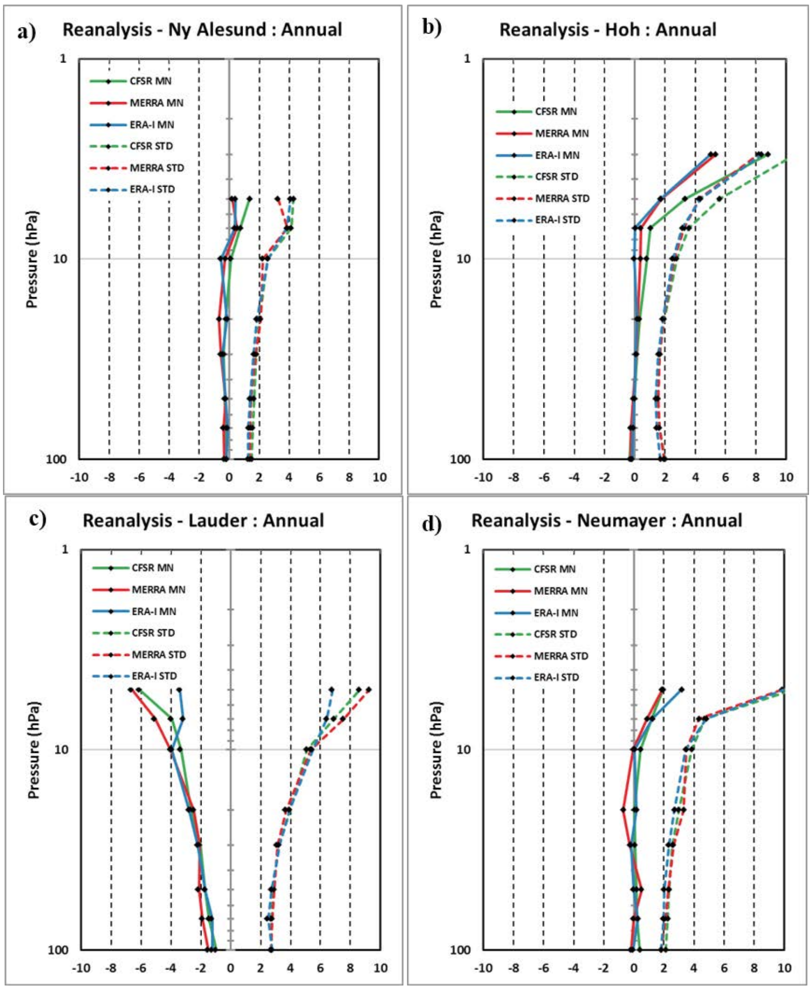


Figure 3.25: *Mean (solid line) and standard deviation (dashed line) of temperature differences (°C) of three reanalysis (CFSR/CFSv2, MERRA, and ERA-I) from ozonesonde observed temperatures at a) Ny-Ålesund, b) Hohenpeissenberg, c) Lauder, and d) Neumayer. Ozonesonde data is from the NDACC.*

at all altitudes at Ny Ålesund and Hohenpeissenberg. At Neumayer the difference variability is 2 K at 100 hPa, but increases to 4 K at 10 hPa. At Lauder, the variability of differences is nearly 3 K at 100 hPa and grows to nearly 6 K at 10 Pa. The relatively small mean difference and variability of differences of the reanalyses temperatures from observations at Ny Ålesund and Hohenpeissenberg indicate that over the northern hemisphere land masses the reanalyses temperatures are quite good. The large differences from the Lauder observations could be attributed to poorer analyses by the reanalyses since few radiosonde observations are available to be assimilated in the SH middle latitudes. The larger variability of differences at Neumayer indicates increased uncertainty of the reanalyses over Antarctica.

3.8.2 Long-duration balloon observations

Long duration balloons (LDB) are closed, non-expansive, super-pressure, plastic balloons capable of performing 'horizontal' soundings in the atmosphere. They remain on a constant density surface and are advected by the winds. A balloon's 'lifetime' is limited only by leaks of the lifting gas, atmospheric energy, and political or safety considerations. The first LDB were used by the US Navy to collect meteorological data at 300 hPa upwind of the continental US over the Pacific Ocean. Each balloon flight lasted several days and they were tracked by radio triangulation (*Angell*, 1960). Other LDB programs include:

- The NCAR "GHOST" program from 1967-71. This consisted of 60 flights at pressure levels between 200 hPa and 100 hPa. The emphasis was towards monitoring the circulation in the Southern Hemisphere upper troposphere (Angell, 1972).

- The French-US "Eole" Program from August 1971 to December 1972. This program launched 480 balloons taking over 80,000 observations (or more than 600 observations per day) near the 200 hPa pressure level in the SH mid and high latitudes. Satellites were used to obtain several positions per day for each balloon (*Hertzog et al.*, 2006). Along with winds, the balloons provided temperature and pressure observations.

- The Vorcore Campaign from September 2005 to February 2006 over Antarctica. This program included 27 balloons that flew between 60 hPa and 80 hPa. GPS positioning was used to determine the winds and pressure and temperature observations were communicated with the ground station. Observations were determined at a frequency of 15 minutes (*Boccara et al.*, 2008).

- The French Space Agency (Centre National d'Etudes Spatiales (CNES)) Pre-Concordiasi took place in February 2010 and consisted of three flights, lasting about three months. Observations were taken every 30 seconds of winds, pressure and temperature in the equatorial upper troposphere – lower stratosphere (UTLS) or Tropical Tropopause Layer (TTL) (*Podglajen et al.*, 2014).

The Eole, Vorcore, and Pre-Concordiasi observations of temperature and winds have been compared to ERA-40 and *R1*, ECMWF operations and *R1*, and ECMWF operations, ERA-I and MERRA, respectively. None of these balloon observations were assimilated by the reanalyses or ECMWF operations. The Eole observations showed that both ERA-40 and *R1* were warmer than the balloon observations in the subtropics but colder at higher latitudes. Eole observations also showed that the analysis of meteorological wind fields over the open oceans was much better in *R1* than in ERA-40. *Boccara et al.*, 2008 noted that over Antarctica during the 2005 austral spring the "ECMWF analyses were found to agree closely with the observations with virtually no bias on the zonal and meridional velocities and a small cold bias (-0.42 K) on the temperature. The velocities from the NCEP-NCAR reanalysis [equivalent to *R1*] reanalyses are also very close to the balloon observations although they exhibit larger dispersion. Overall, R-1 displayed a strong warm bias (+1.51 K)". The Pre-Concordiasi observations showed that the scarcity of upper air wind observations by radiosonde or aircraft in the Tropical Tropopause Layer impacted the reanalyses' ability to properly resolve small and mesoscale motions in the wind fields.

3.8.3 Rocketsonde observations

A four decade analysis of radiosondes (1969-present), M-100 rocketsondes (1971 - 1991), RH-200 rocketsondes (2002 - present) and high altitude radiosondes (1969 - present) taking measurements on a weekly to two-week frequency has been performed by *Das et al.* (2016). Observations from all the above were taken at the Thumba Equatorial Rocket Launching Station (TERLS), India (8.5°N, 76.9°E). M-100 rocketsonde observations between 1969 - 1991 at Volgagrad, Russia (49°N, 44°E), Heiss Island (81°N, 58°E), and Molodezhnaya (68°S, 46°E) are also available for analysis. Monthly means were determined between 1970 and 2014. Comparisons were generated between the rocket and radiosonde observations and *R2*, ERA-40, ERA-I, and MERRA for zonal and meridional wind. Comparisons of winds were generated from the surface up to 30 km (~10 hPa) from radiosondes and 30 km to 65 km (~10 hPa - 0.1 hPa) from rocketsondes. RMS errors of the reanalyses zonal wind was about 2 m s^{-1} below 10 km, growing to about 4 m s^{-1} at 30 km, continuing to increase to 6 m s^{-1} by 65 km. Meridional winds had an RMS difference of about 2 m s^{-1} up 10 km, peaking in the troposphere at 15 km at about 2.5 m s^{-1}, declining to 2 m s^{-1} at 20 km, then slowly increasing to 6 m s^{-1} at 50 km. Filtering the observed and reanalysis zonal winds into their annual oscillation, SAO, and QBO components show that the annual oscillation, QBO, and SAO are all reproduced by the reanalyses and their structures in the vertical are comparable with the observations at Thumba. However, MERRA's difference with the annual oscillation observations increase from 0 m s^{-1} to 10 m s^{-1} from 50 km to 65 km, and the SAO winds above 40 km are overestimated by MERRA by 5 m s^{-1}. The QBO amplitude growth from 20 km to 25 km is captured by all the reanalyses, but their amplitude with respect to observations is underestimated between 25 km to 30 km.

3.9 Effects of volcanic eruptions and other natural variabilities

A paper by *Mitchell et al.* (2015) examined nine reanalyses to isolate zonal-mean temperature and zonal wind signatures of variability in the stratosphere and troposphere associated with the two volcanic eruptions (El Chichón, and Mt Pinatubo), El Niño–Southern Oscillation (ENSO), the QBO of the equatorial zonal wind, and the 11-year solar cycle. *Mitchell et al.*(2015) also examined in greater detail the seasonal evolution of the 11-year solar cycle signal that operate in the stratosphere and may penetrate downward to influence the troposphere. Employing a multiple linear regression technique with no time lags, *Mitchell et al.* (2015) found that the characteristic signals of all four sources of variability were very consistent between each of the reanalyses over the 1979 - 2009 period. They found that ENSO imparts a high temperature anomaly in the equatorial troposphere and strengthens the westerly winds in the subtropics. They also found that ENSO has an influence upon the Southern Annular Mode (SAM) with associated zonal wind anomalies in the SH mid-latitude and sub-polar regions. Influences in the stratosphere were also found that imparted a temperature anomaly above the tropical tropopause. This may impact the lower region of the QBO. There is also a warm winter anomaly in the NH polar mid / upper stratosphere that is indicative of increased wave propagation into the winter stratosphere during strong El Niño events. The consensus volcanic response is a warm anomaly in the tropical lower stratosphere, cool anomaly in the upper tropical stratosphere, and wide spread cooling in the troposphere. There is a triple temperature and wind anomaly (positive/negative/positive) response over the equator associated with the QBO. The upper positive response indicates that the QBO has an influence upon the SAO. The 11-year solar cycle response is weaker and less statistically significant than the atmosphere's response to other forcings. A tropical warm anomaly under solar maximum conditions may influence planetary wave propagation toward the poles which is most apparent during the winter months.

Furthermore, *Fujiwara et al.* (2015) analysed the volcanic temperature responses to the 1982 El Chichón and 1991 Mount Pinatubo eruptions individually using nine reanalysis data sets (JRA-55, MERRA, ERA-I, CFSR/CFSv2, JRA-25, ERA-40, *R1*, *R2*, and 20CR). They found that the latitude-pressure distribution of volcanic temperature responses was different for different eruptions, but was quite similar at least among the recent four reanalysis data sets (JRA-55, MERRA, ERA-Interim, and CFSR/CFSv2) for each eruption. The *R1*, *R2*, and JRA-25 showed different tropical stratospheric signals particularly for the El Chichón eruption, though the original upper-air temperature observations assimilated are basically common, and this is most probably in association with the use of older analysis systems. The 20CR did not assimilate upper-air observations and gives very different volcanic signals, despite including volcanic aerosols in the forecast model (this is in part due to unknown warming signals in 20CR in 1989 and in 1990 that raised the 36-month averaged base in the volcanic signal definition). They also analysed the response to the 1963 Mount Agung eruption using JRA-55, ERA-40, *R1*, and 20CR, and concluded that the JRA-55 data set is probably the most ideally suited for studies of the response to the Mount Agung eruption because it is the only data set that employs the most recent reanalysis system.

3.10 Summary and conclusions

In this chapter a comparison of monthly zonal mean temperatures and zonal winds from the five more recent reanalyses and several older reanalyses were evaluated and intercompared. Our initial evaluation was to look for temperature discontinuities in the time series of each of the reanalyses. This showed that the earlier reanalyses (ERA-40 and JRA-25) had multiple temporal discontinuities in the 1980s in the stratosphere associated with changes in the biases of the data from the NOAA TOVS and SSU instruments. The *R1* and *R2* did not show such discontinuities because they used NESDIS-generated temperature profiles, not the original radiance data. NESDIS most likely strived to minimize such discontinuities in the profile temperatures. Almost all the reanalyses have a temporal discontinuity in 1998 when the ATOVS observations became available and the reanalyses either switched immediately or transitioned from the TOVS to the ATOVS over several years. The CFSR has temporal discontinuities at the time of switching from one stream to the next. The CFSR bias corrected the top SSU channel 3. The model used by the CFSR had a warm bias in the upper stratosphere and slowly warmed about 5 K during the course of each stream. Because of the presence of the discontinuities and transitions discussed above, great caution should be exercised in using reanalyses for trend analysis and/or trend detection, especially in the middle and upper stratosphere.

So as not to favour any one particular reanalysis, a reanalysis ensemble mean (REM) of three of the more recent reanalyses (MERRA, ERA-I, and JRA-55) was generated. We presented the climatological mean (1981 - 2010) of the temperature and zonal wind REM and showed the altitudes and seasons with the largest variance in the REM. The temperature and zonal winds have the greatest inter-annual variability in the NH polar region from January through March because of the large variability in wave activity, including the frequent occurrence of strong stratospheric warming events. This variability is greatest in the upper stratosphere as planetary-scale wave amplitudes and the associated temperature and zonal wind changes during strong stratospheric warming events are largest in the upper stratosphere. In the SH polar region, the inter-annual variability is not as large in magnitude and is prevalent throughout the stratosphere. Because midwinter wave activity is much smaller in the SH, most of the inter-annual variability in the SH polar region is associated with the springtime transition to summer circulation patterns and polar vortex breakdown when wave activity shows larger inter-annual variability in timing and magnitude.

Time series of the temperature variance in the three REM members showed that the greatest disagreement occurs during the TOVS time period (1979 – 1998) in all latitude zones, and agreement improves during the ATOVS time period (1999 to present). The disagreement in the SH polar latitudes extended lower into the stratosphere than in the NH polar latitudes. The zonal wind variance was smaller than the temperature variance in the polar latitudes, but had a similar temporal difference between the TOVS and ATOVS time periods. In the tropics, the zonal wind variance was much larger than in the polar regions as the disagreement of the SAO and QBO zonal winds was quite large. Thus, improving equatorial winds in future reanalyses is an important goal.

The characteristics of each reanalysis were identified as differences from the temperature and zonal wind REM. The CFSR/CFSv2 had a seasonal warm bias compared to the REM in the upper stratosphere that persisted during both the TOVS and ATOVS time periods. The JRA-55, on the other hand, had a seasonal cold bias that persisted during both the TOVS and ATOVS time periods. ERA-I and MERRA had smaller differences from the temperature REM except that the ERA-I had a warm bias in the SH polar latitudes between 7 hPa and 5 hPa that occurred only during the austral winter and only during the TOVS time period. MERRA-2 had very small differences from the REM except in the upper stratosphere in the polar regions where it had a year-long cool bias at 1 hPa and a warm bias between 2 hPa and 3 hPa. These biases greatly diminished during the ATOVS period. Temperature differences from the REM in the earlier reanalyses (JRA-25, ERA-40, and *R1*) extended throughout the stratosphere and the upper troposphere. These differences occurred through both the TOVS and ATOVS time periods. This illustrates the progress made by the reanalysis centres to improve the analyses from the earlier versions to the later versions. This results in better agreement among the more recent reanalyses.

In the tropics, the individual reanalyses exhibited smaller temperature differences than in the polar latitudes. However, the characteristic biases in the upper stratosphere observed in the polar latitudes were maintained in the tropics. The zonal wind differences from the REM of the individual reanalyses are very large in the SAO region. In the QBO region the differences frequently show dissimilarities in the timing of the descending westerlies and easterlies as well as the amplitude of these winds. Zonal wind differences from the REM were not confined to the stratosphere as several reanalyses also had sizable differences in the troposphere.

Specifically comparing the more recent reanalyses QBO zonal winds (70 hPa – 10 hPa) against the zonal winds observed at Singapore using the FUB data set showed that the CFSR/CFSv2 had the largest RMS differences from the Singapore winds than the other reanalyses at most levels and during both the TOVS and ATOVS periods However, MERRA-2 10 hPa zonal winds were nearly twice as large as the other reanalyses during the TOVS period, mostly due to an overly aggressive gravity wave parametrisation. The RMS differences from the Singapore zonal winds were smaller during the ATOVS period for all the reanalyses. The CFSR/CFSv2 had the largest amplitude biases from the Singapore winds as shown by the linear slope of their matched monthly values. The linear slopes of all the reanalyses were furthest from unity at 50 hPa and 30 hPa during the TOVS period.

There are several reasons why the ATOVS period is an improvement over the TOVS period. The primary reason is that the AMSU-A instrument has five narrower channels in the stratosphere instead of the broader three SSU channels. (The MSU channel 4 and AMSU-A channel 9 weighting functions are almost identical.) Another reason is that the SSU was the only instrument monitoring the thermal structure of the stratosphere from 1978 through 1998. From 1999 onward, there are additional satellite instruments monitoring the stratosphere: AIRS, IASI, MLS, and GPSRO. Hence the quantity and quality of data monitoring in the stratosphere increases from 1999 to the present.

The amplitude of the annual temperature cycle (warmest summer month minus the coldest winter month) in the SH polar latitudes is larger than the NH polar latitude temperature amplitude by 5 - 15 K. The region of large amplitude extends throughout the middle and upper stratosphere in the SH polar latitudes. In the NH polar latitudes, the vertical region of large temperature amplitudes is confined to the upper stratosphere and occurs during the years with an SSW. The ERA-I has a noticeably smaller annual temperature amplitude in the SH polar latitudes than the other ensemble members from 3 hPa to 30 hPa. This is due to its warm bias during the SH winter months in this latitude region. The CFSR/CFSv2 temperature amplitude decreases rapidly above 3 hPa due to its warm bias in the upper stratosphere in both SH and NH polar latitudes.

Comparisons against HIRDLS (January 2005 - March 2008) and Aura MLS (2005 - 2014) temperatures concur with the previous characteristics of the various reanalyses in the upper stratosphere. The CFSR has a definite warm bias compared to HIRDLS temperatures, while the JRA-55 has a definite cold bias. Both MERRA and ERA-I have a slight warm bias during the summer months between 3 hPa and 7 hPa. MERRA has a slight cold bias above this between 1 and 2 hPa nearly all year long. MERRA-2 assimilates Aura MLS temperatures at pressures less than 5 hPa and consequently differences are very small.

The NOAA STAR TLS, SSU1, and SSU2 data sets (*Zou and Qian*, 2016; *Zou et al.*, 2014) are a much-improved CDR than the version used in Thompson *et al.* (2012), which pointed out the dissimilarities between the NOAA and Met Office SSU data records. The comparison between the version used in this chapter and the appropriately weighted reanalyses is much better than previous papers using the older version and the Met Office CDR. All of the more recent reanalyses capture the characteristics of the NOAA STAR TLS anomalies. Excluding the CFSR/CFSv2, the other reanalyses (MERRA-2, MERRA, ERA-I, and JRA-55) capture the basic features of the SSU1 and SSU anomalies. We learn from this intercomparison that the GPSRO observations provide an anchor that drives the reanalyses to closer agreement in the middle and lower stratosphere. We also learn that using a long period climatology may not be the best practice to generate anomalies in parts of the atmosphere which are more sensitive to the changes in data sources, which impacts their quality and accuracy over time.

Temperature soundings at four ozonesonde locations extending back to the late 1980's and early 1990's were used to compare against CFSR/CFSv2, MERRA, and ERA-I. The long-term mean differences show that the reanalyses do well below 10 hPa in the NH high (Ny Ålesend) and mid latitudes (Hohenpeissenberg). Similar mean temperature differences were observed over Antarctica (Neumayer) but with increasing standard deviation with height. Larger mean differences and standard

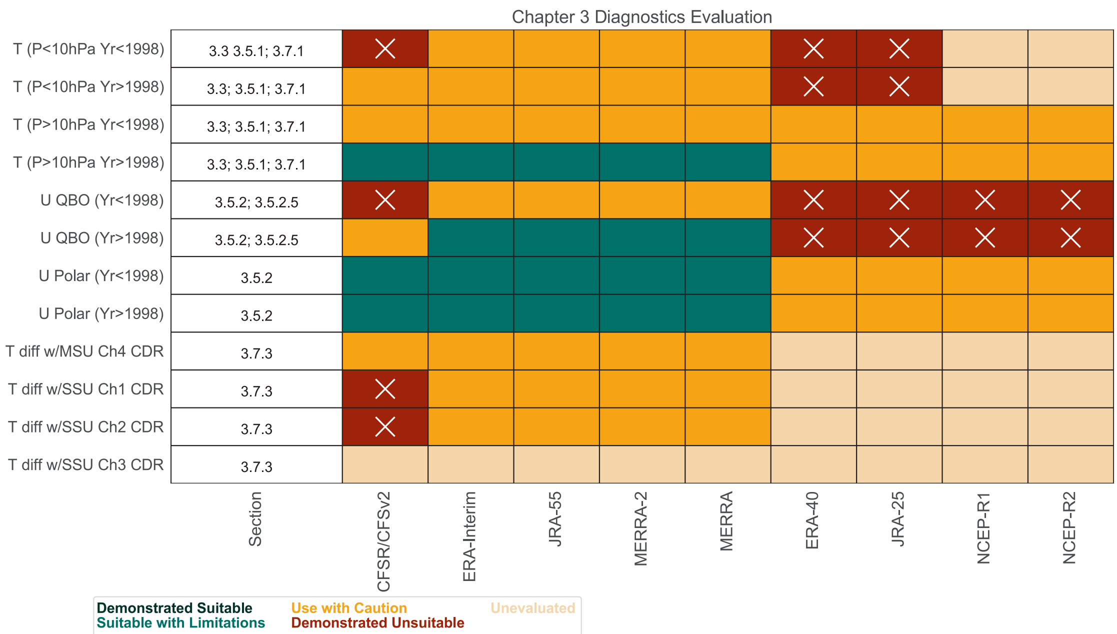


Figure 3.26: *Evaluation of specified diagnostics for each reanalysis. Four evaluation characterizations are provided: "Demonstrated Suitable" (dark green), "Suitable with Limitations" (light green), "Use with Caution" (yellow), and "Demonstrated Unsuitable" (red). Note that the score corresponding to "demonstrated suitable" was not assigned to any of the diagnostics listed here, so the darkest green colour does not appear in this table. Diagnostics relate to: Use of temperatures above/below 10 hPa and before/after 1998; QBO zonal winds and polar zonal winds before/after 1998; and temperature layer differences from the Climate Data Record (CDR) for MSU channel 4, and SSU Channels 1, 2, and 3. Note that the score coresponding to "Demonstrated Suitable" (dark green) was not assigned to any reanalysis for any diagnostic.*

deviations with height were observed over Lauder.

Additional studies extending over several decades have examined other aspects of the dynamical features of reanalyses in the upper troposphere and stratosphere. These studies have used long duration balloons at multiple geographic regions of the globe (*Podglajen et al.*, 2014; *Baccara et al.*, 2008; *Hertzog et al.*, 2006) and historical radiosonde and rocketsonde flight information over India and Russia (*Das et al.*, 2016). Using linear regression techniques, *Mitchell et al.*, (2015) examined reanalyses for impacts on both hemisphere's annular modes and wave activity from multiple sources of variability (ENSO, QBO, volcanoes, and the solar cycle). *Fujiwara et al.* (2015) used multiple reanalyses to examine their temperature response to the El Chichón and Mt. Pinatubo volcanic eruptions.

3.11 Key findings and recommendations

In this chapter we have examined the thermal and dynamical characteristics of the older and the more recent reanalyses. A summary of the diagnostics evaluated in this chapter is provided in **Figure 3.26**. This figure contains assessments of the reanalysis representation of key diagnostics related to temperature and winds and directs the reader towards the appropriate chapter section for further information.

Key Findings:

- More recent reanalysis from all centres are better than their previous version (*e.g.*, JRA-55 vs. JRA-25; MERRA-2 vs. MERRA).
- Due to changes in available data sources, drifts and jumps in the long-term temperature time series can occur. These irregularities are greatest above 10 hPa. Greatest caution is advised when determining trends with reanalysis temperature data sets above 10 hPa.
- The more recent reanalyses have fewer discontinuities in their temperature and wind time series due to better data assimilation techniques and transition among different sets of observations.
- The transition from the TOVS to ATOVS satellite periods, starting around 1998 - 1999, is problematic for all reanalyses. In the stratosphere, the transision from three broad SSU IR channels to 5 narrower AMSU/ATMS microwave layers proves to be problematic for data assimiation.
- The more recent reanalyses agree quite well with each other in the lower and middle stratosphere. All reanalyses have greater differences in the upper stratosphere and lower mesosphere. The latter discrepancies result from differences in model top, vertical resolution, data assimilation techniques and data that is assimiated. *Chapter 2* provides detailed information about each reanalyis model's structure and physics.
- Temperature biases exist among the various reanalyses in the UTLS especially before 1998. Temperatures do not harmonize until GNSS-RO observations are used to lock in the temperatures after 2005.
- The reanalysis QBO winds show improvement over time. Separating them into the TOVS and ATOVS periods, the ATOVS period agree much better with the Singapore radiosonde observations than during the TOVS period. We expect that future reanalyses will have QBO winds that agree with observations as their forecast models improve to produce a spontaneous QBO in the tropics.

Recommendations:

- Users of any reanalysis should proceed with greatest caution when intercomparing reanalyses, and particularly when attempting to detect trends and/or changes in climate above the tropopause.
- Improving the TOVS time period would be highly beneficial to future reanalyses. However, the TOVS time period may never be as good as the ATOVS period due to the sparsity of data.
- Model improvements, improvements to the variational bias corrections to handle the broad SSU weighting functions, and non-orographic gravity wave parametrisation improvements (so the forecast models can generate a QBO on their own) are some of the ways the TOVS time period can be improved upon.
- It may benefit each 'satellite era' reanalysis to begin their reanalysis several years earlier using just conventional data. This most likely will help harmonize the reanalyses' temperature structure below 10 hPa at the start of assimilating satellite data.

Acknowledgements

We thank Michael Schwartz for providing the MLS monthly temperature data. MF's contribution was financially supported in part by the Japan Society for the Promotion of Science (JSPS) through Grants-in-Aid for Scientific Research (26287117 and 16K05548), and Corwin Wright's was supported by NERC grant NE/K015117/1.

References

Andrews, D.G., J.R. Holton, and C.B. Leovy, 1987: Middle Atmosphere Dynamics. *Academic Press*, 489pp, doi:10.1002/qj.49711548612.

Angell, J.K., 1960: An analysis of operational 300-mb transosonde flights from Japan in 1957-58. *J. of Meteorol.*, **17**, 20 - 35, doi: 10.1175/1520-0469(1960)017<0020:AAOOMT>2.0.CO;2.

Angell, J.K., 1972: Some climatological aspects of the circulation in Southern Hemisphere temperate latitudes as determined from 200-millibar GHOST balloon flights. *Mon. Wea.Rev.*, **100**(2), 107 - 116, doi: 10.1175/1520-0493(1972)100<0107:SCAOTC>2.3.CO;2.

Baldwin, M.P. and L.J. Gray, 2005: Tropical stratospheric zonal winds in ECMWF ERA-40 reanalysis, rocketsonde data and rawinsonde data, *Geophys. Res. Lett.*, **32**, L09806, doi: 10.1029/2004GL022328.

Baldwin, M.P., *et al.*, 2001: The quasi-biennial oscillation, *Rev. Geophys.*, **39**, 179 - 229, doi: 10.1029/1999RG000073.

Boccara, G., A. Hertzog, C. Basdevant, and F. Vial, 2008: Accuracy of NCEP/NCAR reanalyses and ECMWF analyses in the lower stratosphere over Antarctica in 2005. *J. Geophys. Res.*, **113**, D20115, doi: 10.1029/2008JD010116.

Charlton, A.J. and L.M. Polvani, 2007: A new look at stratospheric sudden warmings, Part I: Climatology and modelling benchmarks. *J. Climate*, **20**, 449 - 469, doi: 10.1175/JCLI3996.1.

Compo, G.P., *et al.*, 2011: The twentieth century reanalysis project. *Q. J. Roy. Meteorol. Soc.*, **137**, 1 - 28, doi: 10.1002/qj.776.

Coy, L., *et al.*, 2016: Structure and dynamics of the quasi-biennial oscillation in MERRA-2. *J. Climate*, **29**, 5339 - 5354, doi: 10.1175/JCLI-D-15-0809.1.

Cucurull, L., J.C. Derber, R. Treadon, and R. Purser, 2007: Assimilation of global positioning system radio occultation observations into NCEP's global data assimilation system. *Mon. Weather Rev.*, **35**, 3174 - 3193.

Das, S.S., *et al.*, 2016: Four-decadal climatological intercomparison of rocketsonde and radiosonde with different reanalysis data: results from Thumba Equatorial Station. *Q.J.R. Meteorol. Soc.*, **142**, 91 - 101, doi: 10.1002/qj.2632.

Davis, S.M., *et al.*, 2017: Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP. *Atmos. Chem. Phys.*, **17**, 12743 - 12778, doi: 10.5194/acp-17-12743-2017.

Dee, D.P., *et al.*, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. Roy. Meteorol. Soc.*, **137**, 553 - 597, doi: 10.1002/qj.828.

Fujiwara, M., *et al.*, 2015: Global temperature response to the major volcanic eruptions in multiple reanalysis data sets. *Atmos. Chem. Phys.*, **15**, 13507 - 13518, doi: 10.5194/acp-15-13507-2015.

Fujiwara, M., *et al.*, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp- 17-1417-2017.

Gelaro, R., *et al.*, 2017: The Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate*, **30**, 5419 - 5454, doi: 10.1175/JCLI-D-16-0758.1.

Gille, J., *et al.*, 2008: High resolution dynamics limb sounder: experiment overview, recovery, and validation of initial temperature data. *J. Geophys. Res.*, **113**, D16S43, doi: 10.1029/2007JD008824.

Gibson, J.K., *et al.*, 1997: ECMWF Re-analysis. *Final Project Report*, **Series 1**, ECMWF, Reading, 72 pp., available at: **https://www.ecmwf.int/node/9584** (last access: 05 March 2021).

GMAO – Global Modeling and Assimilation Office, 2015: MERRA-2 inst3_3d_asm_Nv: 3d, 3-Hourly, Instantaneous, Model-Level, Assimilation, Assimilated Meteorological Fields V5.12.4, Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA, doi: 10.5067/WWQSXQ8IVFW8.

Gray, L. and J. Pyle, 1987: Two-dimensional model studies of equatorial dynamics and tracer distributions. *Q. J. Roy. Meteorol. Soc.*, **113**, 635–651, doi: 10.1002/qj.49711347611.

Han, Y., *et al.*, 2006: JCSDA Community Radiative Transfer Model (CRTM) – Version 1, *NOAA Tech. Rep.* **122**, NOAA/NESDIS, USA, 33 pp, available at: **https://repository.library.noaa.gov/view/noaa/1157** (last access: 05 March 2021).

Hertzog, A., C. Basdevant, and F. Vial, 2006: An Assessment of ECMWF and NCEP–NCAR Reanalyses in the Southern Hemisphere at the End of the Presatellite Era: Results from the EOLE Experiment (1971–72), *Mon. Weather Rev.*, **134**, 11, 3367 - 3383, doi: 10.1175/MWR3256.1.

Kalnay, E., *et al.*, 1996: The NCEP/NCAR 40 year reanalysis project, *Bull. Am. Met. Soc.*, **77**, 437 - 471, doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kanamitsu, M., *et al.*, 2002: NCEP–DOE AMIP-II reanalysis (R-2), *Bull. Am. Met. Soc.*, **83**, 1631 - 1643, doi: 10.1175/BAMS-83-11-1631.

Kawatani, Y., *et al.*, 2016: Representation of the tropical stratospheric zonal wind in global atmospheric reanalyses, *Atmos. Chem. Phys.*, **16**, 6681 - 6699, doi: 10.5194/acp-16-6681-2016.

Kistler, R., *et al.*, 2001: The NCEP–NCAR 50 year reanalysis: monthly means CD-ROM and documentation, *Bull. Am. Met. Soc.*, **82**, 247 - 267, doi: 10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2.

Kobayashi, S., *et al.*, 2015: The JRA-55 reanalysis: general specifications and basic characteristics, *J. Meteorol. Soc. Jpn.*, **93**, 5 - 48, doi: 10.2151/jmsj.2015-001.

Lawrence, Z.D., *et al.*, 2015: Comparisons of polar processing diagnostics from 34 years of the ERA-Interim and MERRA reanalyses, *Atmos. Chem. Phys.*, **15**, 3873 - 3892, doi: 10.5194/acp-15-3873-2015.

Leroy, S., C. Ao, and O.P. Verkhoglyadova, 2012: Mapping GPS Radio Occultation Data by Bayesian Interpolation. *J. Atmos. Ocean. Tech.*, **29**, 1062, doi: 10.1175/JTECH-D-11-00179.1.

Livesey, N., *et al.*, 2015: EOS MLS version 4.2x level 2 data quality and description document, *Tech. Rep.* **D-33509** Rev. A, Jet Propul. Lab., Calif. Inst. of Technol., Pasadena, Calif., 2015.

Long, C.S., *et al.*, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP), *Atmos. Chem. Phys.*, **17**, 14593 - 14629, doi: 10.5194/acp-17-14593-2017.

Manney, G.L., *et al.*, 2005: Dianostic comparison of meteorological analyses during the 2002 Antarctic winter. *Mon. Weather Rev.*, **133**, 1261 - 1278, doi: 10.1175/MWR2926.1.

Martineau, P. and S.-W. Son, 2010: Quality of reanalysis data during stratospheric vortex weakening and intensification events. *Geophys. Res. Lett.*, **37**, L22801, doi: 10.1029/2010GL045237.

Martineau, P., S.-W. Son, and M. Taguchi, 2016: Dynamical consistency of reanalysis datasets in the extratropical stratosphere. *J. Climate*, **29**, 3057 - 3074, doi: 10.1175/JCLI-D-15-0469.1.

Mitchell, D.M., *et al.*, 2015: Signatures of naturally induced variability in the atmosphere using multiple reanalysis datasets, *Q. J. Roy. Meteorol. Soc.*, **141**, 2011 - 2031, doi: 10.1002/qj.2492.

Miyazaki, K., *et al.*, 2016: Inter-comparison of stratospheric mean-meridional circulation and eddy mixing among six reanalysis datasets. *Atmos. Chem. Phys.*, **16**, 6131 - 6152, doi: 10.5194/acp-16-6131-2016.

Molod, A., *et al.*, 2015: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8**, 1339 - 1356, doi: 10.5194/gmd-8-1339-2015.

Onogi, K., *et al.*, 2007: The JRA-25 reanalysis. *J. Meteorol. Soc. Jpn.*, **85**, 369 - 432, doi: 10.2151/jmsj.85.369.

Pawson, S. and M. Fiorino, M, 1998a: A comparison of reanalyses in the tropical stratosphere: Part 1: Thermal structure and the annual cycle. *Clim. Dynam.*, **14**, 631 - 644.

Pawson, S. and M. Fiorino, 1998b: A comparison of reanalyses in the tropical stratosphere: Part 2: The quasi-biennial oscillation. *Clim. Dynam.*, **14**, 645 - 658.

Pawson, S. and M. Fiorino, 1999: A comparison of reanalyses in the tropical stratosphere: Part 3: Inclusion of the pre-satellite data era. *Clim. Dynam.*, **15**, 241 - 250.

Pawson, S. and B. Naujokat, 1999: The cold winters of the middle 1990s in the northern lower stratosphere. *J. Geophys. Res.*, **104**, 14209 - 14222, doi: 10.1029/1999JD900211.

Podglajen, A., *et al.*, 2014: Assessment of the accuracy of (re)analyses in the equatorial lower stratosphere. *J. Geophys. Res.*

Atmos., **119**, 11,166–11,188, doi:10.1002/2014JD021849.

Poli, P., S.B. Healy, and D.P. Dee, 2010: Assimilation of global positioning system radio occultation data in the ECMWF ERA–interim reanalysis. *Q. J. Roy. Meteorol. Soc.*, **136**, 1972 - 1990, doi: 10.1002/qj.722.

Randel, W., *et al.*, 2004: The SPARC intercomparison of middle-atmosphere climatologies. *J. Climate*, **17**, 986 - 1003.

Randel, W., *et al.*, 2016: Stratospheric temperature trends over 1979 - 2015 derived from combined SSU, MLS, and SABER satellite observations. *J. Climate*, **29**, 4843 - 4858, doi: 10.1175/JCLI-D-15-0629.1.

Ray, E.A., M.J. Alexander, and J.R. Holton, 1998: An analysis of the structure and forcing of the equatorial semiannual oscillation in zonal wind. *J. Geophys. Res.*, **103**, 1759 - 1774, doi: 10.1029/97JD02679.

Rienecker, M.M., *et al.*, 2011: MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Climate*, **24**, 3624 - 3648, doi: 10.1175/JCLI-D-11-00015.1.

Saha, S., *et al.*, 2010: The NCEP climate forecast system reanalysis. *Bull. Am. Met. Soc.*, **91**, 1015 - 1057, doi: 10.1175/2010BAMS3001.1.

Saha, S., *et al.*, 2014: The NCEP climate forecast system version 2. *J. Climate*, **27**, 2185 - 2208, doi: 10.1175/JCLI-D-12-00823.1.

Schwartz, M.J., *et al.*, 2008: Validation of the Aura microwave limb sounder temperature and geopotential height measurements. *J. Geophys. Res.*, **113**, D15S11, doi: 10.1029/2007JD008783.

Seidel, D.J., *et al.*, 2016: Stratospheric temperature changes during the satellite era. *J. Geophys. Res. Atmos.*, **121**, 664 - 681, doi: 10.1002/2015JD024039.

Simmons, A.J., *et al.*, 2014: Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. *Q. J. Roy. Met. Soc.*, **140**, 329 - 353, doi: 10.1002/qj.2317.

Simmons, A., *et al.*, 2020: Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1, ECMWF Tech. Memo, **859**, 38pp, doi: 10.21957/rcxqfmg0, 2020.

Smith, A.K., R.R. Garcia, A.C. Moss, and N.J. Mitchell, 2017: The semiannual oscillation of the tropical zonal wind in the middle atmosphere derived from satellite geopotential height retrievals. *J. Atmos. Sci.*, **74**, 2413 - 2425, doi: 10.1175/JAS-D-17-0067.1.

Smith A.K, *et al.*, 2019: The equatorial stratospheric semiannual oscillation and time-mean winds in QBOi models. *Q. J. Roy. Met. Soc.*, 1 - 17. doi: 10.1002/qj.3690.

SPARC, 2002: SPARC intercomparison of middle atmosphere climatologies. edited by: Randel, W., M.-L. Chanin, and G. Michaut, *WCRP*-**116**, *WMO/TD*-No. **1142**, *SPARC Rep.* **3**, Verrières-le-Buisson CEDEX, France, 96 pp.

Steiner, A.K., *et al.*, 2020: Consistency and structural uncertainty of multi-mission GPS radio occultation records. *Atmos. Meas. Tech.*, **13**, 2547 - 2575, doi: 10.5194/amt-13-2547-2020.

Susskind, J., J. Rosenfield, and D. Reuter, 1983: An accurate radiative transfer model for use in the direct physical inversion of HIRS and MSU temperature sounding data. *J. Geophys. Res.*, **88**, 8550 - 8568, doi: 10.1029/JC088iC13p08550.

Thompson, D., *et al.*, 2012: The mystery of recent stratospheric temperature trends. *Nature*, **491**, 692 - 697, doi: 10.1038/nature11579.

Uppala, S.M., *et al.*, 2005: The ERA-40 reanalysis. *Q. J. Roy. Met. Soc.*, **131**, 2961 - 3012, doi: 10.1256/qj.04.176.

Young, P.J., *et al.*, 2011: The seasonal cycle and interannual variability in stratospheric temperatures and links to the Brewer–Dobson circulation: an analysis of MSU and SSU Data. *J. Climate*, **24**, 6243 - 6258, doi: 10.1175/JCLI-D-10-05028.1.

Young, P.J., *et al.*, 2012: Changes in stratospheric temperatures and their implications for changes in the Brewer–Dobson circulation, 1979–2005. *J. Climate*, **25**, 1759 - 1772, doi: 10.1175/2011JCLI4048.1.

Zhao, L., *et al.*, 2016: Use of SSU/MSU satellite observations to validate upper atmospheric temperature trends in CMIP5 simulations. *Remote Sens.*, **8**, 13, doi: 10.3390/rs8010013.

Zou, C.-Z. and H. Qian, 2016: Stratospheric temperature climate data record from merged SSU and AMSU-A observations. *J. Atmos. Ocean. Tech.*, **33**, 1967 - 1984, doi: 10.1175/JTECH-D-16-0018.1.

Zou, C.-Z. and W. Wang, 2012: MSU/AMSU radiance fundamental climate data record calibrated using simultaneous nadir overpasses. *Climate Algorithm Theoretical Basis Document (C-ATBD)*, NOAA/NESDIS, USA, available at: **https://www.star.nesdis.noaa.gov/smcd/emb/mscat/documents/MSU_AMSU_CATBD_V1.0.pdf** (last access: 8 March 2021).

Zou, C.-Z., et al., 2014: Recalibration and merging of SSU observations for stratospheric temperature trend studies. *J. Geophys. Res.-Atmos.*,**119**, 13180 - 13205, doi: 10.1002/2014JD021603.

Major abbreviations and terms

20CR	20th Century Reanalysis of NOAA and CIRES
AIRS	AIRS : Atmospheric InfraRed Sounder
AMSU	Advanced Microwave Sounding Unit (AMSU-A for Unit A)
ATOVS	Advanced TIROS Operational Vertical Sounder
Aura	A satellite in the EOS A-Train satellite constellation
CDAS-T574	Climate Data Assimillation System T574 resolution
CDR	Climate Data Record
CFSR	Climate Forecast System Reanalysis of NCEP
CIRA86	COSPAR International Reference Atmosphere, 1986
CIRES	Cooperative Institute for Research in Environmental Sciences (NOAA and the University of Colorado Boulder)
CFSR	Climate Forecast System Reanalysis of NCEP
CFSv2	CFSv2 : Climate Forecast System version 2. Post 2010 version of CFSR
CNES	Centre national d'études spatiales
COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate
COSPAR	Committee on Space Research
CRTM	Community Radiative Transfer Model
DOE	Department of Energy
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño Southern Oscillation
EOS	NASA Earth Observing System
ERA5	ECMWF Reanalysis version 5
ERA-15	ECMWF 15-year reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA-I or ERA-Interim	ECMWF interim reanalysis
GENESIS	Global Environmental and Earth Science Information System
GHOST	Global Horizontal Sounding Technique
GLATOVS	Goddard Laboratory for Atmospheres TOVS forward model
GMAO	Goddard Modeling and Assimilation Office
GPSRO	Global Positioning System radio occultation
GSI	Gridpoint Statistical Interpolation
HIRDLS	High Resolution Dynamics Limb Sounder
IASI	Infrared Atmospheric Sounding Interferometer
ITCZ	InterTropical Convergence Zone
JMA	Japanese Meteorological Agency
JRA-25	Japanese 25-year reanalysis
JRA-55	Japanese 55-year reanalysis
JPL	Jet Propulsion Laboratory
LDB	Long Duration Balloon
MERRA	Modern Era Retrospective-Analysis for Research (MERRA-2 for its version2)
MLS	Microwave Limb Sounder

MSU	Microwave Sounding Unit
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of NOAA
NDACC	Network for the Detection of Atmospheric Composition Change
NESDIS	National Environmental Satellite, Data, and Information Service of NOAA
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
NOAA-*	NOAA polar-orbiting operational meteorological satellite (* indicates the satellite number)
QBO	Quasi-biennial oscillation
R1	NCEP-NCAR Reanalysis 1
R2	NCEP-DOE Reanalysis 2
REM	Reanalysis ensemble mean
RMS	Root mean square
RTM	Radiative Transfer Model
S-RIP	SPARC Reanalysis Intercomparison Project
SAM	Southern Annular Mode
SAO	Semi-annual oscillation
SD	Standard Deviation
SH	Southern Hemisphere
SPARC	Stratosphere–troposphere Processes and their Role in Climate
SSU	Stratospheric Sounding Unit (SSU1 and SSU2 for SSU channel 1 and 2, respectively)
SSW	Sudden stratospheric warming
STAR	Center for Satellite Applications and Research of NESDIS
TERLS	Thumba Equatorial Rock Launching Station
TIROS	Television Infrared Observation Satellite
TLS	Temperature of the lower stratosphere (MSU channel 4 and AMSU channel 9)
TOVS	TIROS Operational Vertical Sounder
TTL	Tropical Tropopause Layer
UTLS	Upper Troposphere and Lower Stratosphere
WMO	World Meteorological Organization

Chapter 4: Overview of Ozone and Water Vapour

Chapter lead authors

Sean M. Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Michaela I. Hegglin	Department of Meteorology, University of Reading	UK

Co-authors

Rossana Dragani	European Centre for Medium-Range Weather Forecasts	UK
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Yayoi Harada	Japan Meteorological Agency	Japan
Chiaki Kobayashi	Japan Meteorological Agency	Japan
Craig Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Eric R. Nash	Science Systems and Applications, Inc	USA
Gerald L. Potter	National Aeronautics and Space Administration, Goddard Space Flight Center	USA
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Tao Wang	National Aeronautics and Space Administration, Jet Propulsion Laboratory	USA
Krzysztof Wargan	(1) National Aeronautics and Space Administration Global Modeling and Assimilation Office (2) Science Systems and Applications, Inc	USA
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China

Abstract. Because of the central role of water vapour (WV) and ozone (O_3) in determining local temperatures and winds in NWP systems, and for climate change more generally, it is important to understand how accurately and consistently these species are represented in existing global reanalyses. This chapter presents the results of WV and O_3 intercomparisons over a range of timescales and different regions of the stratosphere, and evaluates both inter-reanalysis and observation-reanalysis differences. Also provided is a systematic documentation of the treatment of WV and O3 in current reanalyses to aid future research and guide the interpretation of differences amongst reanalysis fields.

The assimilation of total column ozone (TCO) observations in newer reanalyses results in realistic representations of TCO in reanalyses except when data coverage is lacking, such as during polar night. The vertical distribution of ozone is also relatively well represented in the stratosphere in reanalyses, particularly given the relatively weak constraints on ozone vertical structure provided by most assimilated observations and the simplistic representations of ozone photochemical processes in most of the reanalysis forecast models. However, significant biases in the vertical distribution of ozone are found in the upper troposphere and lower stratosphere in all reanalyses.

In contrast to ozone, reanalysis estimates of stratospheric WV are not directly constrained by assimilated data. Observations of atmospheric humidity are typically used only in the troposphere, below a specified vertical level at or near the tropopause. The fidelity of reanalysis stratospheric WV products is therefore mainly dependent on each reanalysis' representation of the physical drivers that influence stratospheric WV, such as temperatures in the tropical tropopause layer, methane oxidation, and the stratospheric overturning circulation. The lack of assimilated observations and known deficiencies in the representation of stratospheric transport in reanalyses result in much poorer agreement amongst observational and reanalysis estimates of stratospheric WV. Hence, stratospheric WV products from the current generation of reanalyses should generally not be used for scientific data analysis.

Davis *et al.* (2017) have published a shortened version of this chapter.

Contents

4.1 Introduction 125
4.2 Description of ozone and water vapour in reanalyses 125
4.2.1 NCEP-NCAR (R1) and NCEP-DOE (R2) 127
4.2.2 CFSR 128
4.2.3 ERA-40 129
4.2.4 ERA-Interim 130
4.2.5 ERA5 131
4.2.6 JRA-25 and JRA-55 131
4.2.7 MERRA 131
4.2.8 MERRA-2 132
4.3 Data 133
4.3.1 Reanalysis data processing 133
4.3.2 SBUV and TOMS/OMI total column ozone 133
4.3.3 SPARC Data Initiative limb satellite observations 133
4.3.4 Aura MLS satellite data 134
4.3.5 SWOOSH merged limb satellite data record 134
4.4 Evaluation of reanalysis ozone products 134
4.4.1 Total column ozone seasonal cycle 134
4.4.2 Zonal mean ozone cross-sections 136
4.4.3 Ozone monthly mean vertical profiles and seasonal cycles 137
4.4.4 Ozone interannual variability 138
4.4.5 Ozone time series in equivalent latitude coordinates 138
4.4.6 Ozone quasi-biennial oscillation 141
4.4.7 Ozone hole area 144
4.4.8 Long-term evolution of ozone 145
4.5 Evaluation of reanalysis water vapour products 150
4.5.1 Zonal mean water vapour cross-sections 150
4.5.2 Water vapour monthly mean vertical profiles and seasonal cycles 151
4.5.3 Water vapour interannual variability 151
4.5.4 Tropical tape recorder in water vapour 153
4.6 Summary 155
4.7 Key findings and recommendations 157
References 159
Major abbreviations and terms 162

4.1 Introduction

Atmospheric reanalyses produce an optimal estimate of the past state of the atmosphere through the use of a forecast model, input observations, and an assimilation scheme. Meteorological observations assimilated and meteorological quantities analysed include variables such as temperature, wind, geopotential height, and humidity fields. In this chapter, we focus on the reanalysis representation of water vapour and ozone in the upper troposphere to stratosphere.

Ozone and water vapour are trace gases of fundamental importance to the radiative budget of the stratosphere. Because of their impact on stratospheric temperatures, winds, and the circulation (*e.g.*, *Dee et al.*, 2011), ozone and water vapour are represented as prognostic variables in almost all current reanalysis systems. However, the degree of sophistication to which ozone and water vapour fields and their variability are represented depends on the reanalysis system, which observations it assimilates, which microphysical and chemical parameterizations it includes, and how those parameterizations affect the trace gas distributions. The accuracy and consistency of analysis and reanalysis ozone and water vapour fields in the upper troposphere and stratosphere has only been addressed for a limited subset of diagnostics and analysis/reanalysis systems by a few studies (*e.g.*, *Jiang et al.*, 2015; *Dessler and Davis*, 2010; *Thornton et al.*, 2009; *Geer et al.*, 2006).

Since atmospheric scientists are interested in using ozone and water vapour fields from reanalyses for studying climate variability and change, we conducted the first comprehensive assessment of how realistically and consistently reanalyses represent water vapour and ozone in the upper troposphere and stratosphere. In particular, the goals of this chapter are to (1) provide a comprehensive overview of how ozone and water vapour are treated in reanalyses, (2) evaluate the accuracy of ozone and water vapour in reanalyses against both assimilated and independent (non-assimilated) observations, and (3) provide guidance to the community regarding the proper usage and limitations of reanalysis ozone and water vapour fields in the upper troposphere and stratosphere.

4.2 Description of ozone and water vapour in reanalyses

In this section, we provide information on how ozone and water vapour are represented in reanalyses. The information compiled here expands on that provided by *Fujiwara et al.* (2017) and *Chapter 2*, which contain a comprehensive overview of the reanalysis systems and their assimilated observations, including a basic discussion of the treatment of ozone (*Section 2.2.3.2*) and water vapour (*Section 2.4.3*).

In most reanalyses, ozone and water vapour are prognostic variables that are affected by the assimilated observations (see **Tables 4.1** and **4.2** for an overview of key aspects of these fields). The assimilated observations affecting the water vapour fields in reanalyses include some combination of radiosonde humidity profiles, GNSS-RO bending angles, and either radiances or retrievals from satellite microwave and infrared sounders such as TOVS, ATOVS, and SSM/I (see *Appendix A* for a list of all abbreviations; see also *Sections 2.2.3.2* and *2.4* in *Chapter 2* for an in-depth discussion of observations assimilated by the various reanalyses). These observational data affect the reanalysis water vapour fields in the lower atmosphere, but radiosonde humidity data are not assimilated above a specified level in the upper troposphere (typically between 300 hPa and 100 hPa, see **Table 4.2**).

Table 4.1: *Key characteristics of ozone treatment in reanalyses.*

Reanalysis	Primary TCO data sources	Vertical profile data sources	Stratospheric O_3 used in radiative transfer	Stratospheric O_3 treatment	Photochemical parameterization
NCEP R1	None	None	Climatology	None	None
NCEP R2	None	None	Climatology	None	None
CFSR	SBUV	SBUV	Analysed	Prognostic	CHEM2D-OPP
ERA-40	TOMS	SBUV	Climatology	Prognostic	CD86
ERA-I	TOMS, SCIAMACHY, OMI	SBUV, GOME, MLS, MIPAS	Same as ERA-40	Same as ERA-40	Same as ERA-40
ERA5	TOMS, OMI, SCIAMACHY	SBUV, MLS GOME, GOME-2, MIPAS	Updated Climatology	Same as ERA-40	Same as ERA-40
JRA-25	TOMS (1979–2004) [a] OMI (2004–)	Nudging to climatological profile	Daily values from offline CTM	Daily values from offline CTM	*Shibata et al.* (2005)
JRA-55	Same as JRA-25	None	Daily values from updated offline CTM	Daily values from updated offline CTM	*Shibata et al.* (2005)
MERRA	SBUV	SBUV	Analysed	Prognostic	*Stajner et al.* (2008)
MERRA-2	SBUV (1980–9/2004) OMI (9/2004–)	SBUV, MLS	Same as MERRA	Same as MERRA	Same as MERRA

[a] Offline CCM nudged to TOMS/OMI data.

Table 4.2: *Key characteristics of water vapour treatment in reanalyses.*

Reanalysis	Assimilation of satellite humidity radiances?	Highest level of assimilated WV observations	Highest level of analyzed WV[1]	Stratospheric WV used in radiative transfer	Stratospheric WV treatment	Stratospheric methane oxidation parameterization?
NCEP R1	No	300 hPa	300 hPa	Climatology	None	No
NCEP R2	No	300 hPa	10 hPa (RH only)	Climatology	None	No
CFSR	Yes	250 hPa	None	Analysed; negative values set to 0.1 ppmv	Prognostic	No
ERA-40	Yes	Diagnosed tropopause. Radiosonde humidity generally used to 300 hPa	Diagnosed tropopause	Analysed	Prognostic	Yes. Relaxation to 6 ppmv WV at stratopause
ERA-I	Yes	Same as ERA-40	Diagnosed tropopause	Analysed	Prognostic	Yes. Relaxation to 6.8 ppmv WV at stratopause
ERA5	Yes	Same as ERA-40	Diagnosed tropopause	Analysed	Prognostic	Same as ERA-I
JRA-25	Yes	100 hPa	50 hPa	Constant 2.5 ppmv	Prognostic[2]	No
JRA-55	Yes	100 hPa	5 hPa	Climatological annual mean from HALOE and UARS MLS during 1991–1997	Prognostic[2]	No
MERRA	Yes	300 hPa	None	Analysed	3-day relaxation to zonal-mean monthly-mean satellite-based climatology	No
MERRA-2	Yes	300 hPa	None	Same as MERRA	Same as MERRA	No

[1] Level above which assimilation-related increments are not allowed.
[2] Water vapour not provided above 100 hPa in pressure level analysis products.

Even though radiosonde humidity data may not be assimilated above a certain level, analysis increments are possible at higher levels unless the vertical correlations of the background errors are set to zero. Where relevant, this cut-off level above which analysis increments are disallowed has been noted in **Table 4.2**.

Because stratospheric water vapour data are not directly assimilated, the treatment of water vapour in the stratosphere is highly variable amongst the reanalyses. For the modern reanalyses, the concentration of water vapour entering the stratosphere is typically controlled by transport and dehydration processes occurring in the forecast model, primarily in the tropical tropopause layer (TTL). Higher in the stratosphere, chemical production of water vapour through methane oxidation is parameterised in some reanalyses, while others use a simple relaxation of the simulated water vapour field to an observed climatology.

As with water vapour, the treatment of ozone is quite different from reanalysis to reanalysis. The ozone treatment in reanalyses ranges from omitting prognostic ozone and using a climatology in the radiation calculations (NCEP R1/R2), to using a fully prognostic field with parameterised photochemistry (CFSR, ERA-40, ERA-I, ERA5, MERRA, MERRA-2), to assimilating ozone with an offline chemical transport model for use in the forecast model radiation calculation (JRA-25, JRA-55).

The primary ozone observations assimilated by reanalyses are satellite nadir UV backscatter-based retrievals of vertically integrated total column ozone (TCO) or broad vertically weighted averages (*e.g.*, SBUV data). These data come from a variety of satellites that have flown since the late 1970s, and reanalyses vary widely in what subset of the available data they assimilate (**Figures 4.1** and **4.2**). Some further differences exist amongst the reanalyses in their usage of different data versions from the same satellite instrument, and from different applications of data quality control and filtering. These differences in usage of input data may affect the reanalysis ozone fields.

Additional observation types using spectral ranges outside of the UV (namely microwave and IR) and exploiting different viewing geometries (such as limb-sounding) have been used, particularly by the newest reanalyses (ERA-I, MERRA-2).

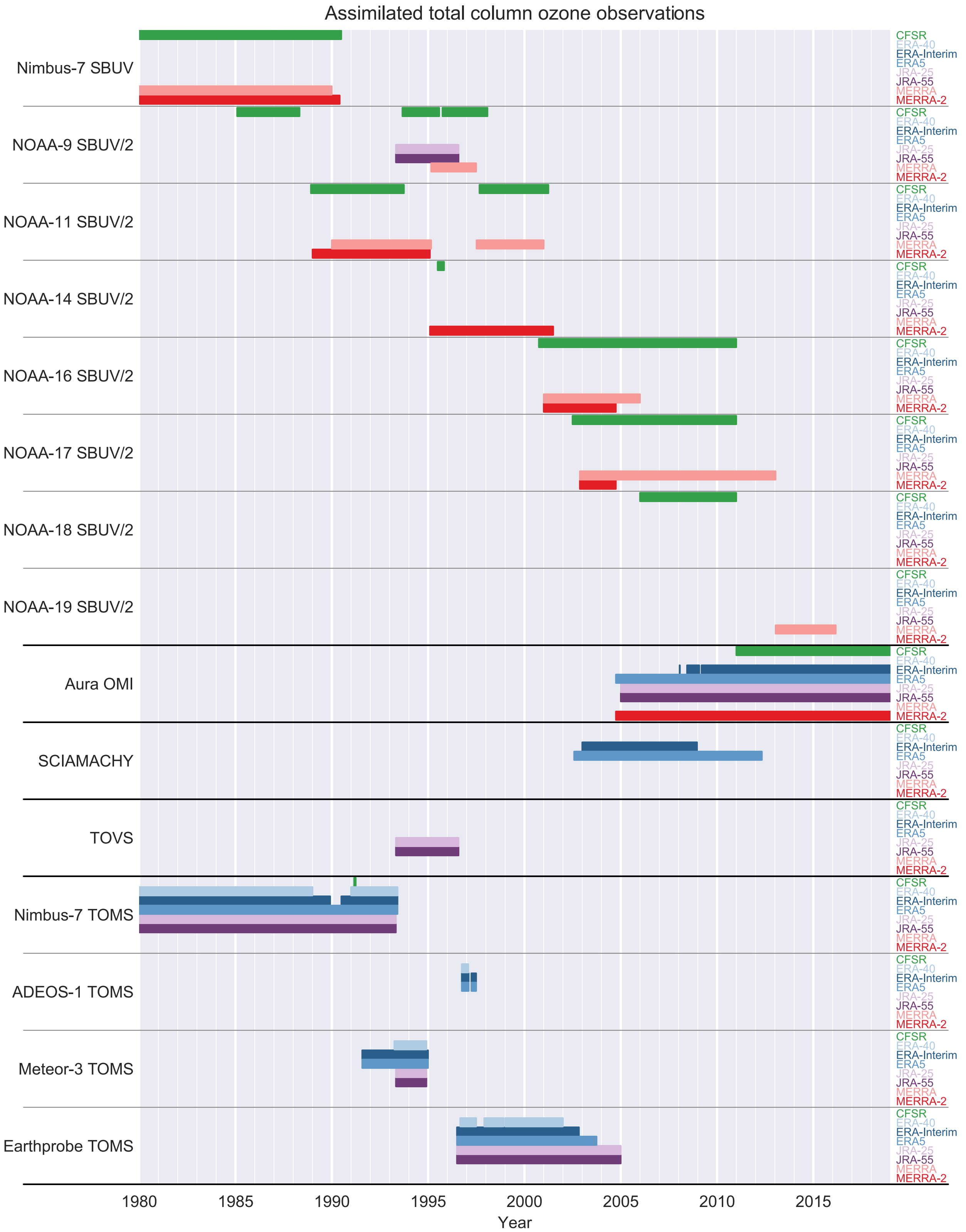


Figure 4.1: *Total column ozone data by instrument as assimilated by the different reanalyses. Updated from Davis et al. (2017).*

The assimilation of additional data, particularly higher vertical resolution limb sounding data, are expected to improve the quality of the ozone in reanalyses. However, the assimilation of new data sets could introduce sudden changes in the reanalysis ozone fields, and these transition times should be considered carefully when deriving or analysing long-term trends.

4.2.1 NCEP-NCAR (R1) and NCEP-DOE (R2)

Neither NCEP-NCAR (*R1*) nor NCEP-DOE (*R2*) assimilates ozone data (*Kanamitsu et al.*, 2002; *Kistler et al.*, 2001; *Kalnay et al.*, 1996). A climatology of ozone was used for radiation calculations.

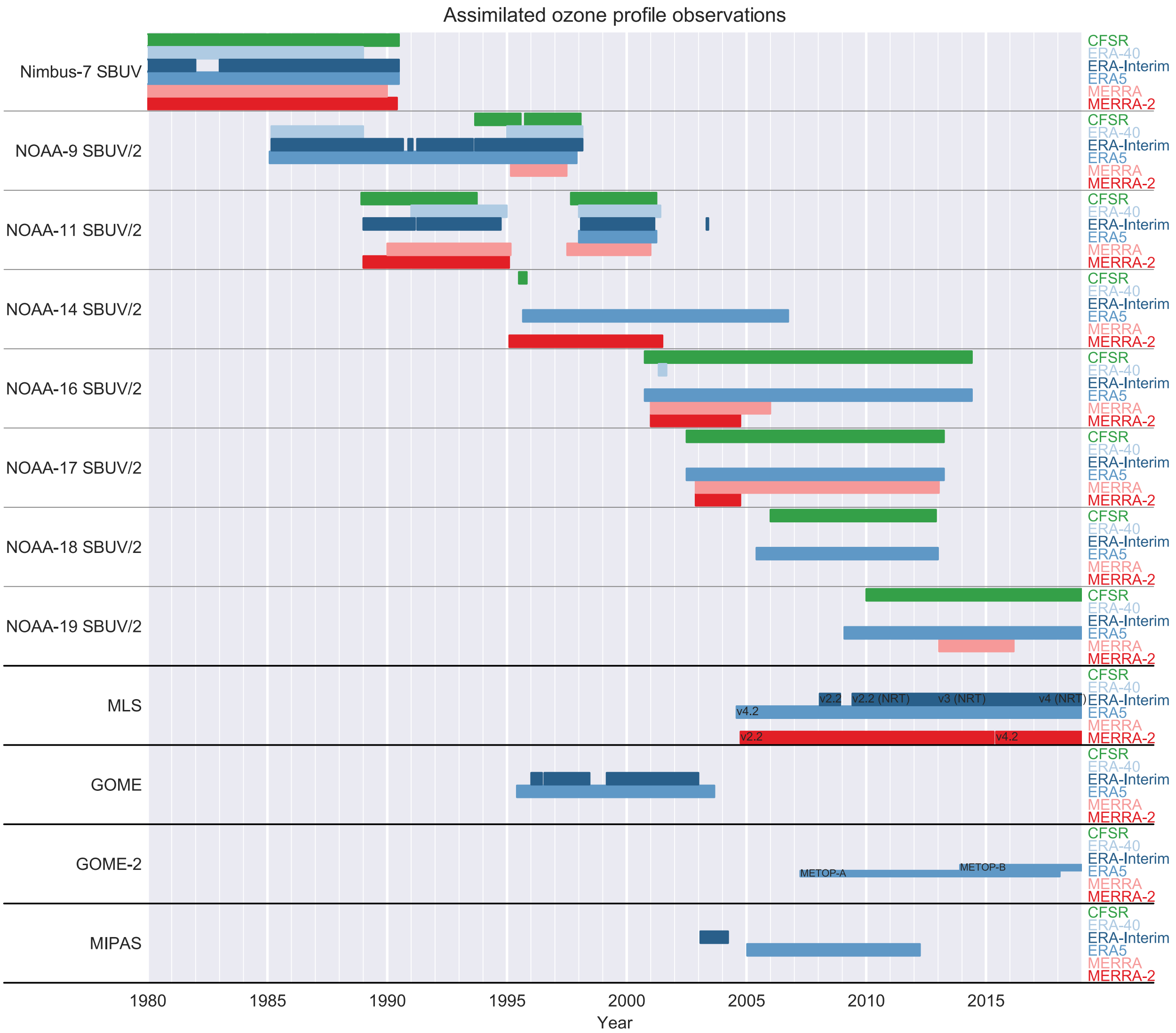


***Figure 4.2:** Ozone vertical profile observations by instrument as assimilated by the different reanalyses. These include higher vertical resolution limb sounders (MLS and MIPAS) and lower resolution nadir sounders (all others). Updated from Davis et al. (2017).*

Humidity information from radiosondes is assimilated in *R1* and *R2*, but humidity information from satellites is not (*Ebisuzaki and Zhang*, 2011). In general, the treatment of water vapour is similar in *R1* and *R2*, with a few notable differences. One major difference is that humidity is not output above 300 hPa in *R1*, whereas it is output up to 10 hPa in *R2*. Another difference is that only relative humidity is output in *R2*, whereas in *R1* both specific humidity and relative humidity are output. It is worth noting that in *R1*, specific humidity is a diagnostics variable, computed from relative humidity and temperature. Several fixes and changes were made in the treatment of clouds in *R2*, and these result in *R2* being ~20 % drier than *R1* in the tropics at 300 hPa (*Kanamitsu et al.*, 2002). As the focus here is on upper levels, we do not assess humidity fields from *R1* or *R2*. It is worth noting that *R1* shows negative long-term humidity trends between 500 hPa and 300 hPa (*Paltridge et al.*, 2009); however, these negative trends appear to reflect suspect radiosonde measurements at these levels and are not found in other reanalyses or satellite data (*Dessler and Davis*, 2010).

4.2.2 CFSR

The Climate Forecast System Reanalysis (CFSR) is a newer NCEP product following the NCEP R1 and R2 reanalyses but with numerous improvements (*Saha et al.*, 2010), including an updated forecast model and data assimilation system. CFSR was originally provided through the end of 2009, but output from the same analysis system was extended through the end of 2010 before transitioning to the CFSv2 analysis system starting in January 2011 (*Saha et al.*, 2014). Because CFSv2 was intended as a continuation of CFSR, in this chapter we refer to both CFSR (*i.e.*, CFSRv1) and CFSv2 as CFSR. However, the system changeover did result in a discontinuity in the water vapour fields that is addressed later in this chapter.

CFSR treats ozone as a prognostic variable that is analysed and transported by the forecast model. The CFSR forecast model uses analysed ozone data for radiation calculations.

In the forecast model, ozone chemistry is parameterised using production and loss terms generated by the NRL CHEM2D-OPP (*McCormack et al.*, 2006). These production and loss rates are provided as monthly mean zonal means, and are a function of local ozone concentration. The rates do not include the coefficients for temperature and overhead ozone column provided by *McCormack et al.* (2006), nor heterogeneous chemistry, although late 20th century levels of CFCs are used indirectly because CHEM2D-OPP is based on the CHEM2D middle atmospheric photochemical transport model, which includes ODS levels representative of the late twentieth century.

CFSR assimilates version-8 SBUV profile and TCO retrievals (*Flynn et al.*, 2009) from Nimbus-7 and SBUV/2 profiles and TCO retrievals from NOAA-9, -11, -14, -16, -17, -18, and eventually NOAA-19 (**Figures 4.1** and **4.2**). The ozone layer and TCO values assimilated by CFSR have not been adjusted to account for biases from one satellite to the next, although the use of SBUV version 8 is expected to minimize satellite-to-satellite differences. Despite the fact that CFSR assimilates TCO retrievals and SBUV ozone profiles, differences have been found between CFSR and SBUV(/2) ozone profile data (*Saha et al.*, 2010). Most of these differences are located above 10 hPa, and appear to result from observational background errors that were set too high in the CFSR upper stratosphere by between a factor of 2 (at 10 hPa) and a factor of 60 (at 0.2 hPa). Because of this, assimilated SBUV(/2) ozone layer observations do not alter the CFSR first guess for pressures less than 10 hPa, and the model first guess is used instead. The observational background errors were fixed for CFSv2, starting in 2011.

Water vapour is treated prognostically in CFSR. There are several assimilated observation types that influence the analysis humidity fields in the troposphere, including GNSS-RO bending angles, radiosondes, and satellite radiances. However, as radiosonde humidity data is only assimilated at 250 hPa and greater pressures, there are no specific observations that constrain humidity in the stratosphere. Stratospheric humidity in CFSR is hence primarily governed by physical processes and parameterizations in the model, including dehydration within the TTL. The treatment of water vapour in the model can lead to negative water vapour values around and above the tropopause. These negative values are replaced by small positive values of 0.1 parts per million by volume (ppmv) for the radiation calculations, but are retained in the analysis products. CFSR does not include a parameterization of methane oxidation.

4.2.3 ERA-40

The ERA-40 forecast model included prognostic ozone and a parameterization of photochemical sources and sinks of ozone, as described by *Dethof and Hólm* (2004). This parameterization of ozone production/loss rates is an updated version of the one proposed by *Cariolle and Deque* (1986, hereinafter *CD86*). In *CD86*, the net ozone production rate is parameterised as a function of the perturbation (relative to climatology) of the local ozone concentration, the local temperature, and the column ozone overhead. Compared to the *CD86* formulation, the ozone parameterization in ERA-40 includes an additional term representing heterogeneous chemistry. This loss term scales with the product of the local ozone concentration and the square of the equivalent chlorine concentration, and is only turned on at temperatures below 195 K. The climatologies and coefficients used in the parameterization are derived from a photochemical model and vary by latitude, pressure, and month. The prescribed chlorine loading varies from year to year, from ~ 700 parts per trillion (ppt) in 1950 to ~ 3400 ppt in the 1990s. Instead of the *CD86* ozone photochemical equilibrium values, ERA-40 made use of the *Fortuin and Langematz* (1995) ozone climatology.

The prognostic ozone was not used in the radiation calculations, which instead assumed the climatological ozone distribution reported by *Fortuin and Langematz* (1995). This choice was motivated by concerns that ozone–temperature feedbacks would degrade the temperature analysis if the assimilated ozone observations were of poorer quality than the temperature observations (*Dethof and Hólm*, 2004).

ERA-40 assimilated TOMS TCO and SBUV layer ozone retrievals from the end of 1978 onward (**Figures 4.1** and **4.2**; See also Table 1, *Dethof and Hólm*, 2004; *Poli*, 2010). No ozonesonde measurements were assimilated, and no ozone data at all were assimilated before 1978. Ozone data prior to 1978 are thus primarily products of the photochemical parameterization. In addition, no ozone data were assimilated during 1989 and 1990 because the execution of the first ERA-40 stream (1989 - 2002, see discussion in *Chapter 2*) was started before the ozone assimilation scheme was implemented. Ozone background error covariances were also changed, such that the period January 1991 to October 1996 used an earlier and inferior background error covariance matrix than the rest of ERA-40 (see discussion in *Dethof and Hólm*, 2004). As a result, there are fewer problems with the ozone vertical profiles during the 1979 - 1988 and November 1996 - 2002 time periods.

ERA-40 water vapour products below the diagnosed tropopause are substantially affected by assimilated observations. Three main periods can be identified (*Uppala et al.*, 2005): until 1973, ERA-40 used only conventional in situ surface and radiosonde measurements; from 1973, satellite radiances from VTPR (1973 - 1978) and the TOVS instruments MSU, SSU, and HIRS (1978 - onwards) were used in addition to these conventional data sources; from 1987, 1D-Var retrievals of TCWV from SSM/I radiances were added to the assimilation. Radiosonde humidity measurements were generally used at pressures greater than 300 hPa. No adjustments to the humidity field due to data assimilation were made in ERA-40 above the diagnosed tropopause. Thus, stratospheric water vapour in ERA-40 reflects TTL dehydration, transport, and methane oxidation.

The latter was included via a simple stratospheric parameterization, in which WV was gradually relaxed to 6 ppmv at the stratopause (*Untch et al.*, 1998). This relaxation was later found to produce too low WV concentrations at the stratopause as it was based on earlier studies when atmospheric methane levels were lower (*Uppala et al.*, 2005). ERA-40 stratospheric humidity has also been shown to be too low overall, due primarily to a cold bias in TTL temperatures caused by an excessively strong Brewer-Dobson circulation (*Oikonomou and O'Neill*, 2006).

4.2.4 ERA-Interim

The treatment of ozone and water vapour in ERA-Interim is very similar to that in ERA-40. Notable differences include additional assimilated datasets and an improved treatment of water vapour in the upper troposphere and lower stratosphere (UTLS). Descriptions of the ozone system and assessments of its quality have been provided by *Dee et al.* (2011) and *Dragani* (2011).

As with ERA-40, total ozone from TOMS (Jan 1979 - Nov 1989; Jun 1990 - Dec 1994; Jun 1996 - Dec 2001) and ozone layer averages from SBUV (1979 - present) are assimilated (**Figures 4.1** and **4.2**). ERA-Interim also assimilates TCO from OMI (Jun 2008 - Jan 2009, Mar 2009 - present) and SCIAMACHY (Jan 2003 - Dec 2008), and ozone profiles from GOME (Jan 1996 - Dec 2002), MIPAS (Jan 2003 - Mar 2004), and MLS (Jan - Nov 2008, Jun 2009 - present). Details on the data versions and data providers are provided in Table 1 of *Dragani* (2011). A change in the assimilation of SBUV ozone profiles was implemented in January 2008. Before January 2008, assimilated SBUV profiles were low vertical resolution products derived over six vertical layers (0.1 - 1 hPa, 1 - 2 hPa, 2 - 4 hPa, 4 - 8 hPa, 8 - 16 hPa and 16 hPa - surface) from NOAA version 6 (v6) retrievals. These data were replaced by native 21-vertical-level SBUV profiles from v8 retrievals.

The assimilation of ozone profile retrievals from Aura MLS started in 2008 (**Figure 4.2**) using the reprocessed v2.2 MLS retrievals (215 - 0.1 hPa), followed by the near-real-time v2.2 product (68 - 0.1 hPa) from June 2009 through December 2012, followed by a "v3+" near-real-time product (same levels as the reprocessed v2.2 with an additional level at 178 hPa) from January 2013 to 17 March 2017. After 17 March 2017 until present the MLS v4 near-real-time product has been assimilated.

The ozone forecast model used in ERA-Interim has the same basic formulation as that used in ERA-40 but some aspects of the parameterization have been upgraded substantially, especially the regression coefficients. An account of the changes is provided by *Cariolle and Teyssédre* (2007). As in ERA-40, the radiation scheme in ERA-Interim does not use the prognostic ozone field.

A preliminary assessment of the temperature and wind fields revealed unrealistic temperature and horizontal wind increments generated near the stratopause by the 4D-Var assimilation scheme in an attempt to accommodate large local adjustments in ozone concentrations (*Dragani*, 2011; *Dee*, 2008). As an ozone bias correction was not available in ERA-Interim to limit the detrimental effect of ozone assimilation on temperature and wind fields, the sensitivity of the latter to ozone changes was switched off in ERA-Interim. This change affected the period from 1 February 1996 onwards and the ten years from 1979 through 1988 that were run at a later stage.

Through December 1995, ERA-Interim ozone analyses perform better than their ERA-40 counterparts with respect to independent ozone observations in the upper troposphere and lower stratosphere, but perform slightly worse on average in the middle stratosphere (*Dee et al.*, 2011). The assimilation of GOME ozone profiles (Jan 1996 - Dec 2002) improves the agreement between ERA-Interim analyses and independent data, such that ERA-Interim outperforms ERA-40 throughout the atmosphere (including the middle stratosphere) from January 1996 through the end of ERA-40 in September 2002 (*Dragani*, 2011).

The ERA-Interim humidity analysis is substantially modified from that in ERA-40 due to changes in both model physics and assimilated observations. A non-linear transformation of the humidity control variable was introduced to make humidity background errors more Gaussian (*Uppala et al.*, 2005; *Hólm*, 2003; *Hólm et al.*, 2002). This transformation normalizes relative humidity increments by a factor that depends on background estimates of relative humidity and vertical level. A 1D-Var assimilation of rain-affected radiances over oceans was also added as part of the 4D-Var outer loop (*Dee et al.*, 2011), which helps to constrain the spatial distribution of total column water vapour (TCWV). The ERA-Interim humidity analysis also benefits from several changes in the model physics, including changes in the convection scheme that lead to increased convective precipitation (particularly at night), reduced tropical wind errors, and a better representation of the diurnal phasing of precipitation events (*Bechtold et al.*, 2004). The non-convective cloud scheme was also updated.

Perhaps of most relevance for humidity in the UTLS, the revised cloud scheme contains a new parameterization that allows supersaturation with respect to ice in the cloud-free portions of grid cells with temperatures less than 250 K (*Tompkins et al.*, 2007). The inclusion of this parameterization results in substantial increases in relative humidity in the upper troposphere and in the stratospheric polar cap relative to ERA-40 (*Dee et al.*, 2011). Methane oxidation in the stratosphere is included via a parameterization like the one used in ERA-40 but with relaxation to 6.8 ppmv at the stratopause (rather than 6 ppmv as in ERA-40), based on an analysis of UARS data by *Randel et al.* (1998).

As with ERA-40, no adjustments due to data assimilation are applied in the stratosphere (above the diagnosed tropopause). ERA-interim tropospheric humidity is affected by the assimilation of radiosonde humidity measurements, radiances from the TOVS (through 5 Sep 2006) and ATOVS (from Aug 1998) instrument suites, and TCWV retrievals based on rain-affected radiances from SSM/I (from Aug 1987). Recent ERA-Interim humidity analyses may also be affected by the assimilation of GNSS-RO bending angles (from May 2001) and/or AIRS all-sky radiances (from April 2004).

4.2.5 ERA5

The treatment of ozone in ERA5 is the same as that used in ERA-Interim, but with substantial updates to the assimilated data. Reprocessed retrievals are assimilated from TOMS (1979 - 2003), SBUV v8.6 (1979 - present), CCI MIPAS (2005 - 2012) and SCIAMACHY (2003 - 2012), Aura MLS v4.2 (2004 - present) and OMI-DOAS (2004 - present). ERA5 also assimilates IR ozone-sensitive radiances that were not used in ERA-Interim, and uses variational bias correction (see *Section 2.2.3.2*) during the ozone analysis. Analysed ozone is not used in the radiation calculations, which instead use an in-house ozone climatology from Copernicus Atmosphere Monitoring Service interim reanalysis (CAMSiRA, *Flemming et al.*, 2017).

Water vapour in ERA5 is similar to ERA-Interim. Notable differences are that the parameterization of supersaturation with respect to ice in cloud-free portions of grid cells has been extended to all temperatures less than 273 K (as opposed to only temperatures less than 250 K in ERA-Interim) and a more consistent treatment of potentially negative water vapour values in the stratosphere has been added.

4.2.6 JRA-25 and JRA-55

Ozone observations were not assimilated directly in the JRA-25 and JRA-55 systems (*Kobayashi et al.*, 2015; *Onogi et al.*, 2007). Instead, daily three-dimensional ozone fields were produced separately and provided to the JRA forecast model (*i.e.*, to the radiation scheme). Daily ozone fields in JRA-55 for 1978 and earlier are interpolated in time from a monthly mean climatology for 1980 - 1984. Daily ozone fields in both systems for 1979 and later are produced using an offline chemistry climate model (MRI-CCM1, *Shibata et al.*, 2005) that assimilated satellite observations of TCO using a nudging scheme. Assimilated TCO retrievals are taken from TOMS on Nimbus-7 and other satellites for the period 1979 - 2004 and from Aura OMI after the beginning of 2005. Different versions of MRI-CCM1 and different preparations of the ozone fields have been used for JRA-25 and JRA-55. For JRA-25, MRI-CCM1 output were also nudged to climatological ozone vertical profiles to account for a known bias in tropospheric ozone that produces a bias in stratospheric ozone after nudging to observations of total ozone. This procedure produced reasonable peak ozone-layer values in the final ozone product. This vertical-profile nudging was not necessary for JRA-55, which used an updated version of MRI-CCM1. JRA-55 produces improved peak values in vertical ozone profiles relative to JRA-25, as well as a clear ozone quasi-biennial oscillation (QBO) signature.

As with other modern reanalyses, JRA-25 and JRA-55 humidity fields are affected by the assimilation of radiosonde humidity measurements and satellite radiances. The JRA-25 assimilation analysed the logarithm of specific humidity (*Onogi et al.*, 2007). Stratospheric humidity was dry-biased and generally decreased with time in JRA-25, in part due to the lack of parameterised methane oxidation. The JRA-25 forecast model radiation calculations assumed a constant value of 2.5 ppmv in the stratosphere. Water vapour in the UTLS shows evidence of discontinuities at the start of 1991, which corresponds to the transition between the two major processing streams of JRA-25. *Onogi et al.* (2007) reported sudden jumps of +0.7 ppmv at 150 hPa and +0.9 ppmv at 100 hPa associated with this transition.

The treatment of water vapour in JRA-55 is similar in most respects to that in JRA-25. JRA-55 does not contain a parameterization of methane oxidation. Differences include a change in the upper boundary above which the vertical correlations of humidity background errors are set to zero, preventing spurious analysis increments at higher levels. This boundary is set at 5 hPa in JRA-55, and 50 hPa in JRA-25. Forecast model radiation calculations in JRA-55 use an annual mean climatology of stratospheric water vapour derived from UARS HALOE and UARS MLS measurements made during 1991 - 1997 in the stratosphere, rather than the constant 2.5 ppmv used in JRA-25. The introduction of an improved radiation scheme in JRA-55 greatly reduced lower stratospheric negative temperature biases that were present in JRA-25 during the TOVS period before 1998 (*Fujiwara et al.*, 2017; *Kobayashi et al.*, 2015), which may have beneficial impacts on JRA-55 stratospheric humidity products. Water vapour concentrations at pressures less than 100 hPa are not provided in the standard pressure-level products of these two reanalyses (although these concentrations are provided in model-level products), and are therefore not evaluated in this chapter.

4.2.7 MERRA

Ozone is a prognostic variable in MERRA (*Rienecker et al.*, 2011), and is subjected to assimilation, transport by assimilated winds (more precisely, the odd-oxygen family is the transported species), and parameterised chemistry. The MERRA general circulation model (GCM) uses a simple chemistry scheme that applies monthly zonal mean ozone production and loss rates derived from a 2-dimensional chemistry model (*Stajner et al.*, 2008).

Ozone data assimilated in the reanalysis include partial columns and total ozone (defined as the sum of layer values in a profile) from a series of SBUV instruments (*Flynn et al.*, 2009) on various NOAA platforms (**Figures 4.1** and **4.2**). Version 8 of the SBUV retrievals (Flynn, 2007) is used but the native 21 vertical layers are combined into 12 layers (each 5 km deep) prior to assimilation. All other assimilated data, including radiance observations, are explicitly prevented from impacting the ozone analysis directly.

Background error standard deviations for ozone are specified as ~4% of the global mean ozone on a given model level. Horizontal background error correlation lengths vary from ~400 km in the troposphere to ~800 km at the model top. Assimilated ozone fields are fed into the forecast model radiation scheme and are used in the radiative transfer model for radiance assimilation.

Water vapour is also a prognostic assimilated variable in MERRA; however, unlike ozone, moisture fields in the stratosphere are relaxed to a 2-D monthly climatology with a relaxation time of 3 days. This climatology is derived from water vapour observations made by the UARS HALOE and Aura MLS instruments (*e.g.*, *Rienecker et al.*, 2011 and references therein). This climatological constraint is introduced gradually over the layer between the model tropopause and 50 hPa, where pressure-dependent blending between the climatology and the GCM water vapour is applied. Water vapour above the tropopause does not undergo physically meaningful variations on timescales longer than the 3-day relaxation timescale except in the lowermost stratosphere where the climatology is given a smaller weight. No attempt was made to account for methane oxidation or trends in stratospheric methane concentrations.

MERRA assimilates specific humidity measurements from radiosondes at pressures above 300 hPa and marine surface observations. Moisture fields are affected by microwave radiance data from SSM/I and AMSU-B/MHS, infrared radiances from HIRS, the GOES Sounder, and AIRS, and rain rates derived from TMI and SSM/I. Background error statistics for water vapour were derived using the National Meteorological Center method and applied using a recursive filters methodology (*Wu et al.*, 2002). The moisture control variable is pseudo-relative humidity (*Dee and Da Silva*, 2003).

4.2.8 MERRA-2

The key differences between the treatment of ozone in MERRA-2 (*Gelaro et al.*, 2017) and that in MERRA are in the observing system and background error covariances. From January 1980 to September 2004, MERRA-2 assimilates v8.6 SBUV retrievals of partial columns on a 21-layer vertical grid (*Bhartia et al.*, 2013) and total ozone computed as the sum of individual layer values. Compared to the v8 retrievals used in MERRA, the v8.6 algorithm uses improved ozone cross-sections and an improved cloud height climatology. These updates result in better agreement with independent ozone data and make SBUV more suitable for long-term climatologies (*Frith et al.*, 2014; *McPeters et al.*, 2013). Starting in October 2004, SBUV data was replaced by a combination of TCO from Aura OMI (*Levelt et al.*, 2006) and stratospheric profiles from Aura MLS (*Waters et al.*, 2006). The OMI data consist of TCO retrievals from collection 3 and are based on the v8.5 retrieval algorithm, which is an improvement of the v8.0 algorithm extensively evaluated by *McPeters et al.* (2008). The assimilation algorithm makes use of the OMI averaging kernels to account for the sensitivity of these measurements to clouds in the lower troposphere (*Wargan et al.*, 2015). MLS data are from v2.2 between October 2004 and May 2015 and v4.2 (*Livesey et al.*, 2017) afterwards. Users of the MERRA-2 ozone product should therefore be aware that the reanalysis record may show a discontinuity in 2004 with two distinct periods as follows: the SBUV period (1980 - September 2004) and the EOS Aura period (from October 2004 onward). The analysis is expected to be of higher quality during the latter period due to the higher vertical resolution of Aura MLS profiles relative to SBUV profiles and the availability of MLS observations during night.

Ozone background error variance in the MERRA-2 model follows *Wargan et al.* (2015). The background error standard deviation at each grid point is proportional to the background ozone at that point and time. This approach introduces a flow dependence into the assumed background errors and allows a more accurate representation of shallow structures in the ozone fields, especially in the UTLS. As in MERRA, the ozone analyses are radiatively active tracers in both the forecast model and the radiative transfer model used for assimilation of satellite radiances. *Bosilovich et al.* (2015) provided a preliminary evaluation of the MERRA-2 ozone product. A more comprehensive description and validation, including comparisons with MERRA, is given in *Wargan et al.* (2017).

The treatment of stratospheric water vapour in MERRA-2 is similar to that in MERRA, with a 3-day relaxation to the same climatological annual cycle. The main innovation is the introduction of additional global constraints that ensure the conservation of the dry mass of the atmosphere and rescale the water vapour tendency to remove the globally integrated mean from the analysis increment (*Takacs et al.*, 2016).

In addition to the moisture data assimilated in MERRA, MERRA-2 assimilates GNSS-RO data and radiances from the recently introduced infrared sensors IASI, CrIS, and SEVIRI. Radiances from these recent IR instruments are not highly sensitive to stratospheric water vapour, but stratospheric water vapour is not explicitly prevented from being affected by the assimilation of these observations. Changes in the MERRA-2 observing system relative to MERRA are described in more detail by *Bosilovich et al.* (2015) and *McCarty et al.* (2016).

The moisture control variable in the MERRA-2 assimilation scheme is pseudo-relative humidity normalised by the background error standard deviation. Background error covariances used in MERRA-2 have been significantly retuned relative to those used in MERRA (*Bosilovich et al.*, 2015).

4.3 Data

In this section, we describe the approach we use to process the reanalysis ozone and water vapour fields, and the observations used to evaluate them. We note that some of these observational data are assimilated by the reanalyses. While comparisons between reanalyses and observations would ideally be based on independent observations, this is not always possible given the paucity of water vapour and ozone data in parts of the atmosphere. However, comparison to assimilated observations can serve a useful purpose by providing an internal consistency check on the ability of reanalysis data assimilation systems to exploit the data they assimilate.

4.3.1 Reanalysis data processing

Most of the comparisons presented in this chapter are based on monthly mean reanalysis fields calculated from the "pressure level" data sets provided by each reanalysis centre, and processed into a standardised format as part of the CREATE project (**https://esgf.nccs.nasa.gov/projects/create-ip/**). The one exception to this is JRA-25 ozone data, which we have processed ourselves. This was done because the pressure level data product provided by JMA ("fcst_phy3m25") used incorrect hybrid model level coefficients when converting from model levels to pressure levels. The JRA-25 ozone data used here were computed directly from the 6-hourly model level data product ("fcst_phy3m"). To facilitate intercomparison amongst reanalyses, the pressure level-based datasets have been re-gridded to a common horizontal grid (2.5° lon x 2.5° lat) and a common set of 26 pressure levels (1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, 1, 0.7, 0.5, 0.3, 0.1 hPa). Unless otherwise noted, climatological comparisons follow the WMO convention in using the 30-year 1981 - 2010 climatological norm (*Arguez and Vose*, 2011).

Reanalysis TCO data are monthly means computed from the 6-hourly TCO fields. All of the models provided 6-hourly TCO on various native horizontal grids, except for JRA-25. For JRA-25, 6-hourly ozone mass mixing ratios were provided on model levels. The mixing ratios were integrated for each horizontal grid point to get TCO, and then monthly means were computed. For each reanalysis, the climatologies and departures from climatology were calculated and are presented on each data set's native horizontal grid. For comparisons to the SBUV and TOMS/OMI data, each model was interpolated to the native horizontal grid of each of the observational data sets.

4.3.2 SBUV and TOMS/OMI total column ozone

Two datasets are used to evaluate the total column ozone in the reanalyses. The first is the SBUV Merged Ozone Data Set (*Frith et al.*, 2014). The second is a combination of TOMS and Aura OMI OMTO3d total ozone observations (*Bhartia and Wellemeyer*, 2002). These two data sets provide a long, coherent span of observations for evaluation. TOMS and OMI data were processed using the TOMS V8 algorithm, while the SBUV data were processed using the TOMS V8.6 algorithm. Because data from SBUV and TOMS (and in many cases OMI) are assimilated by most of the reanalyses, these comparisons are not independent.

Since SBUV sensors measure backscatter solar ultraviolet radiation, only daytime observations are available; wintertime ozone in polar regions is thus poorly constrained by observations. Early NOAA satellites experienced orbital drifts that resulted in reduced daylight coverage over time. For example, the equatorial crossing time for NOAA-11 drifted from ~2 PM in 1989 to ~5 PM five years later, leading to limited SBUV coverage in 1994 (ozone observations were entirely unavailable south of 30° S during that austral winter). A similar orbital drift in the NOAA-17 satellite impacted the quality of the MERRA ozone products in 2012 before the introduction of observations from NOAA-19 SBUV in 2013. Outside of the exceptions described above and occasional short temporal gaps, SBUV provides good coverage of the sunlit atmosphere.

4.3.3 SPARC Data Initiative limb satellite observations

The SPARC Data Initiative (*Hegglin et al.*, 2021; *SPARC*, 2017) data set includes monthly mean zonal mean climatologies of ozone (*Neu et al.*, 2014; *Tegtmeier et al.*, 2013) and water vapour (*Hegglin et al.*, 2013) from an international suite of satellite limb sounders. The zonal monthly mean climatologies have undergone a comprehensive quality assessment and are suitable for climatological comparisons of the vertical distribution and interannual variability of these constituents in reanalyses on monthly to multi-annual timescales. We use a subset of the instrumental records available, as specified below.

The observational multi-instrument mean (MIM) for ozone averaged over 2005 - 2010 is derived using the SPARC Data Initiative (in the following abbreviated as SDI) zonal monthly mean climatologies from ACE-FTS (v2.2), Aura MLS (v2-2), MIPAS (v220), and OSIRIS (v5-0). These instruments provide data for the full 6 years considered and show inter-instrument differences with respect to the MIM that are generally smaller than ± 5 % throughout most of the stratosphere. Hence, temporal inhomogeneities that could affect the MIM are avoided and the standard deviation in the MIM is relatively small. Differences from the MIM in the lower mesosphere and tropical lower stratosphere are somewhat higher (± 10 %) (*Tegtmeier et al.*, 2013).

The evaluation of the ozone QBO signal for 2005 - 2010 is based on the instruments OSIRIS, GOMOS, and Aura MLS, which produce the most consistent QBO signals (*Tegtmeier et al.*, 2013).

The observational MIM for water vapour averaged over 2005 - 2010 is derived using the SDI zonal monthly mean climatologies from Aura MLS (v3.3), MIPAS (V5r_H2O_220), ACE-FTS (v2.2), and SCIAMACHY (v3.0). These instruments show inter-instrument differences that are generally within ± 5 % of the MIM throughout most of the stratosphere (*Hegglin et al.*, 2013). Differences from the MIM in the tropical upper troposphere increase to ± 20 %.

4.3.4 Aura MLS satellite data

The evolution of ozone in the reanalyses is compared with that observed by Aura MLS. This instrument measures millimeter- and submillimeter-wavelength thermal emission from Earth's atmosphere using a limb viewing geometry. *Waters et al.* (2006) provide detailed information on the measurement technique and the Aura MLS instrument. Vertical profiles are measured every 165 km along the suborbital track with an along-track horizontal resolution of 200 ~ 500 km and a cross-track footprint of 3 ~ 9 km. Here we use version 4.2 (hereafter v4) MLS ozone measurements from September 2004 through December 2013. The quality of the MLS v4 data has been described by *Livesey et al.* (2017). The vertical resolution of MLS ozone is about 3 km and the single-profile precision varies with height from approximately 0.03 ppmv at 100 hPa to 0.2 ppmv at 1 hPa. The v4 MLS data are quality-screened as recommended by *Livesey et al.* (2017). V4 stratospheric (pressures less than 100 hPa) ozone values are within ~ 2 % of those in version 2.2 (v2), which is the version assimilated in MERRA-2 (until 31 May 2015, after which v4 data are used) and ERA-Interim. At pressures greater than 100 hPa, v4 MLS ozone shows high and negative biases with respect to v2 at alternating levels, indicating improvement of vertical oscillations seen in v2 (*Livesey et al.*, 2017) and v3 (*Yan et al.*, 2016).

4.3.5 SWOOSH merged limb satellite data record

The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database is a monthly-mean record of vertically resolved ozone and water vapour data from a subset of limb profiling satellite instruments operating since the 1980s (*Davis et al.*, 2016). The SWOOSH version 2.6 data used here include individual satellite source data from SAGE-II (v7), SAGE-III (v4), UARS MLS (v5/6), UARS HALOE (v19), and Aura MLS (v4.2), as well as a merged data product. A key aspect of the merged product is that the source records are homogenised to account for inter-satellite biases and to minimize artificial jumps in the record. The homogenization process involves adjusting the satellite data records to a "reference" satellite using coincident observations during time periods of instrument overlap. SWOOSH uses SAGE-II as the reference for ozone and Aura MLS as the reference for water vapour. SWOOSH merged product data are used for time series evaluations that start before 2004, prior to the availability of Aura MLS. After August 2004, the SWOOSH merged product is essentially the same as the v4.2 Aura MLS data.

4.4 Evaluation of reanalysis ozone products

4.4.1 Total column ozone seasonal cycle

In this section, we compare SBUV TCO data to reanalysis products over the 1981–2010 climatology period. **Figure 4.3** shows the seasonal cycle in total column ozone from SBUV as a function of latitude and month. Also shown are the differences between TOMS/OMI and SBUV, and between the different reanalyses and SBUV. The climatological TCO fields of the TOMS/OMI and the reanalyses are given as line contours in the difference plots. **Figure 4.4** shows the equivalent comparison for TOMS/OMI data. The reanalyses all reproduce the major features of the seasonal cycle and latitudinal distribution of TCO. This agreement is not surprising given that all of reanalyses shown in **Figures 4.3** and **4.4** assimilate TCO data from one of the two satellites (**Figure 4.1**). As such, the comparisons here do not represent independent validation of ozone in reanalyses but rather represent a test of the internal consistency of the ozone data assimilation system. Hence it is not surprising that MERRA and MERRA-2 generally perform better against SBUV than against TOMS/OMI, while ERA-Interim and JRA-55 generally perform better against TOMS/OMI than against SBUV, since MERRA and MERRA-2 assimilate SBUV (but not TOMS/OMI), while ERA-Interim and JRA-55 primarily assimilate TOMS/OMI (but not SBUV).

Although the reanalysis TCO fields look quite similar, a handful of widespread biases are revealed by considering the differences between reanalyses and observations. The agreement between the two observational TCO data sets is within approximately ± 6 DU (2 - 3 %), with SBUV generally having smaller values in the tropics and larger values at high latitudes relative to TOMS/OMI. Differences between the reanalyses and the TCO observations are generally slightly larger than the difference between the two observational data sets. ERA-40 produces substantially larger TCO values than observed, particularly at higher latitudes. JRA-25 contains significantly smaller TCO values than observed (~10 DU less), except during the springtime at high southern latitudes.

For reanalyses that only (or mainly) assimilate UV-based retrievals, the winter hemisphere high latitudes remain largely unconstrained by data assimilation. The impact of the TCO observations may also be limited by filtering choices. For example, assimilated observations are filtered to exclude low solar elevation angles (less than 10 ° for TOMS

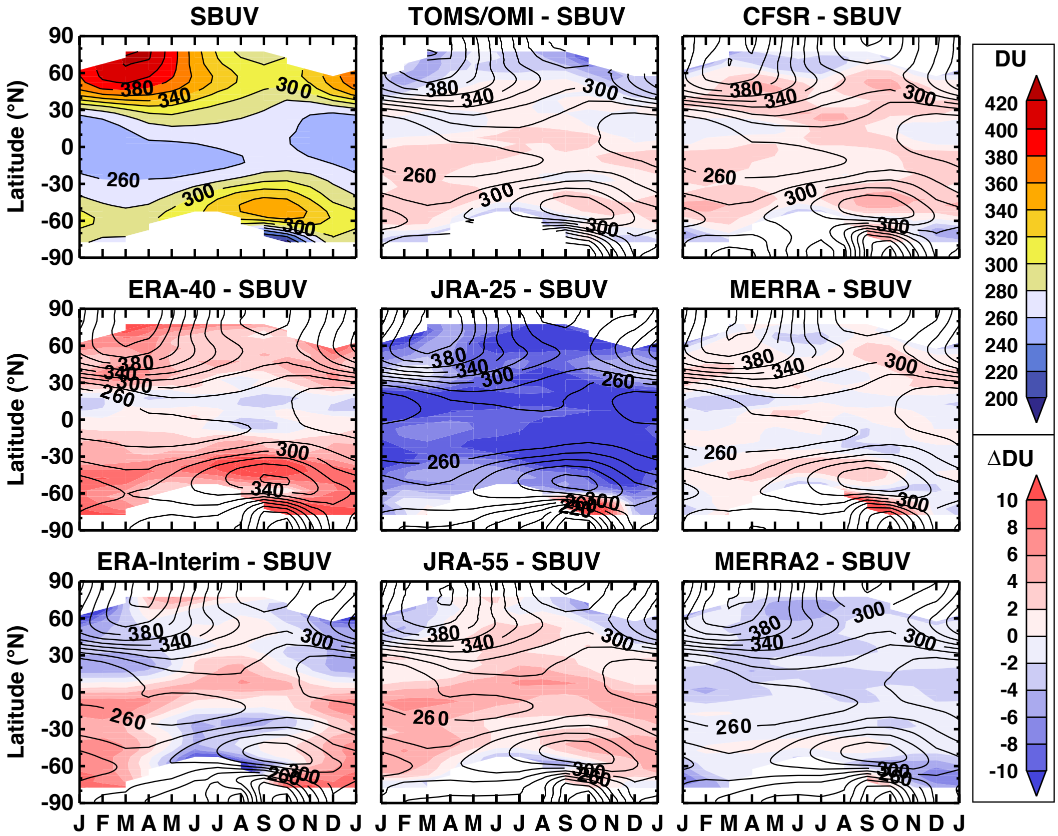


Figure 4.3: *Zonal- and monthly-mean total column ozone climatology over 1981 - 2010 from SBUV observations (uppermost left panel), along with the absolute differences between each reanalysis and SBUV. The difference between TOMS/OMI and SBUV is also shown (uppermost middle panel). Line contours show each reanalysis' respective climatology. Both climatology and observational reference to calculate differences for ERA-40 are for the time period Jan 1981 - Aug 2002 in order to avoid sampling issues. Reproduced from Davis et al. (2017).*

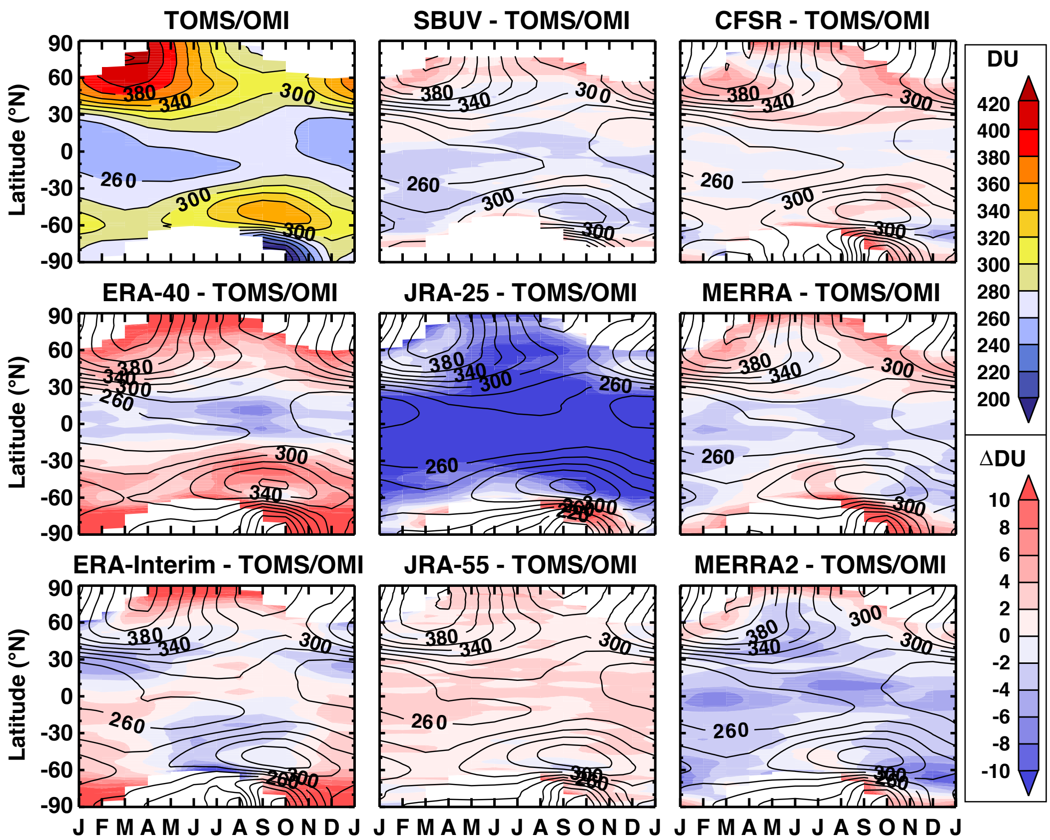


Figure 4.4: *Same as* ***Fig. 4.3****, except using TOMS/OMI as the observational data set. Reproduced from Davis et al. (2017).*

and less than 6° for SBUV) in both ERA-40 and ERA-Interim. This filtering further limits observational impacts on the ozone analyses at higher latitudes. Hence, for ERA-Interim, before the start of the Aura MLS assimilation in 2008, high latitude ozone fields essentially reflect the effects of transport and the ozone parameterization used. For ERA-40, *Dethof and Hólm* (2004) showed that the ozone model produces positive biases in ozone concentrations at high latitudes ranging from ~20 DU in the summer hemisphere to ~50 DU in the winter hemisphere, which is broadly consistent with the comparison shown in **Figure 4.3**.

4.4.2 Zonal mean ozone cross-sections

In this section, we compare zonal mean multi-annual mean latitude-altitude cross-sections of ozone between the different reanalyses and the SDI MIM. We perform the comparison for 2005 - 2010 using the subset of instruments described in *Section 4.3.3*. This shorter period has been chosen to avoid sampling issues that could be introduced by changes in instrument availability, which could alter sampling patterns, or trends in the constituents, such as the increase in ozone depletion from the 1970s to the mid 1990s. ERA-40 is excluded from this and all other comparisons with the SDI MIM because it ended in 2002.

Figure 4.5 shows multi-annual zonal mean ozone from the SDI MIM and the relative differences between each reanalysis and the SDI MIM (calculated as $100 \times (R_i - MIM)/MIM$, where R_i is the reanalysis field). Also indicated using contours are the climatological ozone distributions of the reanalyses. The reanalyses all capture the general zonal mean distribution of ozone, including the global maximum in ozone volume mixing ratio in the tropical middle stratosphere and the tropopause-following isopleths immediately above the tropopause. Among the reanalyses, MERRA-2 best reproduces this overall structure, with relative differences within ±5% throughout the middle and upper stratosphere. MERRA, CFSR, ERA-Interim, and ERA5 also perform generally well, but with MERRA overestimating concentrations in the ozone maximum (~10 hPa) relative to the SDI MIM. ERA-Interim shows relatively good agreement in the middle stratosphere with biases smaller than ±5% but includes a negative bias with magnitudes greater than 10% in the upper stratosphere. ERA5 generally improves over ERA-interim in the middle stratosphere at all latitudes and in the UTLS at mid- to high-latitudes. However, the differences to the MIM increase around the tropical tropopause when compared to ERA-interim. All reanalyses (except ERA5, which shows generally smaller differences from the MIM) show biases exceeding ±10% in the lowermost stratosphere, at pressures greater than 100 hPa. JRA-55 is an evident improvement relative to JRA-25, particularly in the polar regions. Negative biases in JRA-55 have approximately halved in the middle and upper stratosphere, compared to JRA-25. However, JRA-55 also shows somewhat higher positive biases around the tropical upper troposphere and lower stratosphere than JRA-25. It is worth noting that the diurnal cycle in ozone (*e.g.*, *Parrish et al.*, 2014; *Sakazaki et al.*, 2013) has not been explicitly accounted for in the observational MIM. Neglecting the diurnal cycle potentially contributes to differences between the reanalyses and observations in the upper stratosphere and lower mesosphere.

Most reanalyses have a positive bias in ozone in the Southern Hemisphere (SH) lower stratosphere. This indicates an inability to simulate Antarctic ozone depletion accurately due to a combined effect of limited data coverage, data filtering, and limitations of the

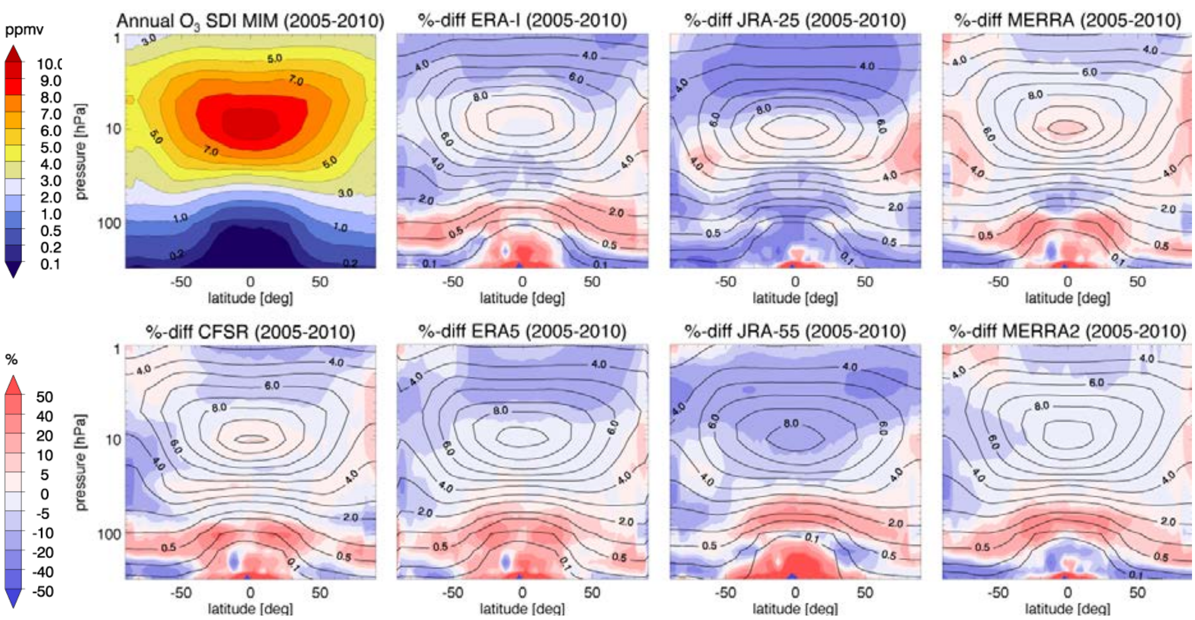


*__Figure 4.5:__ Multi-annual zonal mean ozone latitude-altitude cross-sections averaged over 2005 - 2010 for the SPARC Data Initiative multi-instrument mean (SDI MIM) (upper left), along with the relative differences between reanalyses and observations as (R_i–MIM)/MIM*100, where R_i is a reanalysis field. Also shown in contours are the respective zonal mean climatologies for the different reanalyses. Updated from Davis et al. (2017).*

reanalyses' chemistry schemes at high latitudes (*Section 4.4.1*). A dipole is apparent in the CSFR and ERA-Interim biases, with a positive bias near ~100 hPa located below a negative bias near ~10 hPa. This dipole may reflect a lack of information about the vertical location of the ozone hole in the TCO and SBUV observations assimilated by these systems. In contrast, MERRA includes a significant positive bias (>10 %) at Southern high latitudes that extends throughout the stratosphere.

4.4.3 Ozone monthly mean vertical profiles and seasonal cycles

Figures 4.6a and **b** show vertical profiles of ozone for January (2005 - 2010 average) for the reanalyses and the SDI MIM at two different latitudes, 40° N and 70° S, respectively, along with the relative differences for each reanalysis with respect to the MIM. In addition, **Figures 4.6c-e** and **f-h** show the seasonal cycles of ozone for three different pressure levels at 40° N and 70° S, respectively. The vertical profiles and the seasonal cycles reveal seasonal information on reanalyses-observation differences that expands upon the annual zonal mean evaluation presented in *Section 4.4.2*. In general, the results shown reinforce the conclusions of the previous section.

Most reanalyses resolve the vertical distribution in January reasonably well at both latitudes, in particular in the middle stratosphere between around 50 hPa and 5 hPa. MERRA-2, MERRA, CFSR, and ERA5 perform particularly well. At 70° S, JRA-25 is a clear outlier that produces too little ozone in the vicinity of the maximum. JRA-55 and ERA-Interim also underestimate ozone concentrations above the ozone maximum by between 10 % and 20 % but are not as strongly biased as JRA-25 (which produces differences of more than 30 %). All reanalyses show larger percentage differences from the MIM in the lower part of the profile at pressures greater than 100 hPa. The reanalyses seem to overestimate ozone at around 150 hPa by 20 % in the Southern high latitudes, possibly related to not capturing accurately enough the extent of ozone depletion during spring.

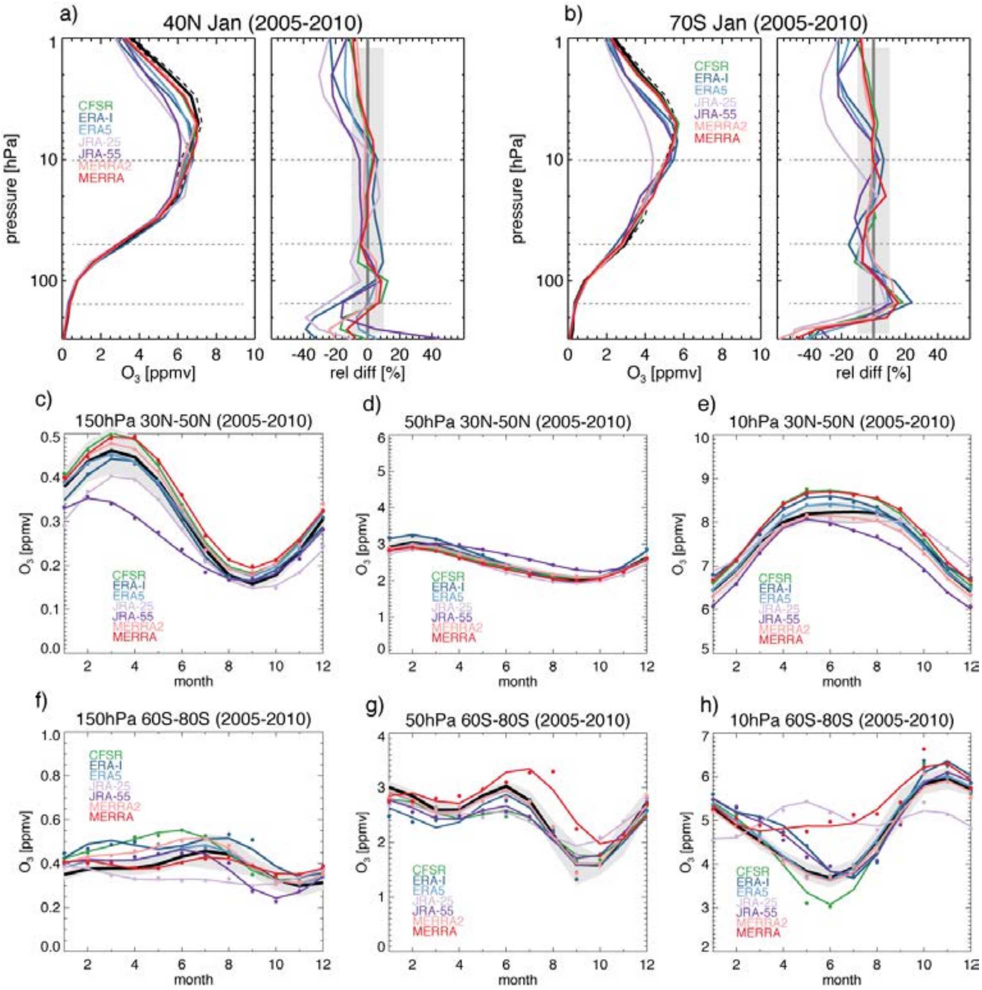


Figure 4.6: *Multi-annual mean vertical ozone profiles over 2005 - 2010 for January at (a) 40° N and (b) 70° S from the SPARC Data Initiative multi-instrument mean (SDI MIM) (black) and the six reanalyses (coloured). Absolute values are shown in the left and relative differences in the right panels for each comparison. Relative differences are calculated as (R_i–MIM)/MIM*100, where R_i is a reanalysis profile. Black dashed lines provide the ±1-sigma uncertainty (as calculated by the standard deviation over all instruments and years available) in the observational mean. Horizontal dashed lines in grey indicate the pressure levels (150, 50, and 10 hPa) for which seasonal cycles are shown in panels (c) and (d) for the two latitude ranges 30°-50° N and 60°-80° S, respectively. Grey shading indicates observational uncertainty (±1-sigma) calculated as the standard deviation over all instruments and years available. Updated from Davis et al. (2017).*

Below 200 hPa at both latitudes, all reanalyses underestimate observed ozone values. An exception to this is ERA5, which shows much smaller differences to the MIM of less than ± 10 % at 40 ° N.

The agreement between the reanalyses and observations varies by month, as can be seen in **Figures 4.6c-e** and **f-h**, which show the annual cycle for selected pressure levels (150, 50, and 10 hPa) and somewhat extended latitude bands of 30 ° N - 50 ° N and 60 ° S - 80 ° S, respectively. The agreement in the ozone seasonal cycle between the SDI observations and the reanalyses is better in the Northern Hemisphere (NH) mid-latitudes (where the seasonal cycles have a simple sinusoidal structure) than in the SH high latitudes. In the NH at 50 hPa and 150 hPa, ozone reaches its annual maximum during boreal spring and its annual minimum during autumn, attributable to the strong seasonality in the Brewer-Dobson circulation. The seasonal cycle is shifted at 10 hPa, with a maximum in summer and a minimum in winter, attributable mostly to ozone photochemistry. Most of the reanalyses produce a fairly accurate ozone evolution at these levels with exceptions as follows: At 150 hPa, JRA-55 shows a strong negative bias when compared to both observations and the other reanalyses during the NH winter/spring months. All the other reanalyses (except ERA5, which shows nearly perfect agreement with the observations) tend to overestimate the absolute ozone values, but agree rather well with the seasonal cycle in the observations in terms of amplitude and phase. At 50 hPa, the seasonal cycle produced by JRA-55 shows a more gradual decline in ozone concentrations into autumn relative to both observations and other reanalyses. ERA-Interim, MERRA, and CFSR at 10 hPa tend to overestimate ozone during spring and early summer, while JRA-55 (JRA-25) tends to underestimate (overestimate) ozone during fall and winter. ERA5 tends to agree also at these other levels best with the observations.

Seasonal cycles in SH high latitudes have a more complex structure than those in the NH mid-latitudes due to generally weaker downwelling in the Brewer–Dobson circulation and the influence of Antarctic ozone depletion. As a consequence, the reanalyses have more difficulty in capturing the seasonal cycle. At 10 hPa, MERRA-2 and ERA5 show the best agreement with the observations. CFSR also follows the observations relatively well, but overestimates the amplitude of the seasonal cycle, primarily because of values that are too low during May through July. MERRA and JRA-25 are outliers in that they do not contain the strong annual minimum observed during late austral autumn and early winter. At 50 hPa, MERRA and JRA-25 agree better with observations than at 10 hPa, but still underestimate austral springtime ozone depletion.

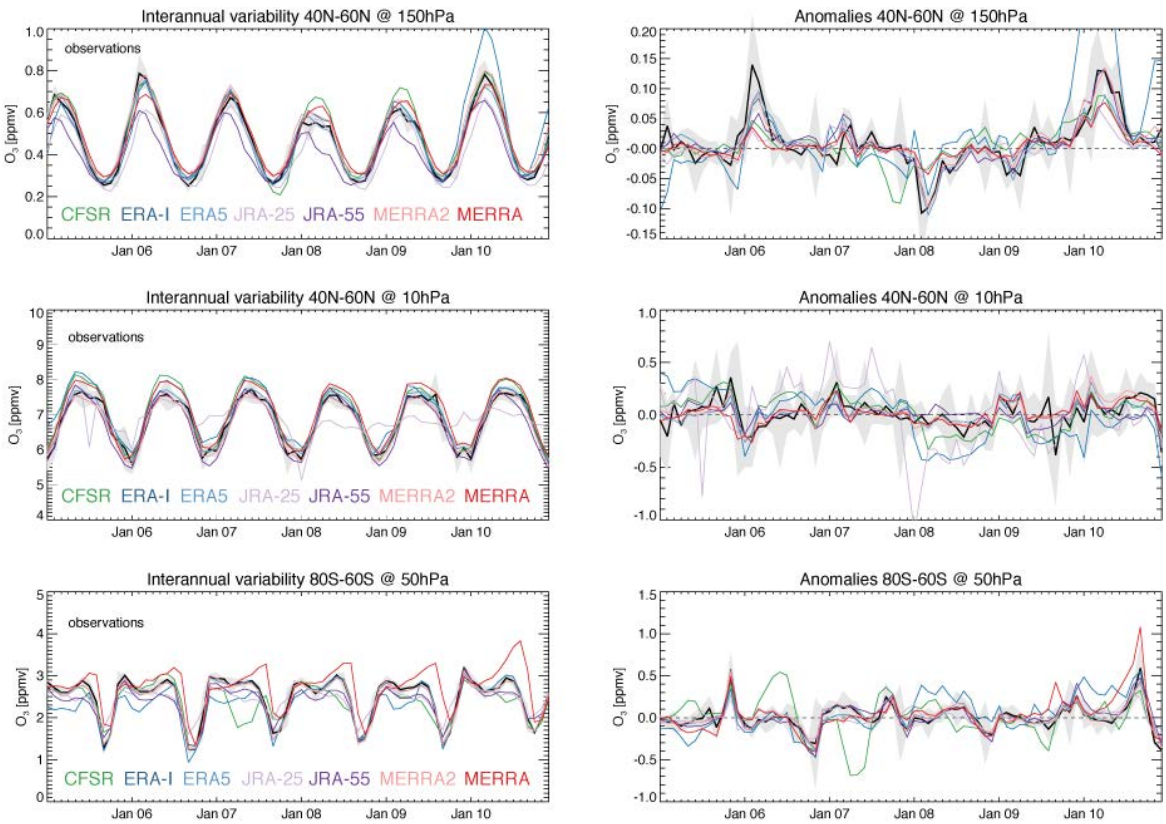


Figure 4.7: *Interannual variability (left column) and deseasonalized anomalies (right column) for ozone during 2005 – 2010 for the SPARC Data Initiative multi-instrument mean (SDI MIM, black) and the six reanalyses (coloured). Results are shown for three different pressure levels and latitude ranges (top to bottom: 150 hPa at 40°– 60 °N, 10 hPa at 40 ° – 60 °N, and 50 hPa at 60°– 80 °S). Grey shading indicates observational uncertainty (± 1-sigma) calculated as the standard deviation over all instruments and years available. Updated from Davis et al. (2017).*

They are also phase shifted, with the MERRA peak at 50 hPa occurring one month later than the SDI observations, and the JRA-25 peak occurring one month earlier. Finally, at 150 hPa, the seasonality in the reanalyses varies widely and is inconsistent with that in the observations, with the exception of MERRA and ERA5, which produce the most realistic seasonal cycle amplitude. MERRA-2 and ERA5 show the closest agreement with observations at all levels, with the exception of MERRA-2 at 150 hPa, which is the next to lowest valid level of the MLS v2.2 ozone retrievals that it assimilates.

4.4.4 Ozone interannual variability

Figure 4.7 shows time series of interannual variability of ozone and its anomalies in the SDI MIM and reanalyses during 2005 - 2010. The anomalies, which are calculated for each reanalysis by subtracting multi-year monthly means averaged over 2005 - 2010 from the monthly mean time series, are a good indicator of how well physical processes (such as transport) are represented in reanalyses. Time series are shown for the SH high latitudes (averaged over 60°S-80°S) at 50 hPa, and for the NH mid-latitudes (40°N - 60°N) at 150 hPa and 10 hPa. In all cases, MERRA-2 and ERA5 produce the closest match with the SDI MIM in terms of both the absolute values and the structure of its interannual variability. This agreement highlights the benefit of assimilating vertical profile observations from a limb-viewing satellite instrument. Although it has to be noted that the comparison is not done against truly independent observations in this case, since Aura MLS (v2.2) is included in the SDI MIM. MERRA-2 (which assimilates v2.2 for the time period of the comparison) is an evident improvement over MERRA, which tends to disagree with the absolute ozone values of the observations at 150 hPa and to overestimate them at 10 hPa, and to underestimate interannual variability at both levels in the NH mid-latitudes. JRA-55 also shows clear improvement relative to JRA-25 with respect to the amplitude and structure of interannual variability, at least at 10 hPa in the NH mid-latitudes. Large excursions seen in JRA-25, such as the sudden drop in ozone at the beginning of 2008, are not present in JRA-55 or in the observations.

Although ERA-Interim ozone mean values mostly agree well with observations, the amplitude of its interannual variability is larger than observed. In particular, ERA-Interim overestimates the negative anomaly in NH midlatitudes at 10 hPa, and the positive anomaly in SH high latitudes at 50 hPa during 2008. The largest differences appear to affect ERA-Interim from mid-2009 when the assimilation of Aura MLS data restarted with the (v3) NRT product after months of data unavailability. All these problems seem to be resolved in ERA5, with ERA5 showing similarly good agreement with the observations as MERRA-2. The improvement may be at least partially explained by the use of a newer version of Aura MLS data (v4.2) in the assimilation system. Finally, CSFR also produces large interannual excursions during certain years (*e.g.,* during spring 2006 and 2007 at 50 hPa in SH high latitudes). This issue may be related to SBUV only offering measurements between September to March, so that the assimilation system is not well constrained during the remainder of the year.

4.4.5 Ozone time series in equivalent latitude coordinates

Equivalent latitude (EqL) is a common vortex-centred coordinate used in studies of the stratosphere (*e.g.*, *Manney et al.*, 1999; B*utchart and Remsberg,* 1986; and references therein). This coordinate is also useful as a geophysically-based coordinate in the UTLS (*e.g.*, *Santee et al.*, 2011), although interpretation becomes more complicated in this context (*e.g.*, *Pan et al.*, 2012; *Manney et al.*, 2011). The equivalent latitude of a potential vorticity (PV) contour is defined as the latitude of a circle centred about the pole enclosing the same area as the PV contour (see *Hegglin et al.*, 2006 for a visual illustration). **Figure 4.8** shows the time series of v4 MLS ozone (*Section 4.3.4*) for late 2004 through 2013 in the lower stratosphere (520 K), along with differences between MERRA, MERRA-2, ERA-Interim, CFSR, and JRA-55 and MLS ozone at the same level. MLS ozone is interpolated to isentropic surfaces using temperatures from MERRA. The EqL ozone time series are then produced using a weighted average of MLS data in EqL and time, with data also weighted by measurement precision (*e.g.*, *Manney et al.*, 2007; *Manney et al.*, 1999). **Figures 4.9-4.10** show the equivalent evaluation for the 350 K and 850 K potential temperature levels.

Figure 4.8 reveals that MERRA-2 matches MLS more closely over the full period than do the other reanalyses. This is expected because the stratospheric ozone reanalyses in MERRA-2 are largely constrained by the MLS stratospheric ozone profiles (v2 for the period shown here) and OMI column ozone beginning in October 2004 (in fact, at 850 K, a suggestion of poorer agreement can be seen in September 2004). This agreement is especially apparent during Antarctic winter and spring, when other assimilated ozone products (*e.g.*, SBUV/2 and TOMS) cannot provide measurements due to darkness and simplified chemical parameterizations cannot adequately represent heterogeneous loss processes. The improved vertical resolution of MLS relative to SBUV/2 also better constrains the structure of the ozone hole, which is vertically limited. ERA-Interim also shows close agreement with MLS during the periods when it assimilates MLS ozone products (2008 and mid-2009 through present).

Biases in the reanalyses that do not assimilate MLS and OMI ozone vary in magnitude and sign, not only among the reanalyses but also with altitude and latitude (see also **Figures 4.9 - 4.10**). Positive biases in MERRA and CFSR ozone during Arctic winter may be partially related to inadequate representations of ozone chemistry and an overall lack of measurements.

We speculate that the latter is dominant due to the appearance of these biases even during years with minimal observed chemical ozone loss. JRA-55 biases become strongly negative in the upper stratosphere (**Figure 4.10**). These large biases in JRA-55 suggest that column ozone alone is insufficient to properly constrain the CTM used in the offline calculation close to observations. Each of the reanalyses except MERRA-2 shows a quasi-biennial pattern in the tropical differences from MLS, indicating deficiencies in the reanalysis representation of the QBO (see *Chapter 9*).

In the UTLS (*e.g.*, 350 K, **Figure 4.9**), significant biases are present in all reanalyses at middle and high latitudes (*i.e.*, poleward of the latitude at which the tropopause intersects the 350 K isosurface, thus in the lowermost stratosphere), but are relatively small. MERRA-2 biases are slightly smaller than those in the other reanalyses, and the biases in ERA-Interim change character noticeably at the beginning of 2008 when MLS and OMI ozone are first assimilated. Seasonally varying biases just poleward of the tropopause are pervasive in the reanalyses.

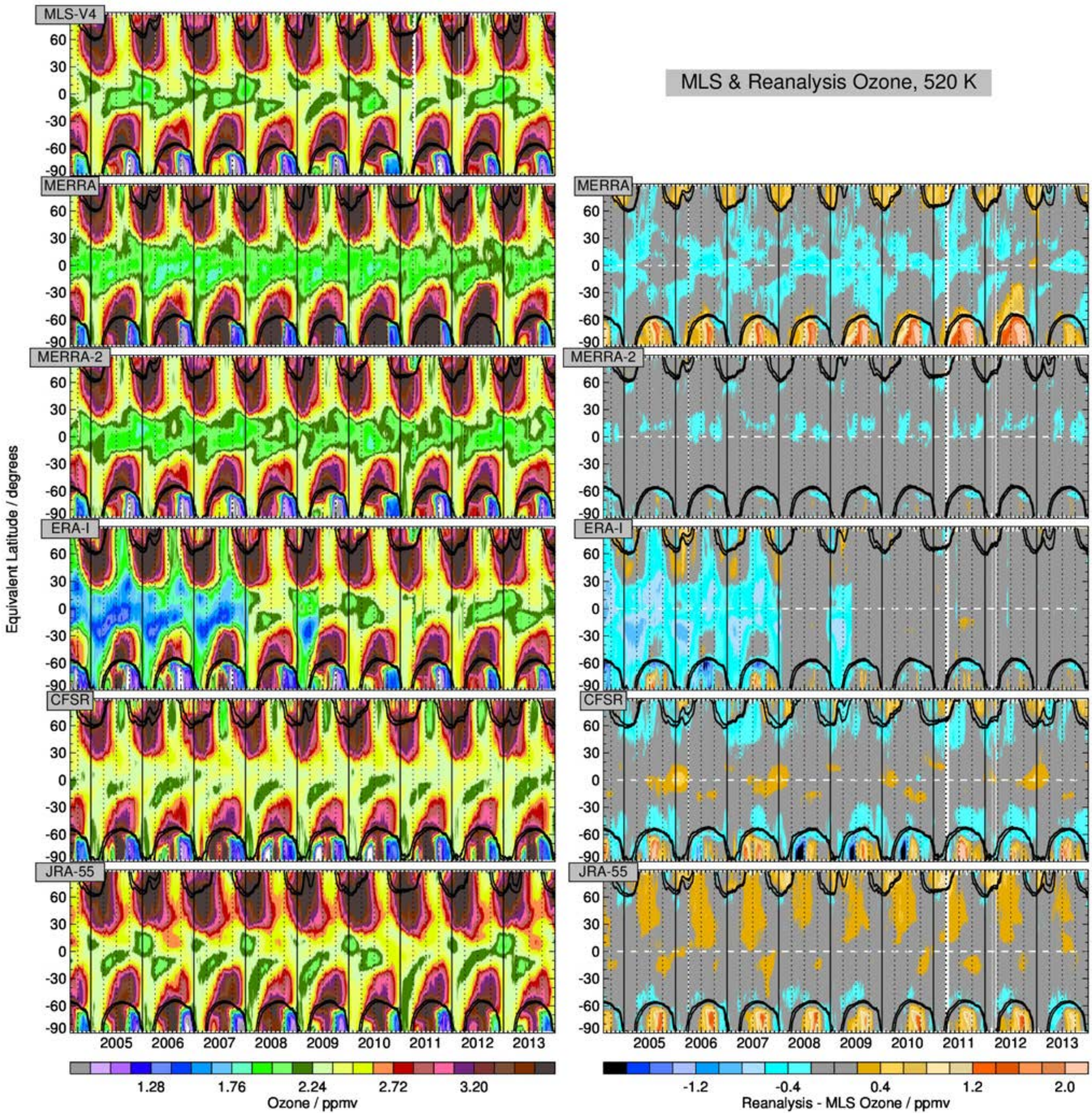


Figure 4.8: Comparison of the equivalent latitude–time evolution of each reanalysis ozone field and MLS on the 520 K isentropic surface (~50 hPa; ~20 km altitude) during the Aura mission September 2004 - December 2013. (Left) Mixing ratios (ppmv) for MLS and the reanalyses MERRA, MERRA-2, ERA-Interim, CFSR, and JRA-55 (top to bottom). (Right) differences (ppmv) between each reanalysis and MLS (R_i – MLS). Overlays are scaled potential vorticity (Manney et al., 1994) contours of 1.4 and 1.6 x $10^{-4}\,s^{-1}$ from the corresponding reanalysis, which are intended to represent the wintertime polar vortex edge. Dynamical fields for the MLS panel are from MERRA. Reproduced from Davis et al. (2017).

It is possible that these biases are caused by variations in the ability of the reanalysis to capture quasi-isentropic stratosphere-troposphere exchange (STE) processes. However, it is worth noting that the small absolute differences on the tropical side of the tropopause in **Figure 4.9** could still be quite large in a relative sense, given the low amount of ozone in that region.

4.4.6 Ozone quasi-biennial oscillation

Variations in transport and chemistry associated with the quasi-biennial oscillation (QBO) in tropical zonal wind are among the largest influences on interannual variability in equatorial ozone. The QBO signal in tropical ozone has a double-peaked structure with maxima in the lower (50 - 20 hPa) and the middle-to-upper (10 - 2 hPa) stratosphere (*Hasebe*, 1994; *Zawodny and Mccormick*, 1991). Ozone is mainly under dynamical control below 15 hPa, where the QBO signal results primarily from changes in ozone transport due to the QBO-induced residual circulation. In contrast, ozone is under photochemical control above 15 hPa. The QBO signal in these upper levels is understood to arise from a combination of QBO-induced temperature variations (*Zawodny and Mccormick*, 1991 ; *Ling and London*, 1986) and QBO-induced variability in the transport of NO_y (*Chipperfield et al.*, 1994). As a result, ozone anomalies in the middle/upper stratosphere show the opposite phase relationship with zonal winds compared to ozone anomalies in the lower stratosphere.

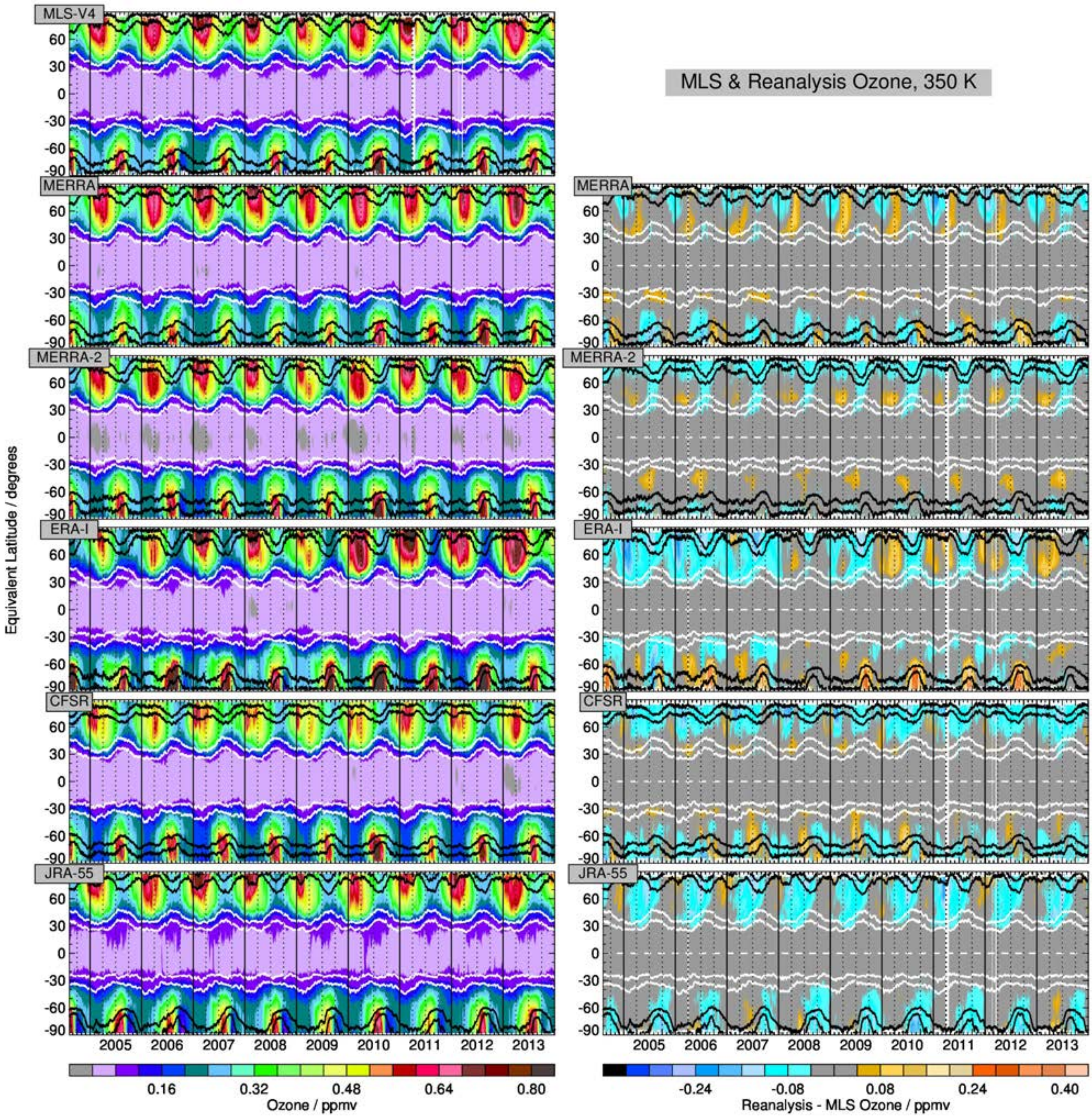


Figure 4.9: *As in* ***Figure 4.8****, but at 350 K. The white contours are PV values of magnitude 1.5 and 4.5 PVU, bounding the range commonly used to define the dynamical tropopause. Reproduced from Davis et al. (2017).*

A realistic characterization of the time–altitude QBO structure is an important aspect of physical consistency in ozone data sets.

Figure 4.11 shows time–altitude cross sections of deseasonalised ozone anomalies from 2005 to 2010 from the SDI MIM, along with the differences between the ozone anomaly fields from the reanalyses and the SDI MIM. The climatological QBO anomaly fields of the reanalyses are given as contours in the difference plots. Combined ozone measurements from the limb-viewing satellite instruments show a downward propagating QBO ozone signal with a shift in the phase relative to zonal winds around 15 hPa, as expected based on the known transition from photochemical to dynamical control of ozone at this level. All reanalyses exhibit some degree of quasi-biennial variability; however, differences are evident in the phase, amplitude, vertical extent, and downward propagation of these signals. The largest deviations from observations are in JRA-25, which displays positive anomalies from 2005 to mid-2007 followed by negative anomalies from mid-2007 through 2010 in place of the QBO signal above 15 hPa. In contrast, ERA-Interim shows predominantly negative anomalies in the 100 - 10 hPa pressure range before 2008 and positive anomalies afterwards. The changes in ERA-Interim coincide with the beginning of the assimilation of Aura MLS profiles beginning in 2008, which caused a shift to positive anomalies. Negative anomalies are present during the first half of 2009 when no MLS data were assimilated, followed by positive anomalies after the reintroduction of MLS data in June 2009 (*Section 4.2.6*).

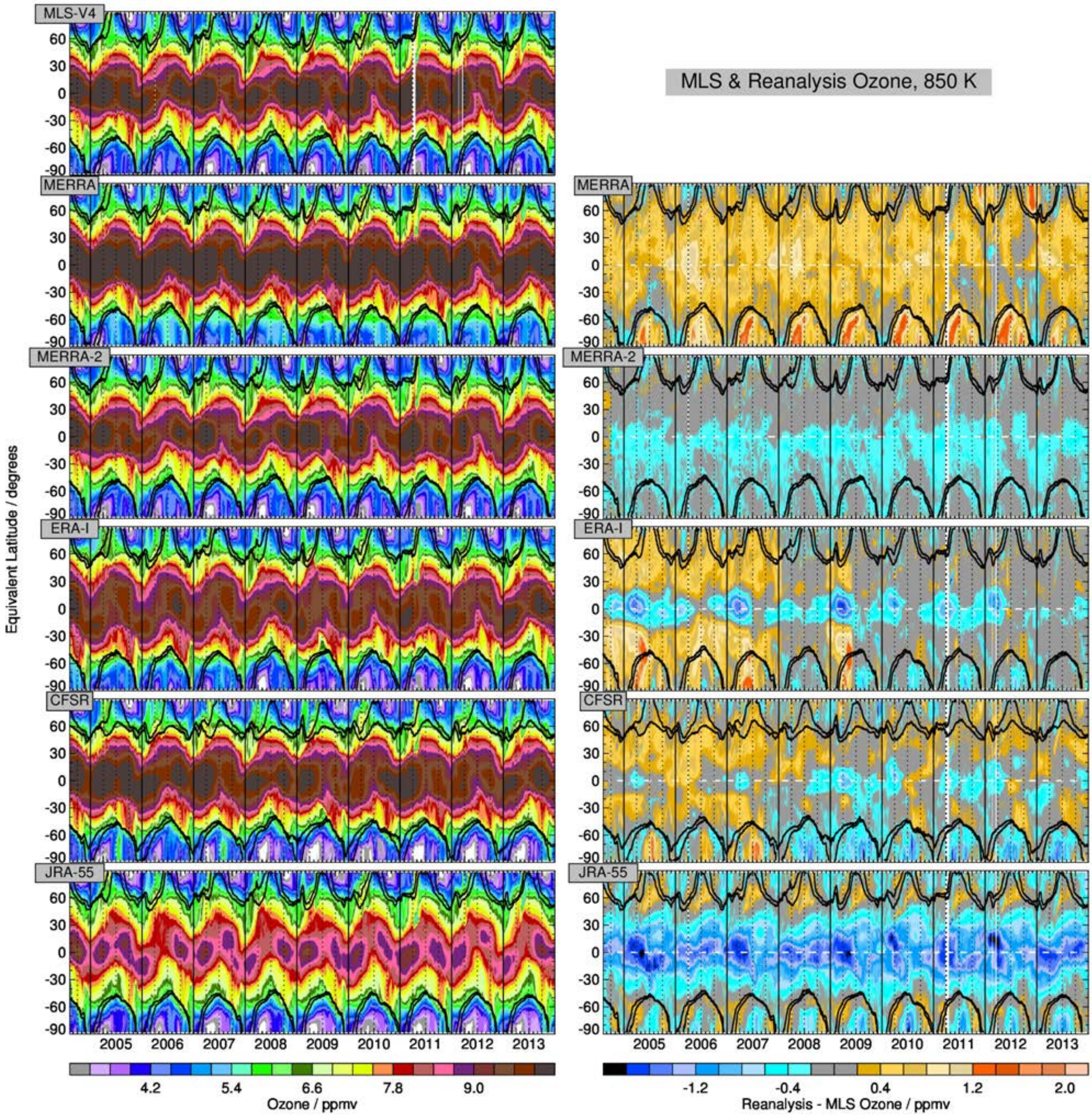


Figure 4.10: *Same as* ***Figure 4.8****, but at 850 K. Reproduced from Davis et al. (2017).*

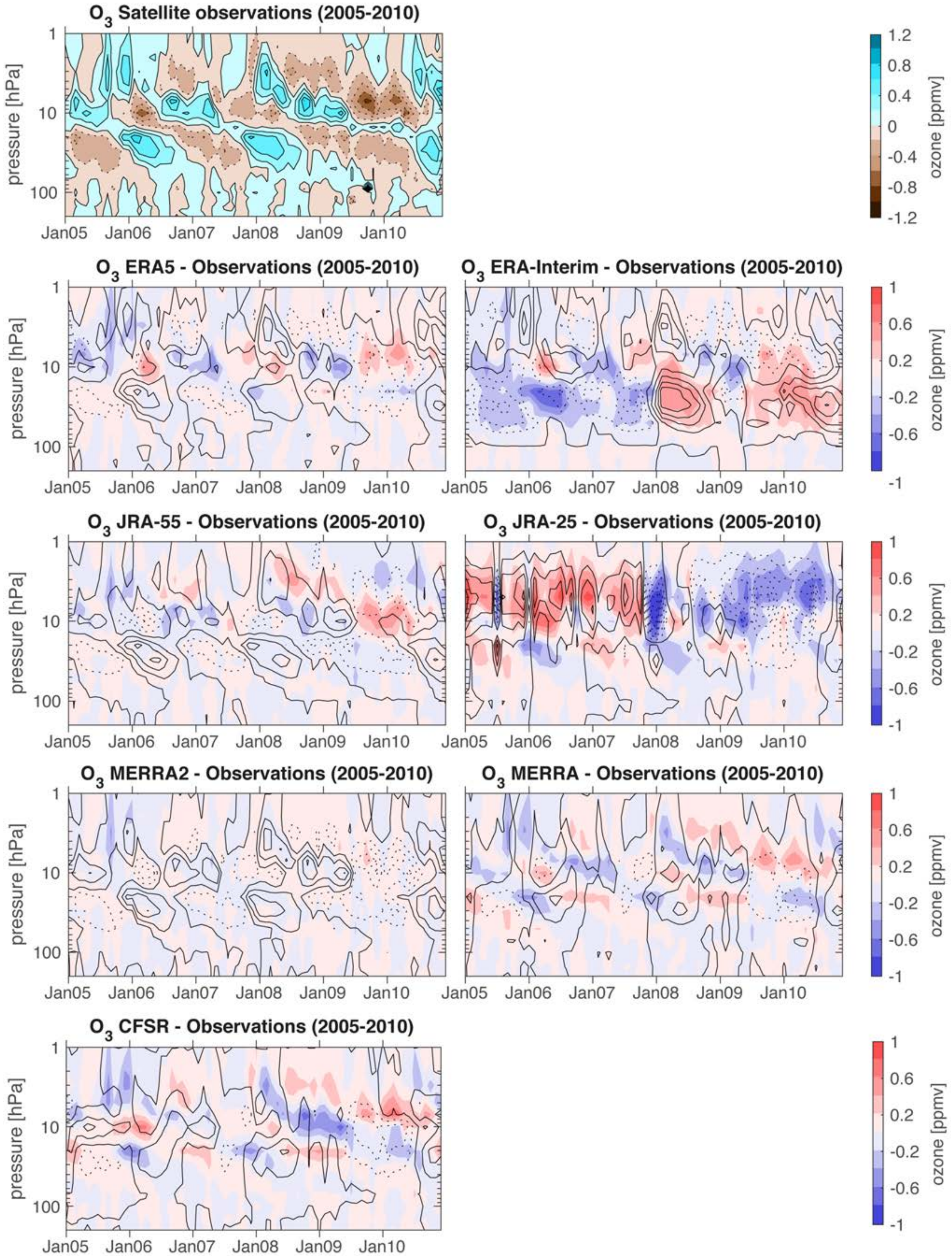


Figure 4.11: *QBO ozone signal from the SPARC Data Initiative observations (upper left) during 2005 - 2010, defined as altitude–time cross-sections of deseasonalized ozone anomalies averaged over the 10° S–10° N tropical band. Observations are based on three satellite data sets. The other panels show the differences in QBO ozone signals between each reanalysis and the observations (R_i–MIM) with the black contours (0.2 ppmv interval, with dotted lines showing negative anomalies and solid lines for positive anomalies) showing the QBO ozone signal generated by each corresponding reanalysis. Updated from Davis et al. (2017).*

The QBO ozone signal is much improved in ERA5 over ERA-Interim, with anomalies that are roughly consistent in amplitude and frequency with the QBO ozone signal in the satellite data. In particular, the lower and middle stratospheric biases seen in ERA-Interim are largely removed. This improvement is at least partially attributable to MLS data being assimilated over the whole Aura mission time period in ERA5.

CFSR and MERRA produce anomalies that are also roughly consistent in amplitude and frequency with the QBO ozone signal in the satellite data. However, no clear downward propagation is apparent in these reanalyses. The vertical structure of the anomalies is also shifted. Instead of a pair of peaks in the lower stratosphere (50 - 20 hPa) and middle-to-upper stratosphere (10 - 2 hPa), a single peak emerges near 15 hPa. This finding may be at least partially explained by the fact that the only vertically resolved ozone measurements assimilated by CFSR and MERRA come from SBUV. SBUV shows only a weak oscillatory behaviour, with a much smaller amplitude and without a properly downward propagating signal, attributable to the instrument's vertically limited and rather low vertical resolution (*Kramarova et al.*, 2013; *McLinden et al.*, 2009). JRA-55 and MERRA-2 produce a phase and amplitude of QBO variability like those observed in the satellite data. Overall, the features of the QBO (including the downward propagation) are much improved in MERRA-2 relative to MERRA (*Coy et al.*, 2016), and in JRA-55 relative to JRA-25. Nearly all reanalysis data sets extend the QBO ozone signal to altitudes below 100 hPa. This upper tropospheric signal is not present (or not captured) in the satellite observations, although it is worth noting that these observations have higher uncertainties that may potentially mask QBO signals below 100 hPa.

4.4.7 Ozone hole area

The Antarctic "ozone hole" is a region of severe ozone depletion that starts in late August or early September and lasts until November or early December. The ozone hole is commonly defined as the area within the 220 DU TCO contour. **Figure 4.12** shows average ozone hole areas based on TOMS/OMI observations and six reanalyses during 1981 - 2010. The average is computed over 21 September - 20 October of each year. This period is chosen to avoid the partial coverage of the SH high latitudes that occurs in TOMS/OMI data during the early part of September. Observationally based ozone hole areas are larger than those produced by the reanalyses in almost all years between 1981 and 2002. The systematic negative bias in reanalysis-based ozone hole areas is consistent with reanalyses generally underestimating ozone loss. Most of the reanalyses track the observations well starting in 2003, causing the timeseries of the differences to (**Figure 4.12b**) to display a long-term trend. This is not a truly independent comparison (all reanalyses except for MERRA assimilate TOMS and/or OMI observations); however, it does show the general consistency among most reanalyses in reproducing realistic interannual and decadal changes in the size of the Antarctic ozone hole, except for a few outliers discussed below.

The newer reanalyses (MERRA-2, ERA-Interim, JRA-55, and CSFR) are all within 1 million km^2 (5.2 %) of the observations, and generally produce root-mean-square (RMS) differences relative to TOMS/OMI of less than 0.9 million km^2 (14.6 %). A notable exception to the latter is MERRA-2 with an RMS of 2.8 million km^2 (44.5 %). This large RMS is attributable to an outlier year in 1994, when MERRA-2 had a very small ozone hole (**Figure 4.12**). JRA-55 produces the smallest RMS difference relative to TOMS/OMI, while MERRA-2 model produces the smallest mean difference relative to these observations.

MERRA did not produce an ozone hole in 1994, and produced very small ozone holes in 1993, 1997, 2009, and 2010. For related reasons, both JRA-55 and MERRA-2 did not produce an ozone hole in 1994, and produced a relatively small ozone hole in 1993. The elimination or reduction of the ozone hole during those years was caused by a lack of ozone observations for constraining the ozone field, as the processes that contribute to the development of the ozone hole are not represented in the parameterised ozone chemistry used in MERRA and MERRA-2. In 1994, orbital drift of the NOAA-11 satellite that provided the SBUV/2 TCO data assimilated by both MERRA and MERRA-2 led to a lack of ozone observations south of ~30 ° S during early Austral spring. NOAA-11 SBUV/2 coverage was also limited in 1993. While both MERRA and MERRA-2 use NOAA-11 SBUV, the version 8.6 data assimilated in the latter allowed less stringent quality screening criteria. Specifically, MERRA-2 uses observations made at solar zenith angles greater than 84 °, excluded in MERRA, if they are otherwise marked as "good". This results in a slightly better coverage of NOAA-11 SBUV in MERRA-2, explaining its better performance in 1993 and even 1994. The MERRA ozone hole was only weakly constrained by observations in late September 1997 because NOAA-11 data only extended to 60 ° S - 75 ° S between 21 September and 20 October. MERRA-2 does not have a negative bias in ozone hole size during 1997 because it used data from NOAA-14 rather than data from NOAA-11. The MERRA ozone hole was also affected by orbital drift in the NOAA-17 satellite and the concomitant loss of SBUV/2 observations at high southern latitudes during the austral springs in 2009 and 2010. MERRA-2 is unaffected during these years because of its assimilation of ozone observations from Aura OMI and MLS.

ERA-40 did not assimilate ozone data in 1989 and 1990. This resulted in a positive bias in ozone concentrations and a very small ozone hole. The ERA-40 model also severely underestimated ozone hole area in 1997, most likely due to a gap in assimilated TCO from the Earthprobe TOMS instrument between August and December that year (**Figure 4.1**; note that NOAA-9 SBUV/2 profiles were assimilated during this timeframe as shown in **Figure 4.2**).

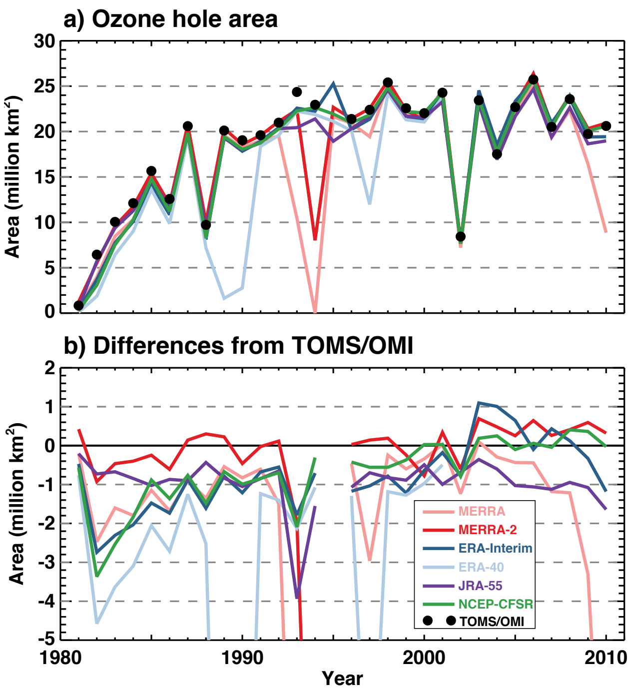


Figure 4.12: *(a) Ozone hole mean area calculated from TOMS/OMI observations and the reanalyses for 21 September through 20 October of 1981-2010. (b) Differences between ozone hole mean areas from reanalyses and TOMS/OMI observations (R_i – observed). Note, no TOMS data were available in 1995. Reproduced from Davis et al. (2017).*

By contrast, the area of the ERA-Interim ozone hole was too large in 1995 (see **Figure 4.14**.). This may be due to a lack of TCO observations for assimilation in ERA-Interim during 1995 (**Figure 4.1**).

4.4.8 Long-term evolution of ozone

Figure 4.13 shows the evolution of deseasonalised TCO anomalies from the reanalyses and assimilated observations from SBUV and TOMS/OMI. Also shown are the differences between the reanalyses and the primary TCO observations they assimilate. Both observational data sets show similar features, including a general trend toward decreasing ozone in the SH high latitudes, consistent with the Antarctic ozone hole depletion discussed in the previous section. However, in **Figure 4.13**, comparison to the data set assimilated by a given reanalysis is done because differences between the TOMS/OMI and SBUV data sets show an apparent step change at the beginning of 2004. For completeness, a comprehensive set of plots showing this step change, as well as reanalysis/observation differences separately for each data source, is provided in **Figures 4.14 - 4.15**.

As expected, reanalyses agree more closely with TCO data that they assimilate than with data that they do not assimilate. For example, MERRA, MERRA-2, and CFSR assimilate SBUV data. The influence of SBUV on these reanalyses can be seen in the QBO-related anomalies in the tropics (particularly after ~1998) that are present in both the SBUV data and in the reanalyses that assimilate it. Differences between these reanalyses and SBUV are smaller in magnitude and more homogeneous in space and time than differences between these reanalyses and TOMS/OMI. The discontinuity in 2004 is particularly pronounced when MERRA and CFSR are compared against TOMS/OMI (**Figure 4.15**). Similarly, differences between the ECMWF reanalyses and TOMS/OMI are generally more homogeneous and smaller in magnitude than differences between the ECMWF reanalyses and SBUV (**Figure 4.14**). The period during which ERA-40 did not assimilate any ozone data (1989 - 1990) is also evident in **Figure 4.13**. The stark contrast between this period and the surrounding years indicates the importance of data assimilation in constraining reanalysis ozone fields.

Figure 4.16 shows differences between reanalysis ozone fields and SWOOSH satellite limb profiler merged ozone data on two pressure levels (10 hPa and 70 hPa). This plot helps to evaluate disruptions in the temporal homogeneity of reanalysis ozone fields caused by changes in the assimilated observational data, and also provides a partially independent dataset for comparison with the reanalyses. The SWOOSH record is based primarily on v4.2 Aura MLS ozone starting in August 2004, so comparisons with reanalyses that assimilate MLS (i.e., MERRA-2 and ERA-Interim) after that time are not independent. However, none of the observations used to construct the SWOOSH record prior to August 2004 were assimilated by these reanalyses.

At 10 hPa, CSFR, MERRA, and MERRA-2 show the best agreement with observations. At this level, ERA-Interim and JRA-25 have positive biases in both SH and NH midlatitudes, while JRA-55 has a negative bias relative to SWOOSH in the tropics. ERA5 shows similar positive biases in NH and SH mid-to high latitudes as found in ERA-Interim through the 1990s, however these drop to near-zero biases with the introduction of vertically resolved ozone in the early 2000s.

Overall, reanalysis ozone products do not exhibit large discontinuities at 10 hPa. As expected, both MERRA-2 and ERA-Interim show extremely good agreement with SWOOSH during the period in which they assimilate Aura MLS ozone data. Biases in these reanalyses undergo a step change when they start assimilating ozone profiles from Aura MLS ozone. For example, MERRA-2 assimilates Aura MLS data from August 2004 (**Figure 4.2**), and at that time biases in 10 hPa ozone relative to SWOOSH drop suddenly to less than 5 % at all latitudes. This reduction is also apparent in ERA-Interim, which assimilates Aura MLS ozone data during 2008 and then from June 2009 through the present, and also in ERA5, which assimilates Aura MLS from 2004 onwards. Similar sudden reductions in ozone biases relative to SWOOSH are seen in ERA-Interim in both early 2008 and the latter half of 2009.

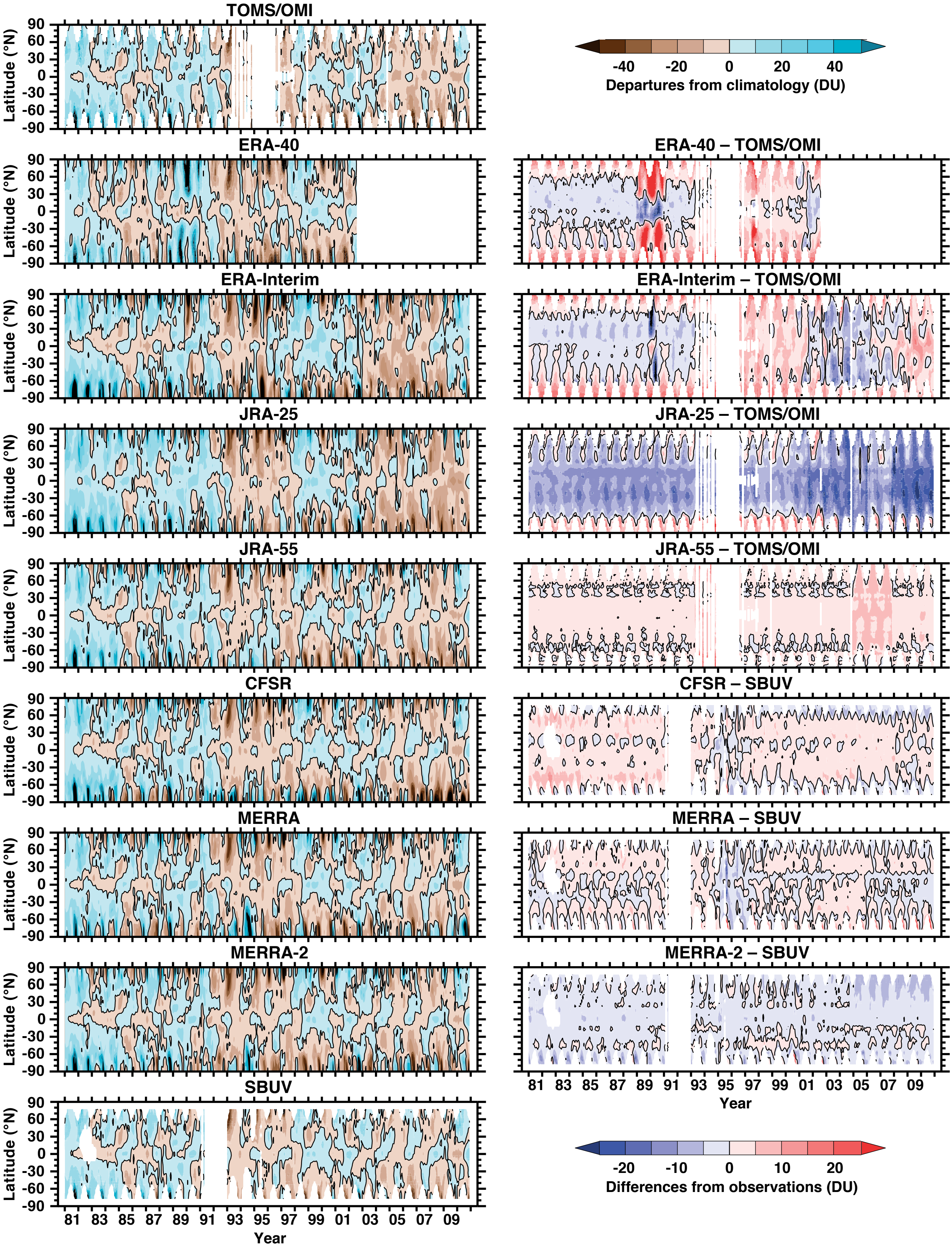


Figure 4.13: *Departures of TCO from the zonal- and monthly-mean 1981 - 2010 climatology for TOMS/OMI (left column, top row), SBUV (left column, bottom row), and reanalyses (left column, other rows). (Right column) Differences between reanalyses zonal- and monthly-mean TCO and the primary TCO observations that they assimilate. The black contour is at 0 DU. Reproduced from Davis et al. (2017).*

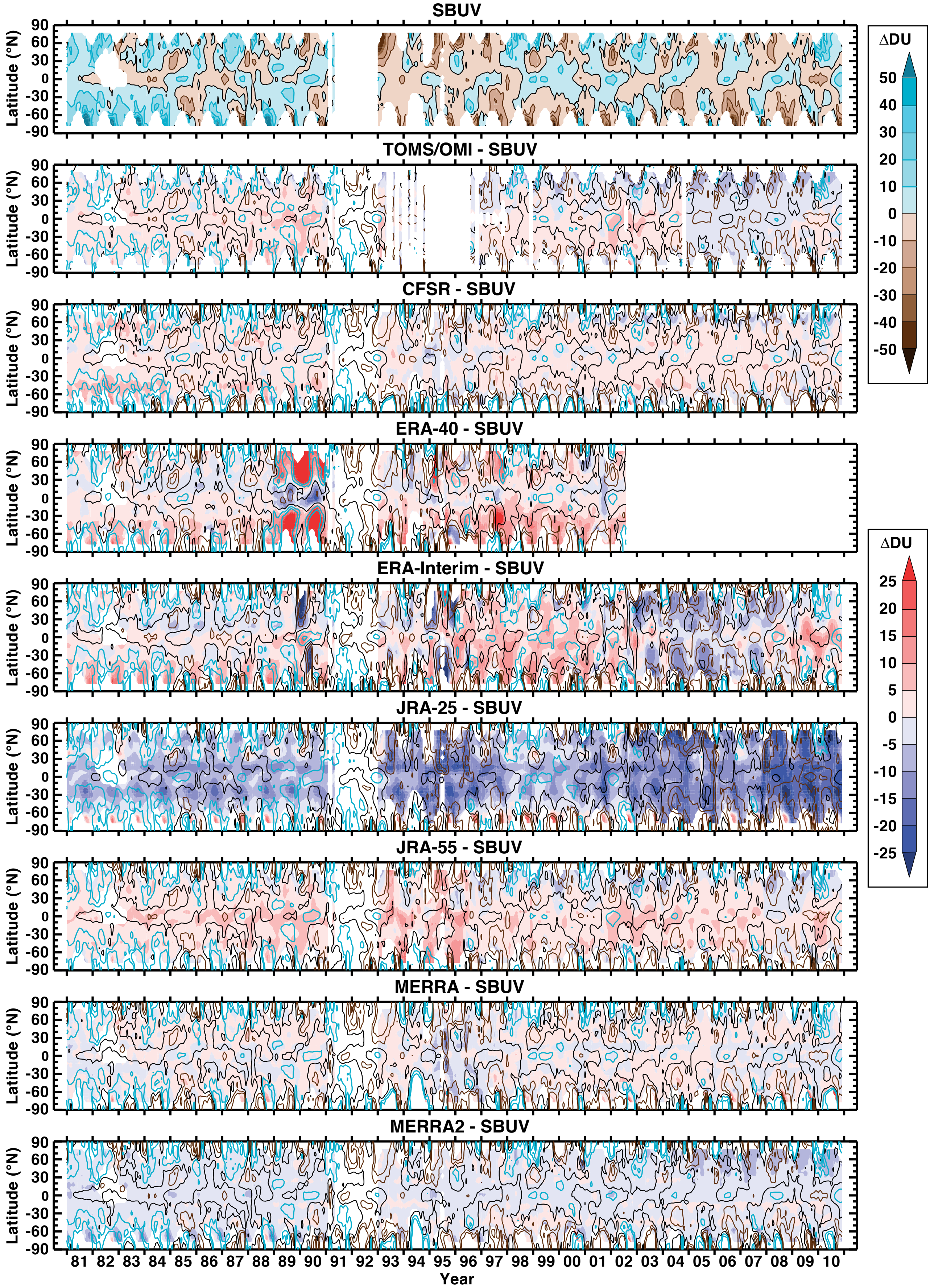


Figure 4.14: *TCO latitude vs. time anomalies for SBUV (top row), differences between TOMS/OMI and SBUV (second from top), and differences between SBUV and the reanalyses (other rows). The black contour is at 0 DU for the SBUV anomaly (top panel) or the dataset being differenced from SBUV (other panels). Anomalies of each dataset being compared to SBUV are contoured cyan (brown) at the 10 (-10) and 20 (-20) DU levels. Reproduced from Davis et al. (2017).*

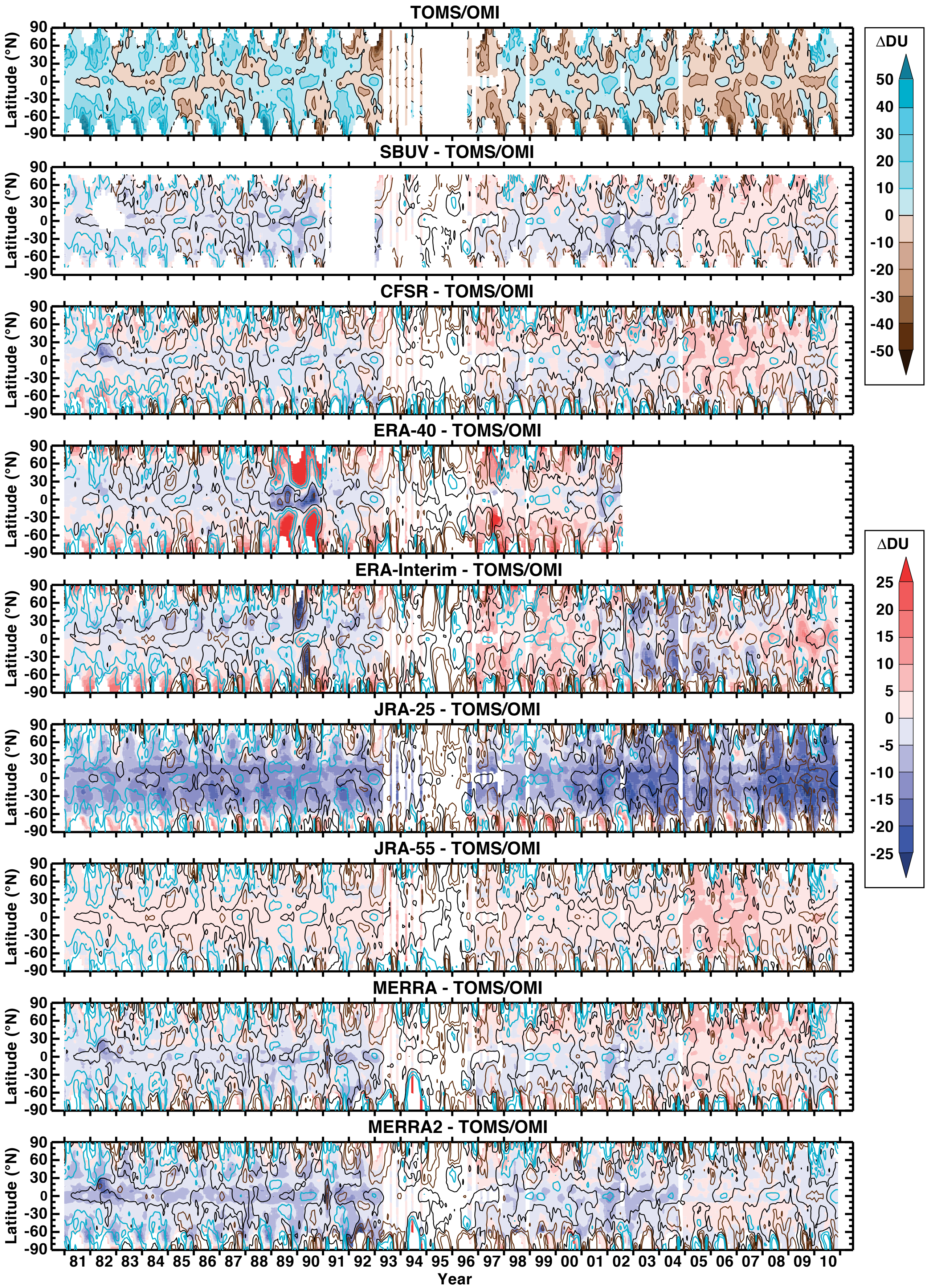


Figure 4.15: *TCO latitude vs. time anomalies for TOMS/OMI (top row), differences between SBUV and TOMS/OMI (second from top), and differences between TOMS/OMI and the reanalyses (other rows). The black contour is at 0 DU for the TOMS/OMI anomaly (top panel) or the dataset being differenced from TOMS/OMI (other panels). Anomalies of each dataset being compared to SBUV are contoured cyan (brown) at the 10 (-10) and 20 (-20) DU levels. Reproduced from Davis et al. (2017).*

Differences between reanalysis ozone fields and SWOOSH are larger at 70 hPa. A strong discontinuity in the MERRA-2 time series occurs in mid-2004 when it begins to assimilate Aura MLS ozone data. The same discontinuity is found in ERA5, although the positive bias is somewhat less pronounced in ERA5 than in MERRA-2. To a lesser extent there is also a discontinuity (in 2008 and again in mid-2009) when ERA-Interim begins assimilating Aura MLS ozone data. The large positive bias in MERRA-2 that starts in mid-2004 is also seen in comparisons to (non-assimilated) ozonesondes (*Wargan et al.*, 2017). This positive bias is related to vertical averaging of the MLS data before assimilation by MERRA-2 (*Wargan et al.*, 2017).

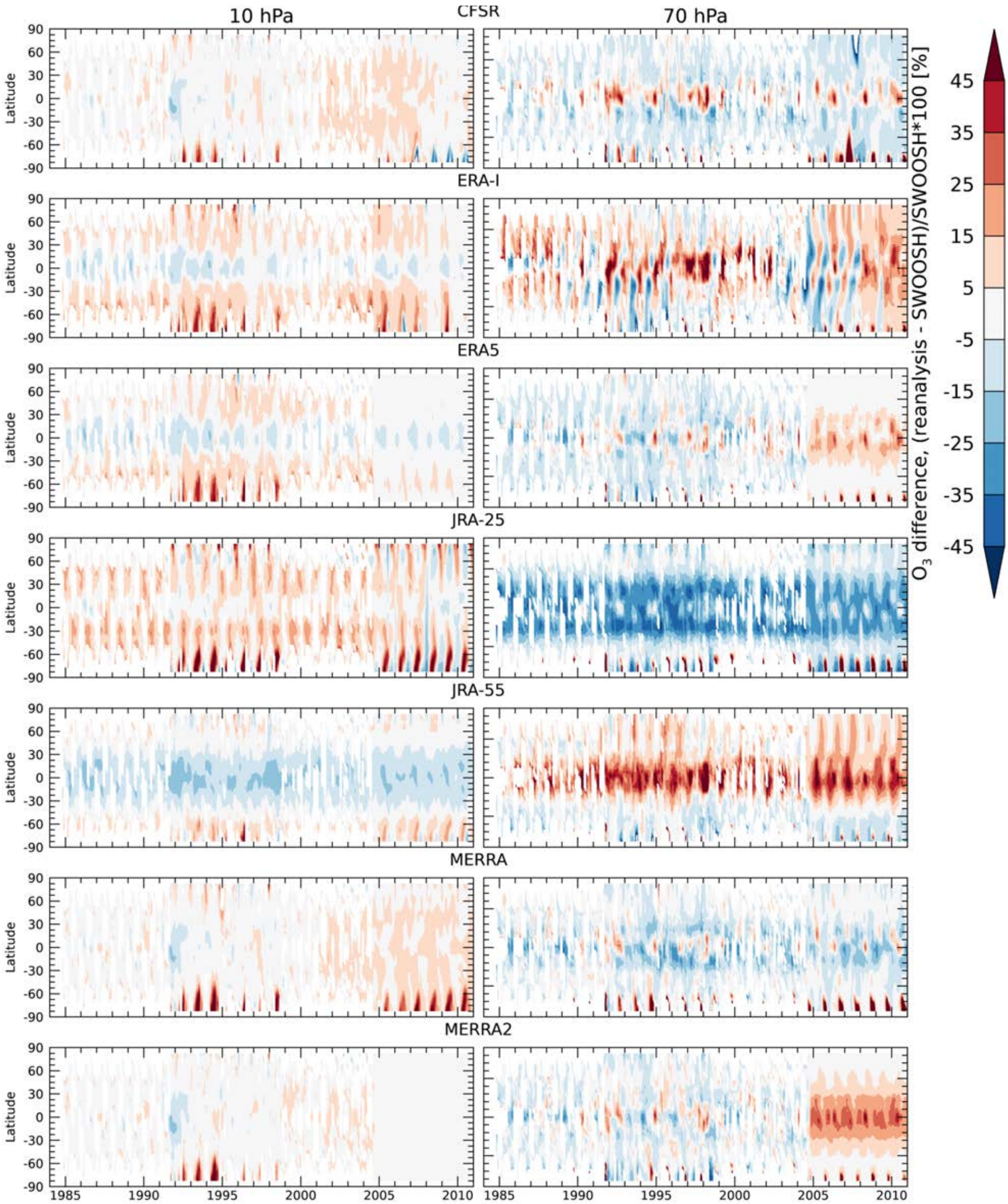


Figure 4.16: *Latitude–time evolution of relative differences between ozone reanalyses and the merged SWOOSH ozone record at 10 hPa and 70 hPa. White indicates missing data, and light grey indicates near-zero differences (e.g., between MERRA2 and SWOOSH after mid-2004). Updated from Davis et al. (2017).*

For the other reanalyses that don't assimilate MLS, there are generally not strong discontinuities that can be tied to observing system changes. There does seem to be a change in the ERA-Interim differences at the beginning of 2003 when it begins to assimilate vertically resolved data from MIPAS and TCO from SCIAMACHY. Beyond the discontinuities discussed above, at 70 hPa differences between the reanalysis ozone fields and SWOOSH are relatively consistent in time, with negative biases prevailing in JRA-25, CSFR, MERRA, and MERRA-2 (pre-Aura MLS), patchy biases in ERA-Interim, and mostly positive biases in JRA-25 and JRA-55 (especially in the tropics).

4.5 Evaluation of reanalysis water vapour products

In this section, we evaluate reanalysis estimates of water vapour in and above the tropopause layer against available observations. In keeping with the S-RIP remit, this section focuses exclusively on evaluations of reanalysis water vapour products in the upper troposphere and stratosphere.

4.5.1 Zonal mean water vapour cross-sections

Figure 4.17 shows multi-annual zonal mean water vapour for 2005 - 2010 from the SDI MIM along with relative differences between each reanalysis and the MIM (calculated as $100{\cdot}(R_i - MIM)/MIM$, where R_i is a reanalysis field). In contrast to ozone, the reanalyses do not consistently capture the zonal mean vertical distribution of water vapour. The pressure-level products provided by JRA-25 and JRA-55 do not include analysed stratospheric water vapour fields, while CFSR produces a stratosphere that is much too dry (negative biases exceeding 60 %). ERA-Interim, ERA5, MERRA, and MERRA-2 show water vapour fields that are close to observations. These three systems resolve the distinct minimum in water vapour mixing ratios just above the tropical tropopause, the second minimum in the lower stratosphere at SH high latitudes, and the increase in water vapour with increasing altitude. The slight negative bias found in ERA-Interim and ERA5 compared to the observations around the stratopause may be a result of relaxing the water vapour towards the UARS climatological value of 6.8 ppmv in this region (see *Section 4.2.3*), which is somewhat lower than the values observed in newer climatologies (*Hegglin et al.*, 2013). In contrast to the other reanalyses, MERRA and MERRA-2 extend up to the lower mesosphere (not shown), and hence capture the water vapour maximum found in the upper stratosphere (*e.g.*, *Hegglin et al.*, 2013).

CFSR is much too dry throughout the stratosphere and does not capture the typical structure of water vapour isopleths. This bias is due in part to the lack of assimilated observations to constrain the water vapour reanalyses at these altitudes and in part to the absence of a methane oxidation parameterization in the forecast model (*Section 4.2.3*). All reanalyses contain positive biases relative to the SDI MIM at pressures greater than 100 hPa (see also *Jiang et al.*, 2015), although this may in part be explained by the increase in measurement uncertainty of satellite limb sounders with decreasing altitude in the upper troposphere (*Hegglin et al.*, 2013). Several studies have shown that Aura MLS contains a dry bias in the upper troposphere/lower stratosphere around 200 hPa (*e.g.*, *Davis et al.*, 2016; *Vömel et al.*, 2007), and similarly a dry bias has been found in the upper troposphere for ACE-FTS (*Hegglin et al.*, 2008). Note, ERA5 shows the best agreement with observations below 100 hPa, with somewhat lower positive biases than the rest of the reanalyses.

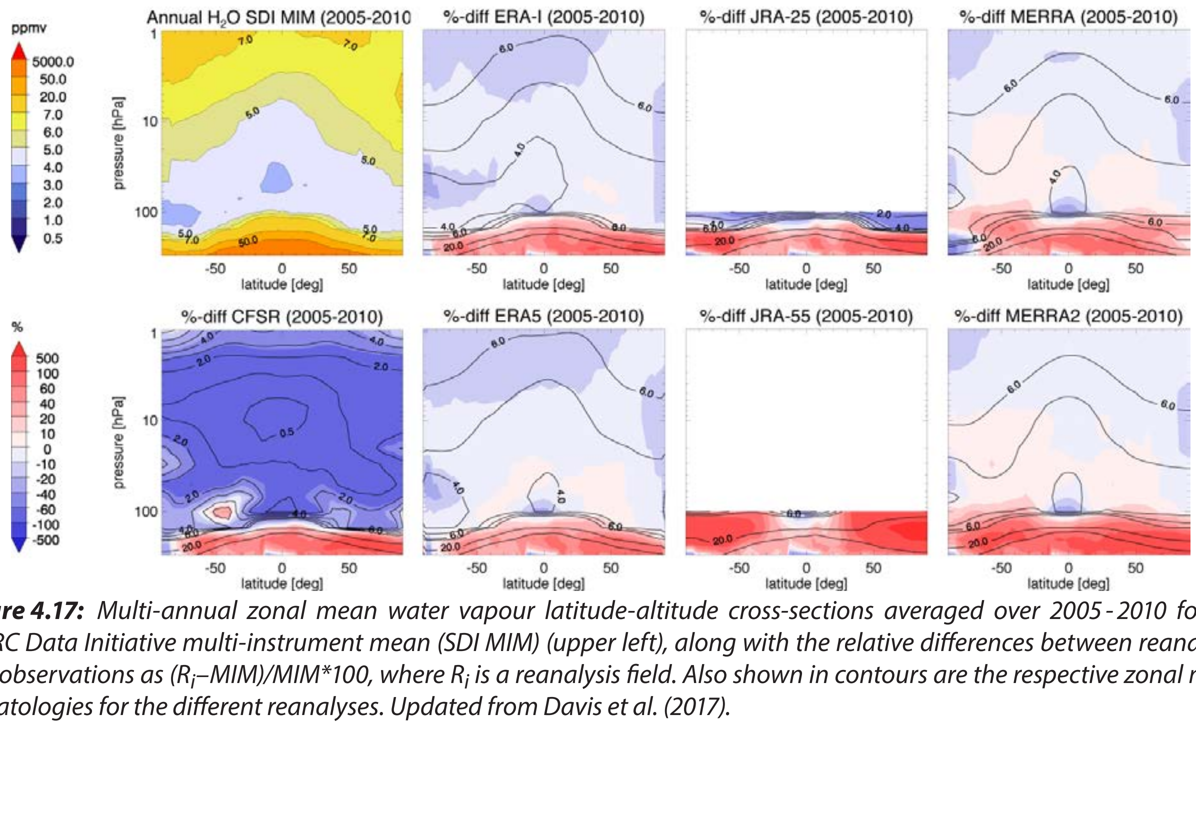


Figure 4.17: *Multi-annual zonal mean water vapour latitude-altitude cross-sections averaged over 2005 - 2010 for the SPARC Data Initiative multi-instrument mean (SDI MIM) (upper left), along with the relative differences between reanalyses and observations as (R_i–MIM)/MIM*100, where R_i is a reanalysis field. Also shown in contours are the respective zonal mean climatologies for the different reanalyses. Updated from Davis et al. (2017).*

4.5.2 Water vapour monthly mean vertical profiles and seasonal cycles

Figures 4.18a and **b** show vertical profiles of water vapour for January (2005 - 2010 average) for the reanalyses and the SDI MIM at two different latitudes 40°N and 70°S, respectively, along with the relative differences for each reanalysis with respect to the MIM. Figures 18c and d show the seasonal cycles of water vapour for three different pressure levels at 40°N and 70°S, respectively. In general, the results shown reinforce the conclusions of the previous section.

The comparisons in **Figures 4.18a** and **b** reveal very good agreement (within ± 10 %) between ERA-Interim, ERA5, MERRA, MERRA-2, and the observations at altitudes above 100 hPa. The 100 hPa level is one of the most important levels for stratospheric water vapour studies, because it is near the level where stratospheric water vapour entry mixing ratios are set in the tropics (*Fueglistaler et al.*, 2009) and because it is near the peak region of the radiative kernel for water vapour in the extratropics (*Gettelman et al.*, 2011). As mentioned in the previous section, water vapour from CSFR is unrealistic in the stratosphere, with values much lower than those observed. The reanalyses show large inconsistencies between their absolute values at altitudes below 100 hPa, leading to sharp increases in their relative differences with respect to the MIM of > 100 %. These relative differences are systematically positive except for in CFSR and JRA-25, pointing towards potential negative biases in the water vapour observations at these altitudes (*e.g.*, *Hegglin et al.*, 2013). The results may also indicate that the reanalyses produce an excessively moist tropical upper troposphere and/or excessive mixing of moist tropospheric air into the extratropical lowermost stratosphere.

The agreement between the reanalyses and observations varies by month, as shown in **Figures 4.18 c-e** and **f-h** for selected pressure levels (250, 100, and 50 hPa) and latitude bands (30°N-50°N and 60°S-80°S). At NH mid-latitudes (30°N-50°N; **Figure 4.18c**) at 250 hPa, all reanalyses are positively biased relative to the observations by more than 100 %, lending further support to the results by *Jiang et al.* (2015), who compared the reanalyses to Aura MLS alone, which is known to have a negative bias around this altitude (*Davis et al.*, 2016; *Hegglin et al.*, 2013; *Vömel et al.*, 2007). JRA-25 and JRA-55 have the smallest positive biases relative to observations at 250 hPa. At 100 hPa and 50 hPa, ERA-Interim, ERA5, MERRA, and MERRA-2 perform best, with approximately correct mean values, but somewhat underestimated seasonal cycle amplitudes. As noted earlier, a significant portion of the agreement in MERRA and MERRA-2 results from the relaxation of stratospheric water vapour towards a climatology that is based in part on Aura MLS data (which are also included in the SDI MIM). The results of ERA-Interim and ERA5 point towards the physical consistency in the parameterisations used to determine this prognostic variable in their reanalyses systems. JRA-55 (JRA-25) has mean values that are much too large (small) at 100 hPa. In addition to being too dry at 100 hPa and 50 hPa, CSFR also has incorrect amplitude and phase of the seasonal cycle at these levels.

At SH high latitudes (60°S-80°S; **Figure 4.18f-h**), all reanalyses show approximately the right phase, but overestimate mean values and amplitudes at 250 hPa, similar to the results at NH mid-latitudes. At 100 hPa and 50 hPa, ERA-Interim and ERA5 capture the phase and amplitude of the observed seasonal cycle best when compared to the other reanalyses, but exhibit a slight negative bias at 50 hPa. MERRA and MERRA-2 also show quite good agreement in terms of mean value, amplitude, and phase at 100 hPa, but overestimate mean values at 50 hPa, and also show a slight shift in the phase of the seasonal cycle with somewhat early minimum followed by an increase in September that occurs about a month earlier than observations. JRA-25 somewhat underestimates the mean value, but shows a similar phase and amplitude as the observations at 100 hPa. JRA-55 on the other hand, strongly overestimates the amplitude of the seasonal cycle at this level with mean values that are much too high. This JRA-55 positive bias at 100 hPa in the extratropics (of both hemispheres) is due to unrealistically large values in its forecast model stratosphere that are unconstrained by observations and impact the 100 hPa level. CSFR shows too low values at both 100 hPa and 50 hPa, but captures the seasonality somewhat better than it does in the NH mid-latitudes.

4.5.3 Water vapour interannual variability

Figure 4.19 shows time series of interannual variability in water vapour and its anomalies based on observations and reanalysis products during 2005 - 2010. At 250 hPa in NH midlatitudes (40°N - 60°N), the reanalyses show a much larger amplitude seasonal cycle, with much larger maxima during summer. Generally, the reanalyses follow the observed interannual variability extremely well, especially JRA-25, JRA-55, MERRA, and ERA5. CSFR seems to exhibit an underlying positive trend in its time series. And as noted previously, all reanalyses are wetter than observations at this level by approximately a factor of two.

At 100 hPa in the tropics (a level that is often used to estimate stratospheric water vapour entry mixing ratios), all reanalyses except CSFR and JRA-25 compare reasonably well with the observed seasonal cycle and anomalies. Perhaps surprisingly, JRA-25 captures the interannual anomalies quite well despite being biased negative in its mean value and seasonal cycle amplitude. CSFR shows no clear interannual variability and produces water vapour mean values as low as 0 ppmv. CSFR begins to produce more realistic water vapour concentrations at these levels in 2010, but with values that are larger than those in the observations and other reanalyses.

This change is discussed further in *Section 4.5.4*. Note that the SDI MIM for this level only includes Aura MLS and ACE-FTS due to known problems in SCIAMACHY and MIPAS data in this region (Hegglin et al., 2013).

At 50 hPa in the SH high latitudes (60°S-80°S), MERRA and MERRA-2 have roughly correct water vapour mean values, whereas ERA-Interim and ERA5 are slightly too low and CFSR is essentially zero before 2010. MERRA and MERRA-2 both place the minimum during austral winter (from dehydration processes in the cold polar vortex) about one month too early. Except for CFSR, the other reanalyses capture the correct structure in the interannual variability, including the prominent positive anomaly in 2010. MERRA and MERRA-2 show less variability than observed, which is unsurprising given their strong relaxation to the climatology.

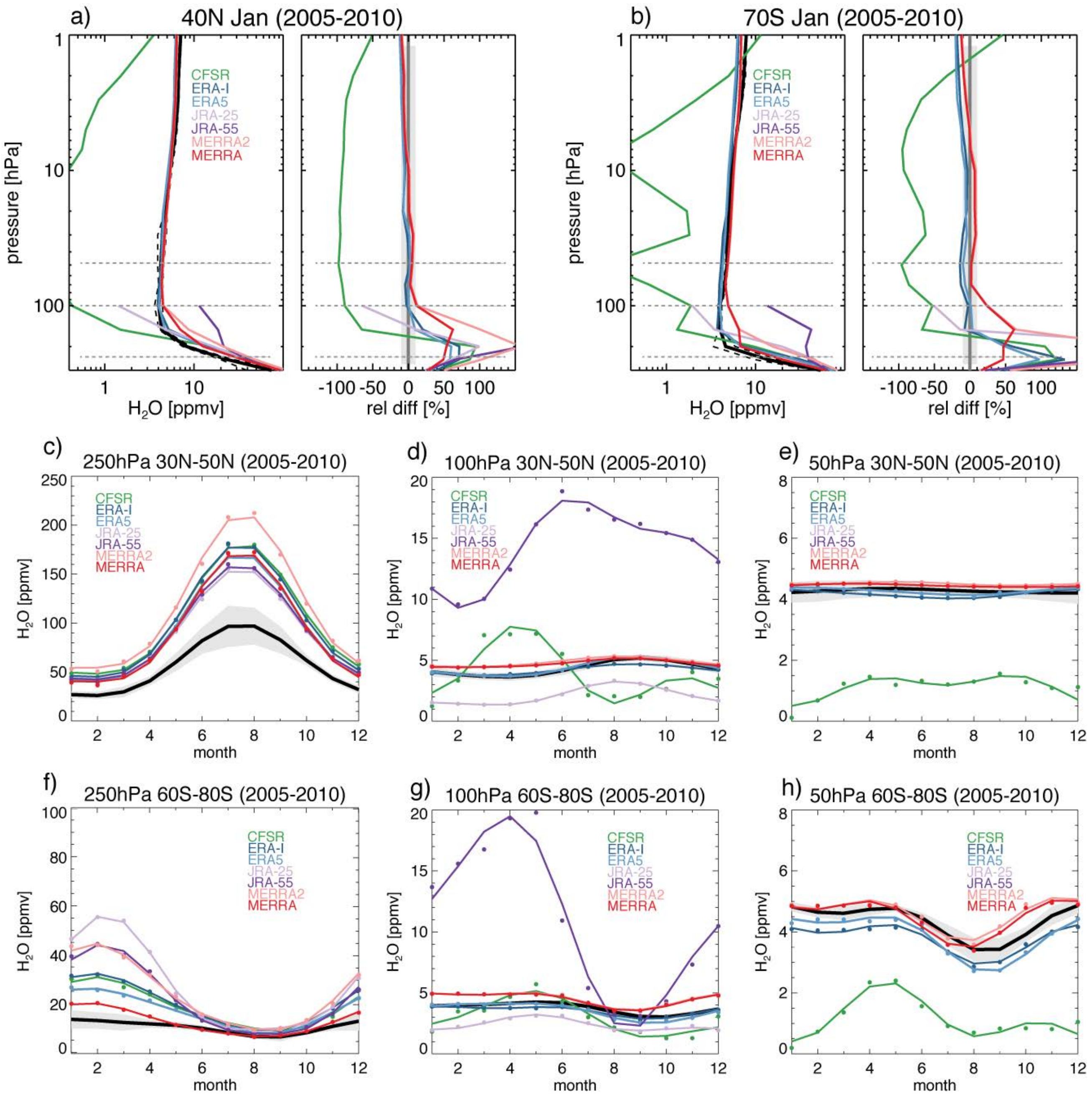


Figure 4.18: *Multi-annual mean vertical water vapour profiles over 2005-2010 for January at (a) 40° N and (b) 70° S from the SPARC Data Initiative multi-instrument mean (SDI MIM) (black) and the six reanalyses (coloured). Absolute values are shown in the left and relative differences in the right panels for each comparison. Relative differences are calculated as (R_i–MIM)/MIM*100, where R_i is a reanalysis profile. Black dashed lines provide the ±1-sigma uncertainty (as calculated by the standard deviation over all instruments and years available) in the observational mean. Horizontal dashed lines in grey indicate the pressure levels (250, 100, and 50 hPa) for which seasonal cycles are shown in panels (c) - (h) for the two latitude ranges 30°-50° N and 60°-80° S. Grey shading indicates observational uncertainty (±1-sigma) calculated as the standard deviation over all instruments and years available. Updated from Davis et al. (2017).*

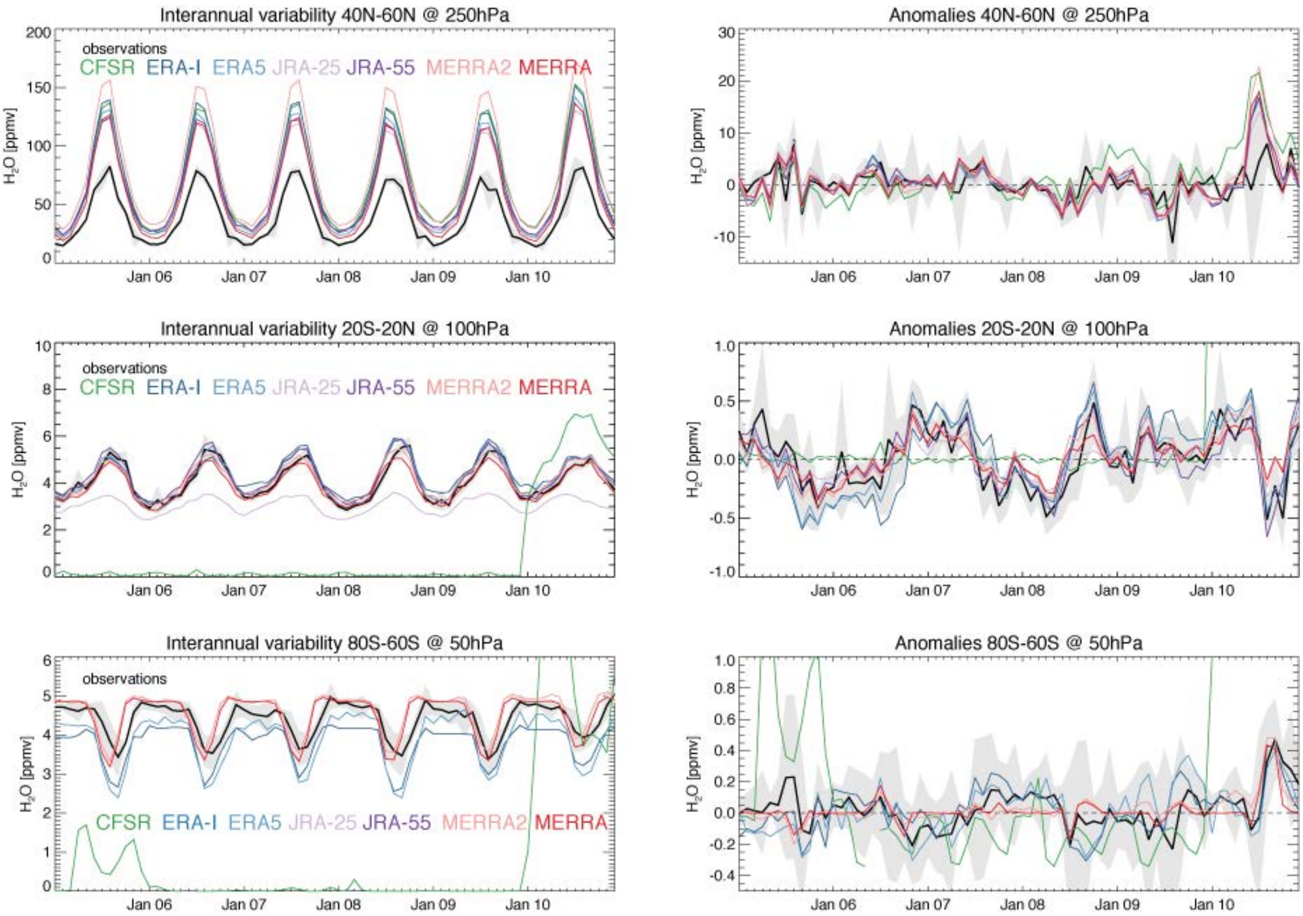


Figure 4.19: *Interannual variability (left column) and deseasonalized anomalies (right column) for water vapour during 2005 - 2010 for the SPARC Data Initiative multi-instrument mean (SDI MIM, black) and the six reanalyses (coloured). Results are shown for three different pressure levels and latitude ranges (bottom to top: 50 hPa at 60 - 80 °S, 100 hPa at 20 °S - 20 °N, and 250 hPa at 40 - 60 °N). Grey shading indicates observational uncertainty (± 1-sigma) calculated as the standard deviation over all instruments and years available. Updated from Davis et al. (2017).*

4.5.4 Tropical tape recorder in water vapour

Representations of the tropical tape recorder (*Mote et al.,* 1996) provide an additional illustration of problems in reanalysis stratospheric water vapour products. **Figure 4.20** shows the time–height evolution of water vapour in reanalyses and the merged SWOOSH observations averaged over the 15 °S - 15 °N tropical band. Anomalies are calculated separately for each data set, relative to the mean seasonal cycle at each level for the period 1992 - 2014 (except ERA-40, which is 1992-2002), when all reanalyses (except ERA-40) overlap. Variations in these fields reflect changes in the mixing ratio of water vapour entering the tropical lower stratosphere, as driven by variations in tropical tropopause temperatures and the subsequent vertical propagation in the ascending branch of the stratospheric overturning circulation. Interannual variability in both water vapour entry mixing ratios and ascent rate (the vertical slope of the signal) is superimposed on this mean seasonal cycle. Although reanalyses do not reproduce observed water vapour concentrations in the stratosphere, most reanalyses do produce a tropical tape recorder signal.

As previously discussed, CFSR (**Figure 4.20a**) produces water vapour concentrations near zero in the stratosphere for most of the record, although unrealistically wet values appear above 20 hPa at certain times (*e.g.;* 1995 and 1999). These upper stratospheric wet anomalies (and several others that occurred before 1992) all correspond to transitions in the main CFSR production stream (see **Figure 2.2** and **Figure 2** of *Fujiwara et al., 2017*). We hypothesize that these wet anomalies are a remnant of a wet bias in the model initialisation that remains after the ~1-year spinup. Additional step changes in water vapour are evident at the beginning of 2010 and at the beginning of 2011. The latter step change corresponds to the transition from CFSR (CDAS-T382) to CFSv2 (CDAS-T574) at the beginning of 2011. As discussed in *Section 4.2.2*, CFSv2 is intended as a continuation of CFSR but has differences in model resolution and physics relative to the original system. Although the reasons for the step change at the beginning of 2010 are not known definitively, we note that CFSR was extended for the year 2010 following its original completion over the 1979 - 2009 time period. This extension used the original CDAS-T382 system but with some slight changes to the forecast model.

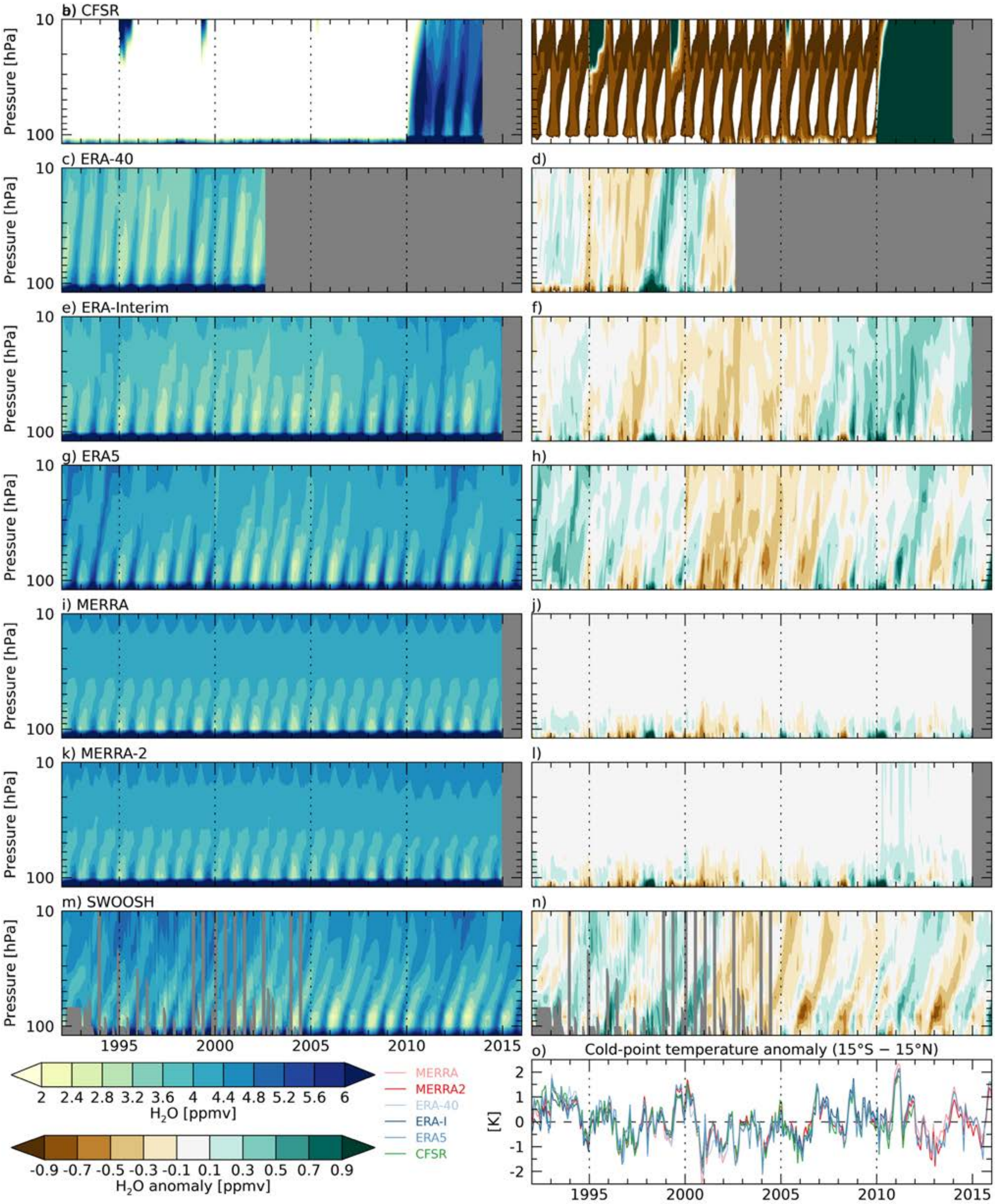


Figure 4.20: *The tropical tape recorder signal as represented in reanalyses and the SWOOSH merged satellite product, defined as the height–time evolution of water vapour averaged over the 15° S - 15° N tropical band. Both absolute values (left column) and anomalies relative to the mean water vapour seasonal cycle at each level (right column) are shown. Anomalies are computed separately for each data set. Monthly mean anomalies in tropical (15° S - 15° N) cold-point tropopause temperatures calculated from 6-h data on the native vertical resolution of each reanalysis model are shown for context (o). Regions with no data are gray, and off-scale data are white. Updated from Davis et al. (2017).*

It is likely that the CFSR 2010 run was performed without a sufficiently long spin-up period, or that a change to the model configuration resulted in the observed water vapour discontinuity beginning in 2010.

ERA-40 and ERA-Interim (**Figure 4.20c, e**) are generally drier than the SWOOSH observations (**Figure 4.20m**), although the ERA-Interim represents an evident improvement over ERA-40 in this respect. ERA-5 (Figure 4.20g) agrees best with SWOOSH, both in magnitude and variability. Both MERRA and MERRA-2 (**Figure 4.20i,k**) are close in magnitude to SWOOSH, but this agreement is expected given that both systems relax stratospheric water vapour to a climatology based on Aura MLS and HALOE (*Sections 4.2.7, 4.2.8*).

The reanalyses all produce tape recorder slopes that are more steeper than suggested by the observations, indicating that vertical upwelling in the tropical stratosphere is too strong in reanalyses. Although biases and differences in tropical stratospheric upwelling have been addressed quantitatively for a subset of reanalyses elsewhere (*Abalos et al.*, 2015; *Jiang et al.*, 2015), the SWOOSH data shown in **Figure 4.20** enable a comparison that extends beyond the Aura MLS record. This extension allows for comparison to ERA-40, and shows that ERA-Interim benefits from a much-improved representation of stratospheric water vapour and its variability relative to its predecessor.

Figure 4.20 also shows interannual variability in tropical stratospheric water vapour as represented by the anomaly from the mean seasonal cycle at each level. Interannual variability in the tape recorder signal is related to interannual variability in cold-point tropopause temperatures (**Fiugre 4.20o**), with warm anomalies at the tropopause corresponding to wet anomalies in the tape recorder and vice versa. Although the reanalyses produce almost identical interannual variations in tropical tropopause temperatures over the period considered here, their interannual variations in stratospheric water vapour differ substantially. The strong relaxation to climatology applied in MERRA and MERRA-2 results in very little interannual variability above 60 hPa because of the short nudging timescale for WV (3 days). ERA-40 produces a very large wet anomaly during the 1997 - 1998 El Niño that coherently propagates upwards. This anomaly is wetter than that suggested by SWOOSH and the other reanalyses. SWOOSH and the reanalyses all show a wet anomaly near 100 hPa in the tropics during the 1997–1998 El Niño, but this anomaly does not correspond to a strong warm excursion in cold-point temperature.

Randel et al. (2006) reported the occurrence of a sudden drop in stratospheric water vapour that persisted for ~5 years during the early 2000s. This drop is evident in the cold-point temperature and SWOOSH water vapour anomalies (**Figure 4.20n,o**). The reanalyses generally capture the drop in stratospheric WV around 2000, with the caveat that the relaxation to a monthly mean climatology in MERRA and MERRA-2 damps the associated signals above the lowermost stratosphere.

4.6 Summary

In this chapter, we described the basic treatment of ozone and water vapour in reanalyses, and presented comparisons both among reanalyses and between reanalyses and observations (both assimilated and independent). Here we briefly summarize the most influential characteristics and differences in the treatment of ozone and water vapour in reanalyses along with the key results of the intercomparisons.

The treatment of ozone and water vapour varies substantially among reanalyses. Some reanalyses prescribe ozone climatologies and do not treat ozone prognostically (*R1*, *R2*), some reanalyses specify ozone as a boundary condition generated by an offline chemical transport model (JRA-25, JRA-55), and some reanalyses treat ozone as a prognostic variable with parameterised photochemical production and loss (CFSR, ERA-40, ERA-Interim, ERA5, MERRA, and MERRA-2). Only ERA-40, ERA-Interim, and ERA5 contain a parameterization of heterogeneous ozone loss processes.

The reanalyses also assimilate different sets of ozone observations, with generally similar observation usage for reanalyses produced by the same reanalysis centre. All reanalyses that assimilate ozone observations rely heavily on total column ozone observations from some combination of satellites carrying the TOMS and SBUV sensors. Several recent reanalyses (including MERRA-2 and ERA-Interim) use the newest generation of vertically resolved ozone measurements (*e.g.*, Aura MLS).

Reanalyses all assimilate tropospheric humidity information via some combination of radiosondes, satellite radiances, GNSS-RO bending angles, and retrievals of atmospheric hydrological quantities (*e.g.*, total column water vapour or rain rate). None of the reanalyses assimilate WV observations in the stratosphere, although information from tropospheric observations may propagate upward in some systems. Beyond these similarities, the treatment of stratospheric water vapour varies substantially among the reanalyses. For example, the specific cut-off altitude up to which radiosonde humidity data are assimilated varies from one reanalysis to another, using either a fixed pressure level or the diagnosed tropopause. ERA-40, ERA-Interim, and ERA5 are the only reanalyses that include a water vapour source from methane oxidation. MERRA and MERRA-2 relax their fields to a water vapour climatology based on satellite observations (*e.g.*, including Aura MLS), while other reanalyses simply do not provide valid data in the stratosphere (*e.g.*, CSFR, JRA-25, JRA-55, R1, R2). These latter reanalyses prescribe a climatology or constant value for stratospheric water vapour as input to the forecast model radiative transfer code.

Given these differences amongst reanalysis treatments of ozone and WV, it is perhaps unsurprising that comparisons between reanalyses and observations also vary widely. Comparisons against assimilated observations of total column ozone (TCO) show that reanalyses generally reproduce TCO well, within ~10 DU (~3%).

Key limitations that result in larger errors and uncertainties include a general lack of TCO data during polar night and the absence of heterogeneous chemistry from most reanalysis ozone schemes (except in ERA-40 and ERA-interim where it is introduced as a simple parameterization activated when the local temperature falls below 195 K). The vertical distributions of stratospheric ozone and WV in reanalyses are unconstrained by observations through most of the record, owing to vertically-resolved data generally not being used in the assimilation systems. The situation for ozone is slightly better than that for WV, because stratospheric ozone observations are assimilated and because the ozone parameterizations are more advanced. Nevertheless, the current parameterisations for stratospheric water vapour implemented in ERA5 show a high level of performance and stark improvements over the water vapour distributions of earlier ECMWF reanalyses.

From the middle to upper stratosphere, reanalysis ozone profiles are within ±20 % of MIM of observations from the SPARC Data Initiative, although the comparisons are not truly independent for MERRA-2, ERA-Interim, and ERA5 because they assimilate data from Aura MLS, one of the instruments that contribute to the SPARC Data Initiative dataset. In the upper troposphere and lower stratosphere, biases increase to ±50 % for ozone.

MERRA-2 and ERA5 perform particularly well for ozone through much of the stratosphere. This is mainly due to the assimilation of the vertically resolved Aura MLS observations, which have helped to address difficulties in reproducing vertical distributions of ozone, particularly during polar night; however, these data are only available since late 2004 and are only assimilated by a few reanalyses. The use of reanalysis ozone for Antarctic ozone hole studies is therefore problematic. The reanalyses produce reasonable ozone holes when observations are available, but the timing and area of reanalysis ozone holes is positively biased when observations are (unavailable or) not assimilated. Also, apart from JRA-55, most reanalyses seem to exhibit a drift in the extent of the ozone hole area when compared to TOMS/OMI observations.

More generally, studies utilizing reanalysis ozone fields to analyse longer-term variations (*e.g.*, trends) should exercise extreme caution. Vertically resolved observations are not available over the entire time period of reanalyses, and the few reanalyses that use these temporally limited observations have significant discontinuities when the assimilation of vertically-resolved observations begins or stops. And while the reanalyses generally do a good job reproducing the TCO observations they assimilate, there remain potentially significant discontinuities associated with both the transition between satellite instruments and the underlying data. Assimilation of vertically-resolved ozone measurements, assimilation of measurements in polar night, and improved chemical parameterization of ozone processes should be pursued by reanalysis centers in order to improve the representation of ozone fields in reanalyses into the future.

None of the reanalyses assimilate observations of stratospheric water vapour, resulting in large differences between reanalyses and independent observations. CFSR has an extreme dry bias in the stratosphere through 2009, with monthly mean values often approaching 0 ppmv.

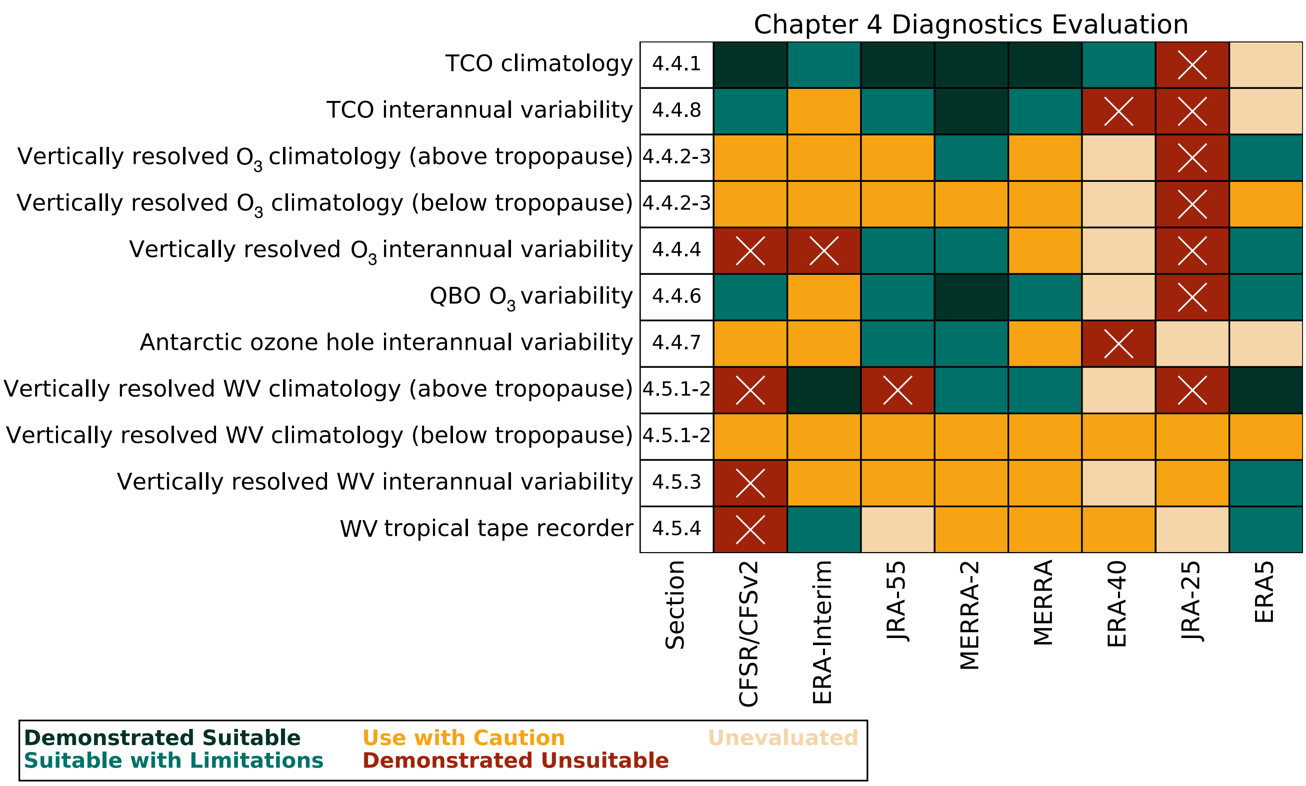


Figure 4.21: *A summary of the diagnostics evaluated in this chapter.*

Although MERRA and MERRA-2 produce reasonable values for stratospheric water vapour, these values represent a strong relaxation to a fixed annual climatology at pressures less than 50 hPa. Hence, mid- and upper-stratospheric water vapour does not undergo physically meaningful variations in MERRA or MERRA-2. ERA-40, ERA-Interim, and ERA5 produce a true "prognostic" water vapour field in the stratosphere. ERA-Interim and ERA5 produce surprisingly reasonable values given that their fields are predominantly controlled by dehydration in the TTL and a very simple parameterization of methane oxidation. In the upper troposphere and lower stratosphere, reanalyses are around a factor of two wetter than the SPARC Data Initiative WV measurements used here, although the observations also have relatively large disagreements in this region. Notably, ERA5, possibly due to further refinements in its parameterization for dehydration since ERA-Interim, shows reduced biases in water vapour in this altitude region.

Because of the lack of assimilated observations and the deficiencies in representation of the relevant physical processes, we recommend that reanalysis stratospheric water vapour fields should generally not be used for scientific data analysis, and stress that any examination of these fields must account for their inherent limitations and uncertainties. However, ERA5 water vapour distributions show promising results and further evaluations should be performed to judge the final quality of this reanalysis. Future efforts toward the collection and assimilation of observational data with sensitivity to stratospheric water vapour, the reduction of reanalysis temperature biases in the TTL, and improvements in the representation of processes that control the entry mixing ratios or subsequent evolution of water vapour in the stratosphere could facilitate more reliable stratospheric water vapour fields in reanalyses.

4.7 Key findings and recommendations

A summary of the diagnostics evaluated in this chapter is provided in **Figure 4.21**. This figure contains assessments of the reanalysis representations of key diagnostics related to water vapour and ozone, and directs the reader towards the appropriate chapter section for further information. The assessments, while inherently subjective, are intended to provide the reader with an overview of the relative quality of the diagnostics. So, for example, across a given diagnostic the relative performance of the different reanalyses can be compared, and for a given reanalysis the performance across different diagnostics can be compared.

Below, we briefly summarize the key findings from this chapter and recommendations for both use of and improvements to reanalysis ozone and water vapour fields.

Key Findings:

- The treatment of ozone and water vapour varies substantially among reanalyses, both in terms of their representation of these species and assimilated observations.

- The latest generation of reanalyses all assimilate satellite total column ozone (TCO) observations, with some including vertically-resolved measurements.

- Currently none of the reanalyses directly assimilate water vapour observations in the stratosphere, although they do assimilate temperature and tropospheric humidity observations that can impact their stratospheric water vapour concentrations.

- Comparisons against assimilated observations of TCO show that reanalyses generally reproduce TCO well in sunlit regions, within ~ 10 DU (~ 3 %).

- The lack of TCO observations in polar night, and lack of representation of heterogeneous chemistry in most reanalyses, leads to relatively larger errors in representing TCO in the Antarctic ozone hole.

- From the middle to upper stratosphere, climatological reanalysis ozone profiles are within ± 20 % of observations.

- Biases are generally larger (~ 50 %) for both water vapour and ozone in the upper troposphere and lower stratosphere.

- Significant discontinuities exist in reanalysis water vapour and ozone time series due to transitions in the observing system.

Recommendations:

- Users should generally use caution when using reanalysis ozone fields for scientific studies and should check that their results are not reanalysis-dependent.
- Reanalysis stratospheric water vapour fields should generally not be used for scientific data analysis (except perhaps for ERA5). Any examination of these fields must account for their inherent limitations and uncertainties.
- In order to improve reanalysis ozone fields, reanalysis centres should work towards improved chemical parameterisations of ozone as well as assimilation of vertically-resolved ozone measurements (e.g., from limb sounders) and measurements in polar night (e.g., from IR nadir sounders).
- In order to improve reanalysis water vapour fields, future efforts should include the collection and assimilation of observational data with sensitivity to stratospheric water vapour, the reduction of reanalysis temperature biases in the TTL, and improvements in the representation of other processes that affect the stratospheric entry mixing ratio.

Code availability

Code for creating the common-grid data files and plots are available from the corresponding author upon request.

Data availability

The reanalysis data files necessary to create the "common grid" data files used here are available through the CREATE project website (**https://cds.nccs.nasa.gov/tools-services/create/**). Reanalysis total column ozone data was downloaded from the NCAR RDA (**https://rda.ucar.edu/**). SBUV data are available at **https://acd-ext.gsfc.nasa.gov/Data_services/merged/**. TOMS/OMI data are available at **https://disc.gsfc.nasa.gov/Aura/data-holdings/OMI/omto3d_v003.shtml.** SPARC Data Initiative data are available at **https://zenodo.org/record/4265393#.YKvAHeso_UI**. Aura MLS satellite data are available at **https://disc.sci.gsfc.nasa.gov/Aura/data-holdings/MLS**. SWOOSH data are available at **https://www.esrl.noaa.gov/csd/swoosh/**.

Acknowledgements

We thank Karen Rosenlof for valuable comments and suggestions on this manuscript. MF's contribution was financially supported in part by the Japan Society for the Promotion of Science (JSPS) through Grants-in-Aid for Scientific Research (26287117 and 16K05548). Work at the Jet Propulsion Laboratory, California Institute of Technology, was carried out under a contract with the National Aeronautics and Space Administration.

References

Abalos, M., B. Legras, F. Ploeger, and W.J. Randel, 2015: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979-2012. *J. Geophys. Res: Atmos*. **120**, 7534 - 7554, doi: 10.1002/2015JD023182.

Arguez, A., and R.S. Vose, 2011: The Definition of the Standard WMO Climate Normal The Key to Deriving Alternative Climate Normals. *Bull. Am. Meteorol. Soc.*, **92**, 699 - 704, doi: 10.1175/2010BAMS2955.1.

Bechtold, P., *et al.*, 2004: The simulation of the diurnal cycle of convective precipitation over land in a global model. *Q. J. Roy. Meteor. Soc.*, **130**, 3119 - 3137, doi: 10.1256/qj.03.103.

Bhartia, P.K., and C.W. Wellemeyer, 2002: TOMS-V8 total O3 algorithm. In: *OMI Algorithm Theoretical Basis Document: OMI Ozone Products*, edited by: Bhartia, P. K., ATBD-OMI-02, 15 - 32.

Bhartia, P.K., *et al.*, 2013: Solar Backscatter UV (SBUV) total ozone and profile algorithm. *Atmos. Meas. Tech.*, **6**, 2533 - 2548, doi: 10.5194/amt-6-2533-2013.

Bosilovich, M., *et al.*, 2015: MERRA-2: Initial Evaluation of the Climate. In: *Technical Report Series on Global Modeling and Data Assimilation*, edited by: Koster, R. D., *NASA/TM–2015-104606*, **43**.

Butchart, N., and E.E. Remsberg, 1986: The Area of the Stratospheric Polar Vortex as a Diagnostic for Tracer Transport on an Isentropic Surface. *J. Atmos. Sci.*, **43**, 1319 - 1339, doi: 10.1175/1520-0469.

Cariolle, D., and M. Deque, 1986: Southern-Hemisphere Medium-Scale Waves and Total Ozone Disturbances in a Spectral General-Circulation Model. *J. Geophys. Res: Atmos.*, **91**, 825 - 846, doi: 10.1029/Jd091id10p10825.

Cariolle, D., and H. Teyssédre, 2007: A revised linear ozone photochemistry parameterization for use in transport and general circulation models: multi-annual simulations. *Atm. Chem. Phys.*, **7**, 2183 - 2196,.

Chipperfield, M.P., L.J. Gray, J.S. Kinnersley, and J. Zawodny, 1994: A Two-Dimensional Model Study of the QBO Signal in SAGE II NO2 and O3. *Geophys. Res. Lett.*, **21**, 589 - 592, doi: 10.1029/94gl00211.

Coy, L., *et al.*, 2016: Structure and Dynamics of the Quasi-Biennial Oscillation in MERRA-2. *J. Clim.*, **29**, 5339 - 5354, doi: 10.1175/JCLI-D-15-0809.1.

Davis, S.M., *et al.*, 2016: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: A long-term database for climate studies. *Earth Syst. Sci. Data*, **8**, 461 - 490, doi: 10.5194/essd-8-461-2016.

Dee, D.P., and A.M. Da Silva, 2003: The choice of variable for atmospheric moisture analysis. *Mon. Weather. Rev.*, **131**, 155 - 171, doi: 10.1175/1520-0493(2003).

Dee, D., 2008: Importance of satellites for stratospheric data assimilation. Seminar on *Recent developments in the use of satellite observations in numerical weather prediction*, ECMWF, Reading, UK, 3 - 7 September 2007.

Dee, D.P., *et al.*, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. Roy. Meteor. Soc.*, **137**, 553 - 597, doi: 10.1002/qj.828.

Dessler, A.E., and S.M. Davis, 2010: Trends in tropospheric humidity from reanalysis systems. *J. Geophys. Res: Atmos.*, **115**, D19127, doi:10.1029/2010JD014192**.**

Dethof, A., and E.V. Hólm, 2004: Ozone assimilation in the ERA-40 reanalysis project. *Q. J. Roy. Meteor. Soc.*, **130**, 2851 - 2872, doi: 10.1256/qj.03.196.

Dragani, R., 2011: On the quality of the ERA-Interim ozone reanalyses: comparisons with satellite data. *Q. J. Roy. Meteor. Soc.*, **137**, 1312 - 1326, doi: 10.1002/qj.821.

Ebisuzaki, W., and L. Zhang, 2011: Assessing the performance of the CFSR by an ensemble of analyses. *Clim. Dynam.*, **37**, 2541 - 2550, doi: 10.1007/s00382-011-1074-5.

Flemming, J., *et al.*, 2017: The CAMS interim Reanalysis of Carbon Monoxide, Ozone and Aerosol for 2003-2015. *Atm. Chem. Phys.*, **17**, 1945 - 1983, doi: 10.5194/acp-17-1945-2017.

Flynn, L., 2007: Solar Backscatter Ultraviolet Instrument (SBUV/2) Version 8 Ozone Retrieval Algorithm Theoretical Basis Document (V8 ATBD), available at: **http://ozoneaq.gsfc.nasa.gov/media/docs/SBUV2_V8_ATBD_020207.pdf** (accessed: April 2021).

Flynn, L.E., *et al.*, 2009: Measurements and products from the Solar Backscatter Ultraviolet (SBUV/2) and Ozone Mapping and Profiler Suite (OMPS) instruments. *Int. J. Remote. Sens.*, **30**, 4259 - 4272, doi: 10.1080/01431160902825040.

Fortuin, J.P.F., and U. Langematz, 1995: An Update on the Global Ozone Climatology and on Concurrent Ozone and Temperature Trends. *P. Soc. Photo: Opt. Ins.*, **2311**, 207 - 216, 1995.

Frith, S.M., *et al.*, 2014: Recent changes in total column ozone based on the SBUV Version 8.6 Merged Ozone Data Set. *J. Geophys. Res: Atmos.*, **119**, 9735 - 9751, doi: 10.1002/2014JD021889.

Fueglistaler, S., *et al.*, 2009: Tropical Tropopause Layer. *Rev. Geophys.*, **47**, RG1004, doi: 10.1029/2008RG000267.

Fujiwara, M., *et al.*, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp-17-1417-2017.

Geer, A.J., *et al.*, 2006: The ASSET intercomparison of ozone analyses: method and first results. *Atmos. Chem. Phys.*, **6**, 5445 - 5474, doi: 10.5194/acp-6-5445-2006.

Gelaro, R., *et al.*, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.*, **30**, 5419 - 5454, doi: 10.1175/jcli-d-16-0758.1.

Gettelman, A., *et al.*, 2011: The extratropical upper troposphere and lower stratosphere. Rev. Geophys., 49, RG3003, doi: 10.1029/2011RG000355.

Hasebe, F., 1994: Quasi-Biennial Oscillations of Ozone and Diabatic Circulation in the Equatorial Stratosphere. *J. Atmos. Sci.*, **51**, 729 - 745, doi: 10.1175/1520-0469(1994)051<0729:Qboooa>2.0.Co;2.

Hegglin, M.I., *et al.*, 2006: Measurements of NO, NOy, N2O, and O3 during SPURT: implications for transport and chemistry in the lowermost stratosphere. *Atmos. Chem. Phys.*, **6**, 1331 - 1350, doi: 10.5194/acp-6-1331-2006.

Hegglin, M.I., *et al.*, 2008: Validation of ACE-FTS satellite data in the upper troposphere/lower stratosphere (UTLS) using non-coincident measurements. *Atmos. Chem. Phys.*, **8**, 1483 - 1499, doi: 10.5194/acp-8-1483-2008.

Hegglin, M.I., *et al.*, 2013: SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders. *J. Geophys. Res: Atmos.*, **118**, 11,824 - 11,846, doi: 10.1002/jgrd.50752.

Hegglin, M.I., *et al.*, 2021: Overview and update of the SPARC Data Initiative: comparison of stratospheric composition measurements from satellite limb sounders. *Earth Syst. Sci. Dat*a, **13**, 1855 - 1903, doi: 10.5194/essd-13-1855-2021.

Hólm, E., *et al.*, 2002: Assimilation and modelling of the hydrological cycle: ECMWF's status and plans. *ECMWF Tech. Memo*, **383**, 2002.

Hólm, E., 2003: Revision of the ECMWF humidity analysis: Construction of a Gaussian control variable. *Proceedings of the ECMWF/GEWEX Workshop on Humidity Analysis*, Reading, UK, 8 - 11 July.

Jiang, J.H., *et al.*, 2015: An assessment of upper troposphere and lower stratosphere water vapor in MERRA, MERRA2, and ECMWF reanalyses using Aura MLS observations. *J. Geophys. Res: Atmos.*, **120**, 11,468 - 11,485, doi: 10.1002/2015JD023752.

Kalnay, E., *et al.*, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Met. Soc.*, **77**, 437 - 472, doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kanamitsu, M., *et al.*, 2002: Ncep-Doe Amip-Ii Reanalysis (R-2), *Bull. Am. Met. Soc.*, **83**, 1631 - 1643, doi: 10.1175/BAMS-83-11-1631.

Kistler, R., *et al.*, 2001: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, *Bull. Am. Met. Soc.*, **82**, 247 - 267, doi: 10.1175/1520-0477.

Kobayashi, S., *et al.*, 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *J. Meteorol. Soc. Jpn.*, **93**, 5 - 48, doi: 10.2151/jmsj.2015-001.

Kramarova, N.A., *et al.*, 2013: Validation of ozone monthly zonal mean profiles obtained from the version 8.6 Solar Backscatter Ultraviolet algorithm. *Atmos. Chem. Phys.*, **13**, 6887 - 6905, doi: 10.5194/acp-13-6887-2013.

Levelt, P.F., et al., 2006: The Ozone Monitoring Instrument, *Ieee T. Geosci. Remote*, **44**, 1093 - 1101, doi: 10.1109/TGRS.2006.872333.

Ling, X.D., and J. London, 1986: The Quasi-Biennial Oscillation of Ozone in the Tropical Middle Stratosphere - a One-Dimensional Model. *J. Atmos. Sci.*, **43**, 3122 - 3137, doi: 10.1175/1520-0469.

Livesey, N.J., *et al.*, 2017: Version 4.2x Level 2 Data Quality and Description Document. *JPL D-33509* Rev. E, available at: **https://mls.jpl.nasa.gov/data/v4-2_data_quality_document.pdf**.

Manney, G.L., R.W. Zurek, A. O'Neill, and R. Swinbank, 1994: On the motion of air through the stratospheric polar vortex. *J. Atmos. Sci.*, **51**, 2973 - 2994, doi: 10.1175/1520-0469(1994)051<2973:OTMOAT>2.0.CO;2.

Manney, G.L., *et al.*, 1999: Polar vortex dynamics during spring and fall diagnosed using trace gas observations from the Atmospheric Trace Molecule Spectroscopy instrument. *J. Geophys. Res: Atmos.*, **104**, 18841 - 18866, doi: 10.1029/1999jd900317.

Manney, G.L., *et al.*, 2007: Solar occultation satellite data and derived meteorological products: Sampling issues and comparisons with Aura Microwave Limb Sounder. *J. Geophys. Res: Atmos.*, **112**, D24S50, doi: 10.1029/2007jd008709.

Manney, G.L., *et al.*, 2011: Jet characterization in the upper troposphere/lower stratosphere (UTLS): applications to climatology and transport studies. *Atmos. Chem. Phys.*, **11**, 6115 - 6137, doi: 10.5194/acp-11-6115-2011.

McCarty, W., *et al.*, 2016: MERRA-2 Input Observations: Summary and Initial Assessment. *NASA Technical Report Series on Global Modeling and Data Assimilation*, *NASA/TM-2016-104606*, **46**, 61.

McCormack, J.P., S.D. Eckermann, D.E. Siskind, and T.J. McGee, 2006: CHEM2D-OPP: A new linearized gas-phase ozone photochemistry parameterization for high-altitude NWP and climate models. *Atmos. Chem. Phys.*, **6**, 4943 - 4972, doi: 10.5194/acp-6-4943-2006.

McLinden, C.A., S. Tegtmeier, and V. Fioletov, 2009: Technical Note: A SAGE-corrected SBUV zonal-mean ozone data set. *Atmos. Chem. Phys.*, **9**, 7963 - 7972, doi: 10.5194/acp-9-7963-2009.

McPeters, R., *et al.*, 2008: Validation of the Aura Ozone Monitoring Instrument total column ozone product. *J. Geophys. Res: Atmos.*, **113**, D15S14. doi: 10.1029/2007jd008802.

McPeters, R.D., *et al.*, 2013: The version 8.6 SBUV ozone data record: An overview. *J. Geophys. Res: Atmos.*, **118**, 8032 - 8039, doi: 10.1002/jgrd.50597.

Mote, P.W., *et al.*, 1996: An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor. *J. Geophys. Res: Atmos.*, **101**, 3989 - 4006, doi: 10.1029/95jd03422.

Neu, J.L., *et al.*, 2014: The SPARC Data Initiative: Comparison of upper troposphere/lower stratosphere ozone climatologies from limb-viewing instruments and the nadir-viewing Tropospheric Emission Spectrometer. *J. Geophys. Res: Atmos.*, **119**, 6971 - 6990, doi: 10.1002/2013JD020822.

Oikonomou, E.K., and A. O'Neill, 2006: Evaluation of ozone and water vapor fields from the ECMWF reanalysis ERA-40 during 1991–1999 in comparison with UARS satellite and MOZAIC aircraft observations. *J. Geophys. Res.*, **111**, D14109, doi: 10.1029/2004JD005341.

Onogi, K., *et al.*, 2007: The JRA-25 reanalysis. *J. Meteorol. Soc. Jpn.*, **85**, 369 - 432, doi: 10.2151/Jmsj.85.369.

Paltridge, G., A. Arking, and M. Pook, 2009: Trends in middle- and upper-level tropospheric humidity from NCEP reanalysis data. *Theor. Appl. Climatol.*, **98**, 351 - 359, doi: 10.1007/s00704-009-0117-x.

Pan, L.L., *et al.*, 2012: Commentary on using equivalent latitude in the upper troposphere and lower stratosphere. *Atm. Chem. Phys.*, **12**, 9187 - 9199, doi: 10.5194/acp-12-9187-2012.

Parrish, A., *et al.*, 2014: Diurnal variations of stratospheric ozone measured by ground-based microwave remote sensing at the Mauna Loa NDACC site: measurement validation and GEOSCCM model comparison. *Atm. Chem. Phys.*, **14**, 7255 - 7272, doi: 10.5194/acp-14-7255-2014.

Poli, P., 2010: List of observations assimilated in ERA-40 and ERA-Interim (v1.0). *ERA Report Series*, **4**, 25p. available at: **https://www.ecmwf.int/en/elibrary/11692-list-observations-assimilated-era-40-and-era-interim-v10** (accessed: April 2021).

Randel, W.J., *et al.*, 1998: Seasonal cycles and QBO variations in stratospheric CH_4 and H_2O observed in UARS HALOE data. *J. Atmos. Sci.*, **55**, 163 - 185, doi: 10.1175/1520-0469(1998)055<0163:SCAQVI>2.0.CO;2.

Randel, W.J., *et al.*, 2006: Decreases in stratospheric water vapor after 2001: Links to changes in the tropical tropopause and the Brewer-Dobson circulation. *J. Geophys. Res: Atmos.*, **111**, D12312, doi: 10.1029/2005jd006744.

Rienecker, M.M., *et al.*, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624 - 3648, doi: 10.1175/JCLI-D-11-00015.1.

Saha, S., *et al.*, 2010: The Ncep Climate Forecast System Reanalysis. *Bull. Am. Met. Soc.*, **91**, 1015 - 1057, doi: 10.1175/2010Bams3001.1.

Saha, S., *et al.*, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185 - 2208, doi: 10.1175/JCLI-D-12-00823.1.

Sakazaki, T., et al., 2013: Diurnal ozone variations in the stratosphere revealed in observations from the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES) on board the International Space Station (ISS). *J. Geophys. Res: Atmos.*, **118**, 2991 - 3006, doi: 10.1002/jgrd.50220.

Santee, M.L., *et al.*, 2011: Trace gas evolution in the lowermost stratosphere from Aura Microwave Limb Sounder measurements. *J. Geophys. Res: Atmos.*, **116**, D18306, doi: 10.1029/2011JD015590.

Shibata, K., M. Deushi, T.T. Sekiyama, and H. Yoshimura, 2005: De- velopment of an MRI chemical transport model for the study of stratospheric chemistry. *Pap. Meteorol. Geophys.*, **55**, 75 - 119, doi: 10.2467/mripapers.55.75.

SPARC, 2017: The SPARC Data Initiative: Assessment of stratospheric trace gas and aerosol climatologies from satellite limb sounders. By M. I. Hegglin and S. Tegtmeier (eds.), *SPARC Report* No. **8**, *WCRP*-**5/2017**, available at **www.sparc-climate.org/publications/sparc-reports/**, doi: 10.3929/ethz- a-010863911.

Stajner, I., *et al.*, 2008: Assimilated ozone from EOS-Aura: Evaluation of the tropopause region and tropospheric columns. *J. Geophys. Res: Atmos.*, **113**, D16S32, doi: 10.1029/2007JD008863.

Takacs, L. L., Suárez, M. J., and Todling, R.: Maintaining atmospheric mass and water balance in reanalyses, Q J Roy Meteor Soc, 142, 1565-1573, 10.1002/qj.2763, 2016.

Tegtmeier, S., *et al.*, 2013: SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders. *J. Geophys. Res: Atmos.*, **118**, 12229 - 12247, doi: 10.1002/2013JD019877.

Thornton, H.E., *et al.*, 2009: The ASSET intercomparison of stratosphere and lower mesosphere humidity analyses, *Atm. Chem. Phys.*, **9**, 995 - 1016, doi: 10.5194/acp-9-995-2009.

Tompkins, A.M., K. Gierens, and G. Radel, 2007: Ice supersaturation in the ECMWF integrated forecast system. *Q. J. Roy. Met. Soc.*, **133**, 53 - 63, doi: 10.1002/qj.14.

Untch, A., A. Simmons, M. Hortal, and C. Jakob, 1998: Increased stratospheric resolution in the ECMWF forecasting system. *Proceedings of the workshop on chemical data assimilation*, KNMI de Bilt, Holland.

Uppala, S.M., *et al.*, 2005: The ERA-40 re-analysis. *Q. J. Roy. Met. Soc.*, **131**, 2961 - 3012, doi: 10.1256/qj.04.176.

Vömel, H., *et al.*, 2007: Validation of Aura Microwave Limb Sounder water vapor by balloon-borne Cryogenic Frost point Hygrometer measurements. *J. Geophys. Res: Atmos.*, **112**, D24S37, doi: 10.1029/2007JD008698.

Wargan, K., *et al.*, 2015: The global structure of upper troposphere-lower stratosphere ozone in GEOS-5: A multiyear assimilation of EOS Aura data. *J. Geophys. Res: Atmos.*, **120**, 2013 - 2036, doi: 10.1002/2014JD022493.

Wargan, K., *et al.*, 2017: Evaluation of the Ozone Fields in NASA's MERRA-2 Reanalysis. *J. Clim.*, **30**, 2961 - 2988, doi: 10.1175/jcli-d-16-0699.1.

Waters, J.W., *et al.*, 2006: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite. *Ieee T. Geosci. Remote.*, **44**, 1075 - 1092, doi: 10.1109/TGRS.2006.873771.

Wu, W.S., R.J. Purser, and D.F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Weather Rev.*, **130**, 2905 - 2916, doi: 10.1175/1520-0493.

Yan, X.L., et al., 2016: Validation of Aura MLS retrievals of temperature, water vapour and ozone in the upper troposphere and lower-middle stratosphere over the Tibetan Plateau during boreal summer. *Atmos. Meas. Tech.*, **9**, 3547 - 3566, doi: 10.5194/amt-9-3547-2016, 2016.

Zawodny, J.M., and M.P. Mccormick, 1991: Stratospheric Aerosol and Gas Experiment-Ii Measurements of the Quasi-Biennial Oscillations in Ozone and Nitrogen-Dioxide. *J. Geophys. Res: Atmos.*, **96**, 9371 - 9377, doi: 10.1029/91jd00517.

Major abbreviations and terms

1D-Var	1-dimensional variational data assimilation scheme
AIRS	Atmospheric Infrared Sounder
ATMS	Advanced Technology Microwave Sounder
ATOVS	Advanced TIROS Operational Vertical Sounder
Aura	A satellite in the EOS A-Train satellite constellation
CAMSiRA	Copernicus Atmosphere Monitoring Service interim Reanalysis
CDAS	Climate Data Assimilation System
CFC	chlorofluorocarbon
CFSR	Climate Forecast System Reanalysis of NCEP
CFSv2	Climate Forecast System, version 2
CHEM2D	The NRL 2-Dimensional photochemical model
CHEM2D-OPP	CHEM2D Ozone Photochemistry Parameterization
CREATE	Collaborative REAnalysis Technical Environment
CrIS	Cross-track Infrared Sounder
CTM	Chemical transport model
ECMWF	European Centre for Medium-Range Weather Forecasts
EOS	NASA's Earth Observing System
EqL	Equivalent Latitude
ERA-15	ECMWF 15-year reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA5	A forthcoming reanalysis developed by ECMWF
ERA-I / ERA-Interim	ECMWF interim reanalysis
GCM	Global Circulation Model

GNSS-RO	Global Navigation Satellite System Radio Occultation
GOME	Global Ozone Monitoring Experiment
HALOE	Halogen Occultation Experiment
HIRS	High-resolution Infrared Radiation Sounder
IASI	Infrared Atmospheric Sounding Interferometer
IR	Infrared
JMA	Japan Meteorological Agency
JRA-25	Japanese 25-year Reanalysis
JRA-55	Japanese 55-year Reanalysis
JRA-55AMIP	Japanese 55-year Reanalysis based on atmosphere-only simulations
JRA-55C	Japanese 55-year Reanalysis assimilating Conventional observations only
MERRA	Modern Era Retrospective-Analysis for Research
MERRA-2	Modern Era Retrospective-Analysis for Research 2
MIM	Multi Instrument Mean
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
MRI-CCM1	Meteorological Research Institute (JMA) Chemistry Climate Model, version 1
MSU	Microwave Sounding Unit
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
NRL	Naval Research Laboratory
NRT	near-real time
ODS	Ozone Depleting Substance
OMI	Ozone Monitoring Instrument
OSIRIS	Optical Spectrograph and InfraRed Imaging System
QBO	quasi-biennial oscillation
ppmv	parts per million by volume
pptv	parts per trillion by volume
R1	NCEP-NCAR Reanalysis 1
R2	NCEP-DOE Reanalysis 2
RDA	Research Data Archive
RH	Relative Humidity
RMS	Root Mean Square
SBUV & SBUV/2	Solar Backscatter Ultraviolet Radiometer
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Chartography
SDI	SPARC Data Initiative
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SH	Southern Hemisphere
SPARC	Stratosphere-troposphere Processes And their Role in Climate
S-RIP	SPARC Reanalysis Intercomparison Project
SSM/I or SSMI	Special Sensor Microwave Imager
SSU	Stratospheric Sounding Unit
STE	Stratosphere-Troposphere Exchange
SWOOSH	Stratospheric Water and OzOne Satellite Homogenized
TCWV	Total Column Water Vapour

TCO	Total Column Ozone
TIROS	Television Infrared Observation Satellite
TMI	Tropical Rainfall Measuring Mission (TRMM) Microwave Imager
TOMS	Total Ozone Mapping Spectrometer
TOVS	TIROS Operational Vertical Sounder
TTL	Tropical tropopause layer
UARS	Upper Atmosphere Research Satellite
UTLS	Upper Troposphere and Lower Stratosphere
UV	Ultraviolet
VTPR	Vertical Temperature Profile Radiometer
WV	Water Vapour

Chapter 5: Brewer-Dobson Circulation

Chapter lead authors

Beatriz M. Monge-Sanz	Atmospheric, Oceanic and Planetary Physics, University of Oxford	United Kingdom
Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany

Co-authors

Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Mohamadou Diallo	Institute of Energy and Climate Research, Stratosphere, Forschungzentrum Jülich	Germany
Florian Haenel	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Paul Konopka	Institute of Energy and Climate Research, Stratosphere,Forschungzentrum Jülich	Germany
Bernard Legras	Laboratoire de Météorologie Dynamique, IPSL, UPMC/ENS/CNRS/Ecole Polytechnique	France
Felix Ploeger	Institute of Energy and Climate Research, Stratosphere, Forschungzentrum Jülich	Germany
Thomas Reddmann	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Gabriele Stiller	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China
Marta Abalos	Department of Earth Physics and Astrophysics, Universidad Complutense de Madrid	Spain
Harald Boenisch	Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology	Germany
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration (NOAA)	USA
Hella Garny	(1) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) (2) Meteorological Institute, Ludwig-Maximilians University Munich	Germany
Peter Hitchcock	Department of Earth and Atmospheric Sciences, Cornell University	USA
Kazuyuki Miyazaki	Tropospheric Composition Group, Jet Propulsion Laboratory, California Institute of Technology *previously at: Japan Agency for Marine-Earth Science and Technology, Japan*	USA
Howard K. Roscoe	British Antarctic Survey	UK
Kaoru Sato	Department of Earth and Planetary Science, The University of Tokyo	Japan
Mengchu Tao	Institute of Energy and Climate Research, Stratosphere, Forschungzentrum Jülich	Germany
Darryn Waugh	Department of Earth and Planetary Sciences, Johns Hopkins University	USA

Abstract. This chapter focuses on the evaluation and comparison of the stratospheric circulation, using diagnostics based on the residual mean meridional circulation (*e.g.*, tropical upwelling), and on stratospheric transport tracers such as the age-of-air (AoA). Off-line chemistry-transport models in Eulerian and Lagrangian frameworks are used to compute tracer diagnostics for major recent reanalyses. Results are compared to those from observation-based datasets derived from satellite, ground-based, balloon, and aircraft observations of long-lived tracers such as SF_6, CO_2, and N_2O. Particular attention is given to comparing past trends in AoA from the different reanalyses with different offline chemistry-transport models (CTMs) driven by the reanalyses.

Dynamics diagnostics show that in recent reanalysis products the Brewer-Dobson circulation (BDC) is consistent in terms of climatological-mean structures with overall coherent interannual variability in metrics such as tropical upwelling at 70 hPa. However, estimates of long-term trends in tropical upwelling are inconsistent among different products, showing either strengthening, weakening, or no trend. Residual circulation transit times (RCTTs), a measure of the integrated circulation strength throughout the stratosphere, show large variability across different products, although long-term trend structures in RCTTs indicate a strengthening of the BDC, especially within its shallow branch.

Our comparison of AoA tracer results has shown that recent reanalyses produce mean AoA in much better agreement with observations than the previous generation of reanalysis, showing the improvement achieved by the reanalysis systems in the representation of the BDC. However significant discrepancies in AoA and tracers distribution among reanalyses still remain. For the overall period (1989 - 2010) our offline results show large spread in values and sign of mean AoA trends, depending on the reanalysis and on the region of the stratosphere. For the MIPAS period (2002 - 2012) only ERA-Interim is in good agreement with the observed trends, independently of the offline model used. We point to possible causes of these discrepancies and provide recommendations for users and for reanalyses centres. Much investigation is still needed on BDC trends, and factors affecting them, including natural variability and changes in the observation system of assimilated data.

Contents

5.1 Introduction 167
5.1.1 General description of the BDC and approach 167
5.1.2 Chapter objectives 167
5.2 Diagnostics description 168
5.2.1 Dynamical variables 168
5.2.2 Transport tracers from offline simulations 169
5.2.2.1 Introduction to offline modelling tools 169
5.2.2.2 Diabatic heating rates 169
5.2.2.3 Mean age-of-air 169
5.2.2.4 Age spectrum 170
5.2.2.5 Stratospheric water vapour tracer 170
5.3 Offline models description 171
5.3.1 Description of the BASCOE model 171
5.3.2 Description of the CLaMS model 171
5.3.3 Description of the KASIMA model 172
5.3.4 Description of the TOMCAT/SLIMCAT model 172
5.3.5 Description of the TRACZILLA model 173
5.4 Description of tracers observations 173
5.4.1 “Standard” mean AoA observations for model intercomparisons 173
5.4.2 Long timeseries of mean AoA in the northern hemisphere 173
5.4.3 MIPAS AoA dataset based on tracer observations 174
5.4.4 BAS Polar tracer observations 174
5.4.5 Stratospheric water vapour tape-recorder observations 174
5.5 Comparison results 175
5.5.1 Results from dynamical variables 175
5.5.1.1 Climatological description 175
5.5.1.2 Tropical upwelling trends 181
5.5.1.3 Tropical outwelling and RCTT trends 183
5.5.2 Results from transport tracers simulations 186
5.5.2.1 Heating rates 186
5.5.2.2 Mean age-of-air from observations 187
5.5.2.3 Mean AoA from offline models 189
5.5.2.4 Age spectrum 195
5.5.2.5 Mean AoA time evolution 195
5.5.2.6 Mean AoA trends 197
5.5.2.7 Impact of other processes on the AoA 202
5.5.2.8 Stratospheric water vapour tracer 205
5.6 Discussion 206
5.7 Conclusions and recommendations 210
References 214
Major abbreviations and terms 218

5.1 Introduction

5.1.1 General description of the BDC and approach

The Brewer-Dobson circulation (BDC) describes the major transport pattern in the stratosphere. The BDC was first postulated by *Brewer* (1949) and *Dobson* (1956) to explain measurements of water vapour and ozone in the stratosphere. The circulation is fundamentally driven by dissipating waves of tropospheric origin and broadly consists of large-scale tropical ascent and winter pole descent. The BDC is much weaker during boreal summer due to the different distribution of land masses and the associated differences in the generation of planetary and gravity waves between both hemispheres.

Modern reanalysis (reanalysis) products[1] include a spatially well-resolved stratosphere, which motivates an assessment of their BDC and its associated trends. Characteristics of the BDC may be obtained based on general circulation metrics, such as the residual mean meridional overturning streamfunction, from variables directly available from the reanalyses, and also from offline model simulations driven by the reanalyses meteorological fields (see *Section 5.2*). Using the residual mean streamfunction, *Iwasaki et al.* (2009) previously compared a general circulation model (GCM) with that from five reanalysis products: JRA-25, ERA-40, ERA-Interim, NCEP/NCAR, and NCEP/DOE. They found consistency in the extratropical winter circulation across the reanalyses, but also large discrepancies in low latitudes, and they found that trends were not reliable. Here we provide new results and discussions on mean streamfunction comparisons including additional modern reanalyses.

For multiannual stratospheric studies, an accurate representation of the BDC is essential for chemistry-transport models (CTMs) to achieve realistic tracer distributions. Offline CTMs take winds and temperatures from general circulation models or from meteorological analyses. The advantage of using analyses is that the CTM simulations are then linked to real meteorology and results are directly comparable to observations. Reanalyses extend this advantage into the past, enabling long-term simulations that provide valuable information on the temporal evolution of the atmospheric composition, helping to understand the present and predict the future. Therefore, CTMs rely on the quality of the reanalyses to obtain accurate tracers distributions. And, in turn, this reliance makes CTMs a powerful tool for evaluating the reanalyses themselves. This use of CTMs was proposed by *Monge-Sanz et al.* (2006) and applied during the preparatory phase of ERA-Interim (*Monge-Sanz et al.*, 2012; 2007; *Dee et al.*, 2011). Here, we have applied and extended such an approach using several complementary CTMs to evaluate modern reanalyses.

Recently, the use of reanalyses to nudge climate models is also becoming an emerging practice to constrain dynamics in climate simulations (*e.g.*, *Orbe et al.*, 2020; *Chrysanthou et al.*, 2019), which increases the need for accurate representation of circulation processes in the reanalyses. In some places, this Chapter includes as a point of comparison results based on chemistry-climate model (CCM) experiments from the Chemistry Climate Model Initiative (CCMI, *Morgenstern et al.*, 2017). Models included are (*cf.*, *Dietmüller et al.*, 2018): CMAM, EMAC, GEOSCCM, MRI-ESM1r1, NIWA-UKCA, SOCOL3, ULAQ-CCM, WACCM. Free-running climate models have the advantage that they provide more physically consistent estimates of metrics. Note that JRA-55AMIP effectively represents a climate model (with prescribed SSTs). ERA-20C and 20CR represent products that can be considered as half-way between free-running models and reanalyses.

5.1.2 Chapter objectives

In this Chapter we evaluate how well existing major reanalyses reproduce the BDC, and we provide an intercomparison among these reanalyses and against existing independent observations.

We have aimed at identifying potential causes for the differences we have found among reanalyses, as well as identifying key elements for a realistic representation of the BDC in the reanalysis systems, with a particular focus on model developments.

The final part of the Chapter provides a set of recommendations for reanalyses users and producers; for users to be aware of potential limitations in the datasets, and for producers to achieve further improvements in future reanalyses.

Beyond the intercomparison of the existing major reanalyses, this Chapter also contributes to increasing our scientific knowledge on stratospheric transport processes and provides an updated overview of studies looking into the BDC pattern using reanalyses. We have devoted a significant part of the Chapter to assess trends and variability in BDC diagnostics using the different reanalyses, aiming at shedding light onto the major research question of the apparent discrepancy between models and observations regarding the evolution of the BDC (*e.g.*, *Waugh*, 2009; *Butchart et al.*, 2006). Our analyses have contributed to identifying processes that affect the representation of the BDC and its evolution, and therefore processes that require further attention in future model development.

[1] To ease discussion of the results, we will distinguish "older" from "more recent" products frequently along the chapter, We generally consider ERA-Interim, MERRA-2, JRA-55, CFSR as "more recent" with ERA5 being the newest product. Full details on production dates for each reanalysis can be found in *Chapter 1* and *Fujiwara et al.* (2017).

The BDC governs the entry and distribution of air masses and constituents from the troposphere into and within the stratosphere. It also plays a major role in the exchange of key constituents, such as ozone, back into the troposphere. Therefore, changes in the BDC will affect the stratospheric concentrations of longer-lived trace gases whose sources are in the troposphere (*e.g.*, CFCs, CH_4, N_2O), as well as of their stratospheric products (*e.g.*, reactive chlorine gases, H_2O, reactive nitrogen gases). BDC changes will also affect the tropospheric concentrations of trace gases with large sources in the stratosphere (*e.g.*, ozone and water vapour). Since these gases have key impacts on atmospheric climate and chemistry, it is essential to understand what changes have occurred to the BDC in the past to be in a better position to predict those that will occur in future.

5.2 Diagnostics description

5.2.1 Dynamical variables

We use standard pressure level output and compute diagnostics consistently across all products (*Martineau et al.* 2018). Notably these pressure level data lack resolution in the shallow branch of the BDC (no level provided between 100 hPa and 70 hPa). Detailed tests were performed initially using ERA-Interim to study the impact of vertical resolution (model versus standard pressure levels), details of the numerical computation methods, and upper boundary conditions. We chose ERA-Interim because model-level diagnostics were available from previous work with slightly different numerical computation methods. These tests included comparing residual velocities computed independently by different groups.

The residual circulation mass streamfunction is defined based on the Transformed Eulerian Mean (TEM) framework as (*e.g.*, *Andrews et al.*, 1987):

$$\psi^* = \psi - ag^{-1}\cos\varphi \frac{\overline{v'\theta'}}{\partial_p\bar{\theta}} \quad (5.1),$$

with the Eulerian mean streamfunction given by:

$$\psi = ag^{-1}\cos\varphi \int_{TOA}^{p} \bar{v}\, dp' \quad (5.2).$$

Here, a is Earth's radius, g is acceleration due to gravity, φ is latitude, p is pressure, v is meridional velocity, and θ is potential temperature. Overbars denote zonal averages and TOA stands for top-of-atmosphere. The upwelling mass flux through a given level (*e.g.*, 70 hPa) is then defined as $\psi^*_{max} - \psi^*_{min}$, *i.e.*, as the difference between the residual streamfunction's maximum and minimum value on that level, which by definition corresponds to the net upward mass flux between the so-called turnaround latitudes. The turnaround latitudes mark those latitudes where residual mean flow is upward equatorward of them and downward poleward of them (Rosenlof, 1995).

Residual (TEM) circulation velocities are calculated directly based on:

$$\bar{v}^* = \bar{v} - \partial_p\left(\frac{\overline{v'\theta'}}{\partial_p\bar{\theta}}\right) \quad (5.3)$$

and

$$\bar{w}^* = \bar{w} + \frac{1}{a\cos\varphi}\partial_\varphi\left(\frac{\overline{v'\theta'}\cos\varphi}{\partial_p\bar{\theta}}\right) \quad (5.4).$$

We use finite centered differences for the numerical derivatives, where the pressure-derivative is computed as $\partial_p = (\partial_p\sigma)\partial_\sigma$, with $\sigma = (p_0/p)^\kappa$, where p_0=1000 hPa, $\kappa = R/c_p \approx 2/7$, and the analytical expression for $\partial_p\sigma = -\kappa\sigma/p$ is used. Issues arise primarily at the upper boundary due to: a) implementation of vertical derivative, and b) missing levels in the output (*e.g.*, ERA-Interim pressure output levels only extend to 1 hPa, but the actual underlying model levels extend to 0.1 hPa). We deal with issue a) by introducing a notional top-of-atmosphere (TOA) layer between the model top (or highest output level) and p_{TOA}=0 hPa. We set the average meridional velocity for this layer to half the velocity at the model/output top, which corresponds to setting the flow to zero at p_{TOA}. To deal with issue b) the following steps for the $\bar{v}^*$ upper boundary condition have been determined to be "optimal" empirically (by comparison to full model levels in the case of ERA-Interim). First, note that $\partial_p\bar{\theta}$ appears inside another vertical derivative, but generally represents a very smooth field. We therefore apply a simple extrapolation beyond the second to last level using a power law:

$$\partial_p\bar{\theta}|_{top} = \partial_p\bar{\theta}|_{top-1}\left(\frac{\sigma_{top}}{\sigma_{top-1}}\right)^\alpha \quad (5.5),$$

where

$$\alpha = \frac{\ln\left(\frac{\partial_p\bar{\theta}|_{top-1}}{\partial_p\bar{\theta}|_{top-2}}\right)}{\ln\left(\frac{\sigma_{top-1}}{\sigma_{top-2}}\right)} \quad (5.6),$$

where the indices "top", "top–1", "top–2" refer to the top output level, the next level below that, and the next further level below that. Furthermore, we assume that the heat flux contribution vanishes at p_{TOA}:

$$\bar{v}^*_{top} = \bar{v}_{top} - \frac{\left(\frac{\overline{v'\theta'}}{\partial_p\bar{\theta}}\right)_{top-1}}{p_{top-1}} \quad (5.7).$$

For $\bar{w}^*$ the horizontal derivatives are also taken as centred differences, using linear extrapolation to obtain values at the poles.

Finally, residual circulation transit times (RCTTs) are obtained as in *Birner and Bönisch*, 2011: residual circulation trajectories are run backward from specified arrival latitudes, pressures, and twice per month. RCTTs provide an integrated measure of the residual velocities (including the effects of transient changes in $\bar{v}^*$ and $\bar{w}^*$ over the transport pathway). They help to diagnostically distinguish different branches of the BDC.

Trajectories are terminated when they intersect the local (time-dependent) tropopause. The RCTT is the transit time along these trajectories (*cf.*, *Rosenlof,* 1995). For the tropopause we use the thermal tropopause level obtained from zonal-monthly-mean temperatures on the provided pressure levels. Furthermore, we set $\bar{w}^*_{top}=0$ and also $\bar{v}^*$ to zero at the poles, to avoid trajectories leaving the domain. Note that boundary conditions at the bottom of the domain do not influence the (stratospheric) RCTT calculations as long as the bottom level is well below the tropopause, which is the case for all data sets used here.

Note that our analyses do not include a temperature-based metric, such as used by *Fu et al.* (2015), even though such a metric has the advantage of being quite well constrained by observations (*e.g.*, satellite data). The fact that temperatures are well constrained by observations results in very close agreement across reanalysis products, and so an intercomparison of these products is in this case less insightful. A related intercomparison for tropical tropopause temperatures is presented in *Chapter 8* of this report.

5.2.2 Transport tracers from offline simulations

5.2.2.1 Introduction to offline modelling tools

Reanalyses are used by a wide range of models to drive offline simulations to study atmospheric composition and transport processes. Since these offline simulations rely on the quality of the meteorological fields used to drive them, offline models can be used as a very valuable tool to assess how realistic meteorological fields from reanalysis are.

In this Chapter we use several offline models with recognised worldwide experience in stratospheric scientific studies and applications. A description of the models we use here (BASCOE, CLaMS, KASIMA, TOMCAT, TRACZILLA) and key references can be found in *Section 5.3*. All of them have taken part in numerous intercomparison projects and international activities (*e.g.*, WMO Ozone Assessments, CCMVal model intercomparisons, StratoClim EU project).

By using several offline models we obtain a certain spread in the performance of the different reanalysis, which helps to overcome the sensitivity that a particular reanalysis may have to a particular offline model configuration. To the extent possible, we will also aim to explain differences in the performance of reanalyses due to differences among the offline models, but this type of research is out of the scope of this S-RIP Report. It is however being pursued as a follow-up project by several co-authors of this Chapter (*Monge-Sanz et al.*, in prep).

5.2.2.2 Diabatic heating rates

The diabatic heating rate field, Q/c_p as described by the equation below, gives information on the model temperature tendencies and is a fundamental component of the temperature budget; this field is used by some offline models in this study to compute vertical velocities.

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \mathrm{grad}\, T - \omega \left(\frac{\kappa T}{p} - \frac{\partial T}{\partial p} \right) = \frac{Q}{c_p} \tag{5.8}$$

In the atmosphere, diabatic heating includes effects of radiative heating, latent heat fluxes and turbulent heat transport, however reanalyses archive total diabatic temperature tendency, and temperature tendency from radiation. This leaves the latent heat flux and the turbulent heat mix as one same contribution to the tendency from the reanalyses data:

$$\frac{Q}{c_p} = \frac{Q_{\mathrm{rad}}}{c_p} + \left(\frac{Q_{\mathrm{lat}}}{c_p} + \frac{Q_{\mathrm{mix}}}{c_p} \right) \tag{5.9}$$

Diabatic heating rates (K/day) generated by the reanalysis forecast models are in general provided at 6-hourly time resolution. This field is based on average temperature tendencies over the length of the assimilation window, therefore, *e.g.,* for a 6-hour window, diabatic heating rates data would be centred at 03:00, 09:00, 15:00, and 21:00 rather than at the standard synoptic times 00:00, 06:00, 12:00, and 18:00.

Offline models operating on isentropic vertical coordinates use these temperature tendencies (heating rates) to calculate the cross-isentropic velocity (*Section 5.3*). It is therefore important to understand the differences that this field exhibits in the different reanalysis datasets we have used, as differences in this field will result in differences in transport and mixing, as well as in thermodynamic differences that impact tropical entry and ascent of atmospheric constituents.

5.2.2.3 Mean age-of-air

The mean age-of-air (AoA) is a standard diagnostic for stratospheric circulation widely used by models in the stratosphere. It gives information on the time spent by air parcels in the stratosphere after entering through the tropical tropopause from lower levels (*e.g.*, *Waugh and Hall*, 2002).

The main advantage of this diagnostic is that it can be computed from measurements of certain atmospheric tracers, *e.g.*, CO_2 and SF_6 tracers. These two long-lived constituents approximately fulfil the linearly conserved conditions, they have linearly increasing concentrations in the troposphere and no stratospheric sources or sinks, and can therefore be used to derive the stratospheric mean AoA. By measuring their concentrations in the stratosphere we can trace back how long air parcels have been residing in the stratosphere.

Doing this at different stratospheric locations provides a picture of the strength of the circulation in this atmospheric region. The annual mean cross-section of the mean AoA obtained in this way should look similar to the one inferred by *Waugh and Hall* (2002) (**Figure 1**). These two gases, CO_2 and SF_6, are complementary as the tropospheric annual cycle of CO_2 can affect age values in the lower stratosphere, while SF6 concentrations in the upper stratosphere are affected by mesospheric loss (*Reddman et al.*, 2001). The effect of the mesospheric SF_6 sink contributes to make mean AoA values older at higher stratospheric altitudes; we will discuss this effect in *Section 5.5.2.3.*

Full-chemistry models can compute the simulated AoA from the corresponding model CO_2 and SF_6 tracers concentrations, or they can use an idealised linear tracer. By using an idealised tracer, results are chemistry independent, then results from different models can be compared only in terms of transport differences, not chemistry differences among models. This use of an idealised AoA tracer has been employed in large stratospheric model intercomparisons like CCMI, CCMVal and CCMVal-2 (*Eyring et al.*, 2006; *Dietmüller et al.*, 2018), where CTMs and CCMs were compared in terms of their performance in the stratosphere.

To compare different reanalyses we can run several simulations with an offline CTM driven by the different datasets, keeping the CTM configuration unchanged so that the transport differences will be due to the different meteorological datasets used to drive the offline model. This approach was applied for instance to advise ECMWF during the preparatory phase of the ERA-Interim production (*Monge-Sanz et al.*, 2007). We are using a similar approach in this part of the Chapter to evaluate the S-RIP reanalyses datasets. The CTMs used for the offline simulations are described in *Section 5.3*.

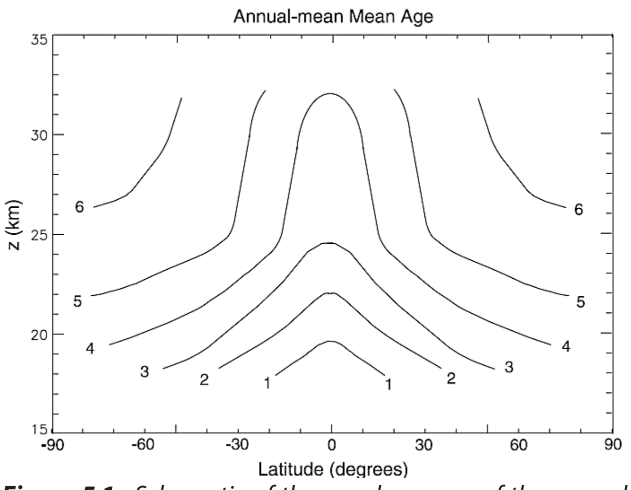


Figure 5.1: *Schematic of the zonal average of the annual mean of the mean AoA distribution (years), as inferred from observations as described in Waugh and Hall (2002). Figure from Waugh and Hall (2002). ©American Geophysical Union. Used with permission.*

Mean AoA trends

A significant part of this Chapter deals with the open and active scientific question that concerns potential trends in the mean AoA. It was initially motivated by the apparent disagreement between most climate models and existing long records of mean AoA observation based datasets (*e.g.*, *Engel et al.*, 2009; *Waugh et al.*, 2009) that was a matter of active debate when the S-RIP project started in 2012. This disagreement was also suggested by global observation datasets of AoA and by CTM simulations driven by ERA-Interim (*Stiller et al.*, 2012; *Monge-Sanz et al.*, 2012; *Diallo et al.*, 2012).

Whether observation datasets are showing long-term changes in the BDC or natural variability, and why most models are not capturing the same trend behaviour are among the scientific questions we address, to the extent possible, with the reanalyses evaluations and intercomparisons performed for this Chapter. Therefore, here we also evaluate the different reanalysis in terms of their ability to reproduce observed time evolution and trends in the mean AoA for the reanalysis period.

5.2.2.4 Age spectrum

The age spectrum is the statistical distribution of transit times for an air parcel from a source location, typically the Earth surface or the tropical tropopause, to a given location in the stratosphere (*Kida*, 1983). This concept was mathematically developed by *Hall and Plumb* (1994), who defined the age spectrum as a Green's function $G(x; x_0; t)$ that, for a tracer mixing ratio x, propagates in time a boundary condition from a source region x_0 (typically the tropical tropopause) into the stratosphere.

The mean age $\Gamma(x; x_0)$ at a certain stratospheric location is then the average over the age spectrum at that location:

$$\Gamma(x, x_0) = \int_0^\infty tG(x, x_0, t)dt \qquad (5.10).$$

The first moment of the spectrum is the mean age-of-air, as described in *Section 5.2.2.3*. Although the age-spectrum gives a more complete view of the stratospheric circulation than the mean age-of-air, it cannot be actually measured. It is the mean age value obtained from the spectra that we can compare against observation based AoA values. Nevertheless, an intercomparison of age-spectra derived with different reanalyses can yield valuable information on the different representation of the stratospheric circulation in each dataset.

5.2.2.5 Stratospheric Water Vapour tracer

A complete overview of stratospheric water vapour (SWV) in the different reanalyses is provided in *Chapter 4* of this Report and in *Davis et al.* (2017).

In the current Chapter we focus on the SWV distributions obtained from offline CTMs driven by the different reanalyses. This gives additional information, for each reanalysis dataset, on their ability to transport real constituents into and within the stratosphere, as well as additional information on the usefulness of each dataset for offline model applications. It also needs to be taken into account that SWV depends on several variables, including tropopause temperature in the different reanalysis, and it will not be only a diagnostic for stratospheric transport.

Entry values of water vapour through the tropical tropopause exhibit a seasonally varying signal, imposed by the seasonality in the tropopause temperatures. This makes water vapour concentration values in the tropical lower stratosphere appear as if they had been marked by a tape-recorder (*Mote et al.,* 1996; 1998). Over the tropical LS region, this so-called "tape recorder" diagnostic (timeseries, amplitude and phase of water vapour concentrations), provides information on the propagation of air masses into the stratosphere. This diagnostic is also one of the standard tests applied to stratospheric models to evaluate the representation of the subtropical mixing barrier. The vertical propagation of the tape recorder signal allows the estimation of the vertical ascent over the tropics. When deriving the tape-recorder with an idealised sinusoidal tracer, we can avoid its dependency on exact tropopause temperatures and the corresponding tape-recorder signal reflects only transport aspects. The phase lag of the tape-recorder signal is a good way to quantify the vertical velocity, while the amplitude decay mainly characterizes the strength of the tropical mixing barrier. Therefore the tape-recorder can be used as an additional way to assess the BDC over tropical latitudes. The SWV tape-recorder is also one of the stratospheric transport diagnostics that we can test against satellite observations such as from the HALOE and MLS instruments, or the merged SWOOSH dataset.

5.3 Offline models description

5.3.1 Description of the BASCOE model

BASCOE is a kinematic transport model (*Skachko et al.*, 2014). Its advection module is the Flux-Form Semi-Lagrangian (FFSL) scheme (*Lin and Rood*, 1996) configured to follow closely the recommendations of *Rotman et al.* (2001). The FFSL advection scheme is run on the native vertical grid of each reanalysis and a common low-resolution latitude-longitude grid with $2° \times 2.5°$ increments. It requires to input the surface pressure and horizontal velocity on a so-called Arakawa C-grid, *i.e.,* the zonal wind u must be staggered in longitude and the meridional wind v must be staggered in latitude. The FFSL algorithm evaluates internally the corresponding mass fluxes and derives the vertical winds (w) from mass conservation. Hence the reanalysis datasets must be pre-processed from spectral or high-resolution gridded fields to the low-resolution C-grid.

Special attention was paid to the preprocessing of the reanalyses to make sure that the different types of wind fields were expressed in a consistent manner for the BASCOE transport algorithm. For the five reanalysis datasets used by BASCOE (ERA-Interim, JRA-55, MERRA, MERRA-2 and CFSR) a preprocessing algorithm based on *Segers et al.* (2002) is used, with additional preliminary derivation of the spherical harmonics coefficients of vorticity, divergence and surface pressure for reanalyses other than ERA-Interim. In all cases, these spectral coefficients are truncated at wavelength 47 to avoid aliasing on the $2° \times 2.5°$ target BASCOE grid.

In the BASCOE simulations the AoA is derived from an idealized tracer with a concentration that increases linearly with time at the surface. To allow quick propagation of this boundary condition to the free troposphere, eddy vertical diffusion is modeled in the lower half of the troposphere with a vertical diffusion coefficient K_{zz} decreasing from $10\,m^2\,s^{-1}$ at the surface to zero at the pressure level halfway between the surface and the tropopause. There is no other representation of convection in the BASCOE model nor any explicit mechanism for horizontal diffusion.

5.3.2 Description of the CLaMS model

The Chemical Lagrangian Model of the Stratosphere (CLaMS) is a Lagrangian transport model with trace gas transport based on the motion of 3-D forward trajectories and an additional parameterization of subgrid scale atmospheric mixing, which relates mixing to deformations in the large-scale flow (*Konopka et al.*, 2004; *McKenna et al.*, 2002). The model uses an isentropic vertical coordinate, with vertical transport driven by the total diabatic heating rate (*Ploeger et al.*, 2010). Meteorological fields from the reanalyses are read in 3-hourly timesteps (horizontal winds and diabatic heating rates).

For this study, CLaMS simulations use fields from ERA-Interim (*Dee et al.*, 2011), JRA-55 (*Kobayashi et al.*, 2015) and MERRA-2 (*Gelaro et al.*, 2017) reanalyses. For driving the CLaMS model simulations, horizontal winds and diabatic heating rates from the reanalysis forecast are used on native model levels and with a horizontal resolution of $1° \times 1°$ in latitude and longitude. The AoA results from the different simulations have been interpolated to potential temperature levels (same for all reanalyses) and monthly zonal mean climatologies have been created.

CLaMS provides evaluation of the representation of SWV, mean age of air (AoA) and age spectrum. The evaluation is based on the modelled quantities as monthly and zonal means from 1979 to 2015. In the stratosphere and the UTLS, potential temperature is employed as the vertical coordinate of CLaMS, and the cross-isentropic velocities are derived from the total diabatic heating rates provided by the reanalysis products, including effects of radiative heating, turbulent heating and heating release.

The model configuration follows the model setup described in *Pommrich et al.* (2014) with 100 km horizontal and 250 m vertical resolutions around the tropical tropopause. The age spectrum diagnostic computation used by CLaMS is described in *Ploeger and Birner* (2016); accordingly AoA spectrum is calculated for each reanalysis from multiple tracer pulses and the mean age value is obtained from the spectrum (*Ploeger et al.*, 2019). It is worth noting that in the CLaMS simulations shown here, an upper boundary condition is imposed for the mean AoA values, by prescribing top of the model values with MIPAS derived AoA.

5.3.3 Description of the KASIMA model

The Karlsruhe Simulation of the Middle Atmosphere (KASIMA) model is a three-dimensional mechanistic model of the middle atmosphere solving the primitive equations including middle atmosphere chemistry (*Kouker et al.*, 1999). For the simulations used here, the model was run on isobaric surfaces from 7 km to 120 km with a vertical resolution of 750 m in the stratosphere, gradually increasing to 3.8 km at the upper boundary. The horizontal resolution in the simulation is 5.4° × 5.4° (T21). The model is coupled to the specific meteorology by using the analyzed geopotential field at the lower boundary (7 km) and applying analyzed vorticity, divergence and temperature fields from ECMWF ERA-Interim below 1 hPa. Above 1 hPa radiative heating rates were calculated using a 2D climatology for ozone and H_2O. The parameterization of the gravity-wave drag is based on the formulation of *Holton* (1982). The parameterization has been modified compared to the version described in *Kouker et al.* (1999) in order to better describe the cross-mesopause transport often observed after sudden stratospheric warmings (SSWs). The spectral distribution of the vertical momentum flux is now described with a Gaussian function of a centroid of 7 m s^{-1} and a standard deviation of 50 m s^{-1} with phase speeds of 0, 20, 40, 60 and 80 m s^{-1}. The filter condition for critical phase speeds has been extended to be applied when the absolute difference between the speeds is less than 10 m s^{-1}. The latter condition effectively prevents gravity waves of low phase speed from propagating and breaking in the lower mesosphere. Only gravity waves of higher phase speed then break at higher altitudes, causing an elevated stratopause to build. In addition, the numerical implementation of the vertical diffusion has been re-formulated for better mass conservation according to *Schlutow et al.*, 2014.

KASIMA has used the following artificial tracers to derive the mean age-of-air: T1 is an idealized tracer exhibiting a linear trend. For T1 the mean age Γ is just the lag time $\Lambda 1$. T2 is a tracer initialized with a time series of mixing ratio data of SF_6 complemented by the data taken from the NOAA HATS (Halocarbons and other Atmospheric Trace Species) data set data set (**https://www.esrl.noaa.gov/gmd/hats/**). No chemical loss is assumed for that tracer and a lag time $\Lambda 2$ is deduced. With this tracer we study non-linearity effects in the trend curve. The tracer T3 is defined as T2, but includes a chemical loss as described by *Reddmann et al.* (2001), using the version including all relevant reactions. Chemical loss of a mean age tracer with a positive trend results in an apparent higher age as the tracer shows a lower mixing ratio than expected. SF_6 exhibits a significant mesospheric loss by electron attachment and subsequent reactions as described by *Reddmann et al.* (2001). Tracer T3 includes these loss mechanisms and the (apparent) lag time is calculated as for the inert tracer case. Whereas tracer T3 should be the most realistic tracer to be compared with SF_6 observations there are caveats as the loss mechanism of SF_6 is subject to significant uncertainties. With the tracer T3 we test the influence of the mesospheric loss on the derivation of stratospheric mean age.

As SF_6 shows a pronounced inter-hemispheric difference in the mixing ratio in the troposphere, the inter-hemispheric difference was imprinted to the mixing ratio of SF_6 at the lower boundary inside the troposphere in the form $A \cdot \tanh(1.5 \sin(\phi))$ with ϕ the geographic latitude and $A = 0.55$ years for the ideal linear tracer, and a mixing ratio difference for tracers T1 and T2 corresponding to an amplitude of the hemispheric difference of approximately 1 year. The tracers were formally initialized for 1965, and the first two years of the ERA-Interim reanalyses were used repeatedly till 1979 to bring the tracers to an approximately steady state. Two years were used for spin-up to include an approximate full QBO period.

5.3.4 Description of the TOMCAT/SLIMCAT model

TOMCAT/SLIMCAT is a 3D offline CTM (*Chipperfield*, 2006). The CTM is flexible in terms of the winds datasets it can use, however, the ECMWF datasets have been the only ones that this model has extensively used since 1999, when they were extended into the stratosphere up to 0.1 hPa, and even more so after the completion of ERA-40 for multiannual runs for long-term chemical investigations (*e.g., Feng et al.*, 2007; *Chipperfield et al.*, 2005). The reanalysis fields are read in typically every 6 h, but this can be adapted to other available frequencies (*e.g., Monge-Sanz et al.*, 2012). In TOMCAT/SLIMCAT, read-in fields are interpolated in time to intermediate time steps (of 60 minutes in the case of the runs considered here).

The horizontal grid of the CTM is completely variable in resolution and in latitudinal regularity. The ECMWF (re) analyses are read in as spectral coefficients, which are then converted to grid-point fields by a spectral transform on to the CTM prescribed latitudinal grid using pre-tabulated integrals of the associated Legendre functions (*Chipperfield*, 2006). In this way the CTM is not restricted only to the usual Gaussian latitudes. Also, the number of vertical levels is flexible and the vertical coordinate can be either $\sigma - p$ (TOMCAT mode) or $\sigma - \theta$ (SLIMCAT mode). Vertical motion is calculated from the divergence of the horizontal winds.

In the case of ECMWF datasets, the divergence field is directly taken from the reanalyses or operational analyses. The conservation of second-order moments non-diffusive advection scheme by *Prather* (1986) is used in the CTM runs.

TOMCAT/SLIMCAT also includes a module for the calculation of particle trajectories, which allows for a Lagrangian as well as the default Eulerian approach. The trajectory position is computed from the same meteorological data used to force the Eulerian simulations; horizontal and vertical motion are calculated at the centre of the Eulerian grid and then interpolated to the trajectory position in that particular grid cell. An explicit fourth-order Runge–Kutta method (*Fisher et al.*, 1993) is used to advance the trajectory position forward (or backward) in time. The same general configuration options (vertical coordinate, vertical motion) are also available for the Lagrangian runs.

5.3.5 Description of the TRACZILLA model

TRACZILLA is a Lagrangian transport model derived from FLEXPART (*Pisso and Legras*, 2008). The simulations used here are performed by launching parcels from a 3-D grid on 32 isentropic levels in the stratosphere (from 300 K to 1420 K) on a 2° by 3° latitude-longitude grid, every ten days over the 22-year period 1989 - 2010. The trajectories are integrated until they cross the tropopause, determining the age of the parcel at that time, or until they reach a maximum age of ten years. In the simulations shown here, the motion of air parcels is governed by horizontal velocity fields and radiative heating rates from three reanalyses: ERA-Interim, MERRA and JRA-55.

For ERA-Interim, TRACZILLA uses meteorological data on model levels up to 0.1 hPa at 1° resolution in latitude and longitude and 3-hourly temporal resolution. For JRA-55, the model uses data on model levels up to 0.1 hPa at 0.56° resolution in latitude and longitude and 6-hourly temporal resolution. In MERRA, we use data on pressure levels up to 0.1 hPa (heating rates not available above) at 0.62° horizontal resolution and 3-hourly temporal resolution. 200 million trajectories have been used in this TRACZILLA S-RIP study.

TRACZILLA calculates the mean age-of-air by averaging over all parcels that cross the tropopause; the contribution of parcels that have not crossed it is calculated based on the well-established approximation of an exponential tail based on *Scheele et al.* (2005). TRACZILLA applies correction techniques to the trajectories calculation: first a uniform horizontal heating is applied on pressure levels to correct the lack of mass conservation when using radiative heatings in the stratosphere. Second, the trajectories which go above 0.5 hPa (*i.e.,* 2300 K) are discarded (clipped). This is a common correction technique in Lagrangian studies, *e.g.*, in *Schoeberl and Dessler* (2011) trajectories were clipped above 1800 K. The clipping level at 0.5 hPa was chosen in order for ERA-Interim to provide the best fit of the reconstructed AoA values to the aircraft and balloon observations derived from CO_2, N_2O and CH_4 during the SOLVE campaign (*Andrews et al.*, 2001).

5.4 Description of tracers observations

This Section provides a brief overview of the independent observation-based datasets we have used to validate the AoA and tracer distributions from the offline model simulations. We also include key references for more detailed descriptions of these measurement datasets.

5.4.1 "Standard" mean AoA observations for model intercomparisons

Mean AoA can be calculated from measurements of concentrations of long-lived tracers with an approximately linear increasing trend at the surface, such as CO_2 or SF_6. Between 1992 and 1998 NASA ER-2 aircraft and high-altitude balloons measured concentrations of CO_2 and SF_6. The ER-2 measurements were part of the campaigns Stratospheric Photochemistry Aerosol and Dynamics Experiment (SPADE), Airborne Southern Hemisphere Ozone Experiment/Measurements for Assessing the Effects of Stratospheric Aircraft (ASHOE/MAESA), Stratospheric Tracers of Atmospheric Transport (STRAT) and Photochemistry of Ozone Loss in the Arctic Regions in Summer (POLARIS). Balloon flights were part of the Observations of the Middle Stratosphere (OMS) experiments.

Multidecadal datasets were compiled from these balloon soundings and aircraft flights using both CO_2 and SF_6 measurements (*e.g.*, *Andrews et al.*, 2001; *Ray et al.*, 2014; *Ray et al.*, 1999) that have been widely used in model intercomparison studies (*e.g., Dietmüller et al.*, 2018; *Eyring et al.*, 2006), and have become a standard reference to monitor model development and circulation processes in the stratosphere (*e.g., Ploeger et al.*, 2019; *Chabrillat et al.*, 2018; *Butchart*, 2014; *Monge-Sanz et al.*, 2012).

5.4.2 Long timeseries of mean AoA in the Northern Hemisphere

For the NH mid-latitudes, a long time series of mean AOA derived from balloon-borne measurements of CO_2 and SF_6 exists that dates back to the mid 1970's *(Engel et al.*, 2009; 2017). The balloon-borne observations used in *Engel et al.* (2009) were taken in the region between 24 km and 35 km over NH midlatitudes, where the vertical gradient in mean AoA was found to be very small. The balloon data were limited to a total of 28 flights over a 30 year period, from 1975 to 2006, and showed a positive trend of 0.24 years per decade for this region.

Although this trend was estimated to be within the observational uncertainty, it pointed to an important potential disagreement between observations and most models (*e.g.*, *Waugh*, 2009). The dataset used in *Engel et al.* (2009) has been more recently extended using the new AirCore in-situ measurements (*Engel et al.*, 2017), which have also helped to narrow the trend uncertainty from the previous dataset. These NH data have become widely used by offline model studies concerned with the active debate of BDC trends in reanalyses (*e.g.*, *Ploeger et al.*, 2019; *Chabrillat et al.*, 2018; *Mahieu et al.*, 2014; *Monge-Sanz et al.*, 2012; *Diallo et al.*, 2012).

Balloon measurements described described here reached maximum altitudes of 31 km, while the aircraft missions reached up to 21 km, which limits the altitude range covered by these datasets to the LS and middle stratosphere. Also the latitude range is limited as the OMS flights covered only three latitude values (65°N, 35°N and 7°S) and the measurements used in *Engel et al.* (2009; 2017) are limited to the NH midlatitudes LS region. We therefore need additional observations that provide mean AoA values derived from global coverage measurements, based on the MIPAS satellite observations.

5.4.3 MIPAS AoA dataset based on tracer observations

Global coverage time series have been derived from satellite observations of SF_6 retrievals from the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS; *Fischer et al.,* 2008) satellite instrument, which provided an updated global dataset for the period 2002 - 2012. MIPAS was an instrument on board the Envisat satellite, measuring the mid infrared emission of the atmosphere against the space background. The measurements were done in limb scanning mode covering an altitude range of cloud top (or about 6 km in cloud-free cases) to about 72 km. The emission signatures of molecules in the atmosphere were used to retrieve the spatial distribution of up to 30 different trace gases and temperature with good global coverage from pole to pole, also during (polar) night. The mission extended from July 2002 to April 2012.

Information on the stratospheric mean AoA is obtained from the spatio-temporal distribution of the SF_6 tracer, measured by MIPAS with a vertical resolution of 4 km to 6 km and a single profile precision of about 10 - 20 %. Although the single profile precision is rather low, the huge number of profiles measured (more than 2 million profiles over the MIPAS mission lifetime) provided very valuable information on AoA from zonal mean distributions. The SF_6 distributions were retrieved from the upper troposphere up to about 50 km. Above 35 km, the systematic errors become larger and the vertical resolution deteriorates; for this reason quantitative analysis of SF_6 and AoA above 35 km is not recommended. Very detailed descriptions of this global dataset can be found in *Stiller et al.* (2012) and *Haenel et al.* (2015).

5.4.4 BAS Polar tracer observations

Stratospheric measurements of polar summer NO_2 were used by *Cook and Roscoe* (2009; 2012) to derive trends in the BDC. NO_2 data were measured from a zenith-sky spectrometer set up at Faraday in the Antarctic (65.25°S, 64.27°W) between 1990 and 1995, and then from the nearby site of Rothera (67.57°S, 68.13°W) since 1996, providing almost continuous measurements of Antarctic NO_2 since 1990 (*Roscoe*, 2004; *Roscoe et al.*, 2001).

Stratospheric column of NO_y over these Antarctic stations were obtained from measurements of NO_2 taken during the period 1990 - 2007. A photochemical box model and observed ozone and temperature profiles were used to determine column values. The years 1991 and 1992 were excluded because of the large amounts of volcanic aerosols from the Pinatubo eruption still present in the stratosphere. Full details and discussions related to this dataset of measurements by the British Antarctic Survey (BAS) can be found in *Cook and Roscoe* (2009; 2012).

5.4.5 Stratospheric water vapour tape-recorder observations

The seasonally varying signal of the water vapour in the tropical stratosphere, the so-called "tape recorder" signal (*Mote et al.,* 1996; 1998), reflects how rapidly air masses are transported upwards from the tropical tropopause into the stratosphere. The tape recorder is thus a good measure for the strength of the BDC over the tropics.

Observationally based values of the time series of $2CH_4 + H_2O$ measured by HALOE from 1992 - 1997 were analysed by *Mote et al.* (1998) with an empirical orthogonal function method. The amplitude and phase of the tape recorder signal were derived from this method, together with estimations from in-situ CO_2 observations (*Boering et al.*, 1996). These data have been extensively used for model validations and intercomparisons.

More recently, the Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database provided by the NOAA Chemical Sciences Laboratory (CSL) extends the coverage period merging vertically resolved water vapor data from the SAGE-II/III, UARS HALOE, UARS MLS, and Aura MLS satellite instruments starting from 1984 to present (*Davis et al.*, 2016). The homogenization process described by *Davis et al.* (2016) is designed to minimize artificial jumps in time and account for inter-satellite biases. The merged SWOOSH data thus provide a long-term SWV time series with reliable representations of interannual to decadal variability. We use the SWOOSH zonal-mean monthly mean time series of merged water vapor mixing ratios to assess offline model simulations of SWV tracer distributions and variability.

5.5 Comparison results

5.5.1 Results from dynamical variables

5.5.1.1 Climatological description

Figure 5.2 summarizes the climatological structure of the BDC for the multi-reanalysis-mean (REM) during December-January-February (DJF) and June-July-August (JJA) in terms of $\overline{w}^*$ (converted to mm s^{-1}) and ψ^*. Here and in the following we define climatologies based on the period 1980 2010 and include MERRA-2, ERA-Interim, JRA-55, CFSR in the REM. Turnaround latitudes are shown based on the extrema of ψ^*. Note that because ψ^* is calculated from $\overline{v}^*$ it is not everywhere consistent with $\overline{w}^*$ (see *Eqs. 5.3, 5.4*), so that the turnaround latitudes do not everywhere match $\overline{w}^*=0$. Overall, these fields are consistent with previous studies showing the climatological BDC structure from individual reanalyses (*e.g.*, *Miyazaki et al.*, 2016; *Abalos et al.*, 2015; *Iwasaki et al.*, 2009).

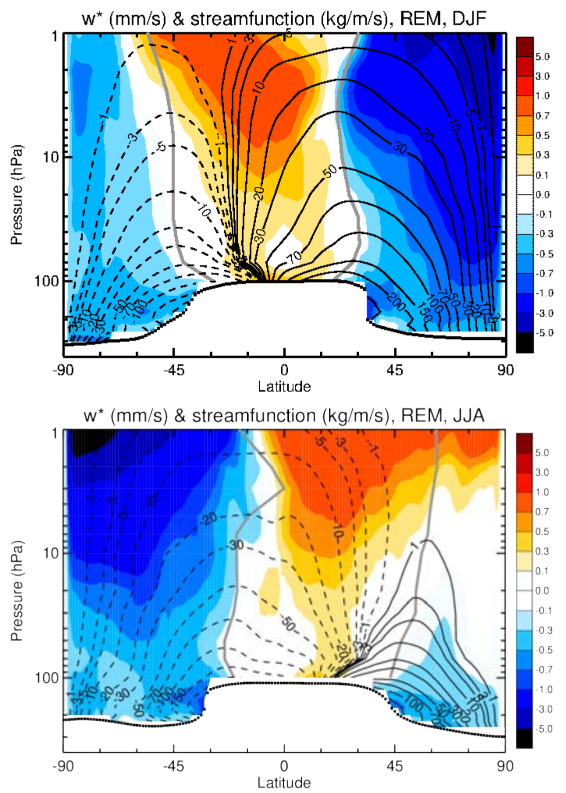


Figure 5.2: *Climatological (1980-2010) REM structures of* $-p\overline{w}^*/H$ *(with scale height H=7 km, color shading) and* ψ^* *(divided by Earth's radius to ease comparison to previous literature, contours) for December-January-February (top) and June-July-August (bottom), respectively. Full gray lines mark the turnaround latitudes based on the* ψ^* *fields. Thick gray dots mark the tropopause location based on climatological temperatures. Fields are only shown above the tropopause.*

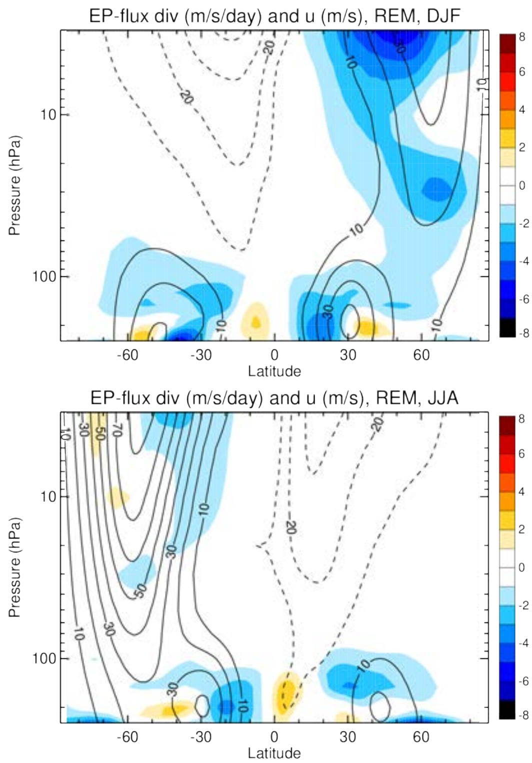


Figure 5.3: *Climatological (1980-2010) REM structures of EP-flux divergence (color shading) and zonal mean zonal wind (contours) for December-January-February (top) and June-July-August (bottom), respectively. Note that the lowest shown level is 250 hPa, which is the tropopause in the extratropics.*

The BDC is driven by wave forcing, which, for the resolved waves, can be quantified based on the Eliassen-Palm flux (EP-flux) divergence shown in **Figure 5.3** for the REM for DJF and JJA together with the zonal mean zonal winds. The structure of EP-flux divergence roughly indicates separate wave forcing for the shallow versus deep circulation branch (*cf.*, *Plumb* 2002; *Konopka et al.*, 2015): the lowermost stratospheric forcing of the shallow branch is present in both seasons and hemispheres, whereas the mid to upper stratospheric forcing of the deep branch is only present during each hemisphere's winter season.

We quantify the wave forcing of the shallow versus deep branch and their seasonality by creating respective hemispheric and vertical averages: 70 - 100 hPa for the shallow branch and 3 - 50 hPa for the deep branch (note that 3 hPa is the highest diagnosed level). Note, that these refer to the total resolved wave forcing[2]; the individual contributions due to Rossby and gravity waves are studied in detail in *Sato and Hirano* (2019). Their seasonal climatological evolutions are shown in **Figure 5.4** for all diagnosed reanalyses. Overall, wave forcing is quite consistent between different reanalysis products. The largest spread is found for the shallow branch forcing in the NH throughout the year, as well as for the deep branch forcing in the NH winter and for the shallow branch forcing in the Southern Hemisphere (SH) spring and summer.

[2] These may include a gravity wave contribution insofar as these waves are resolved.

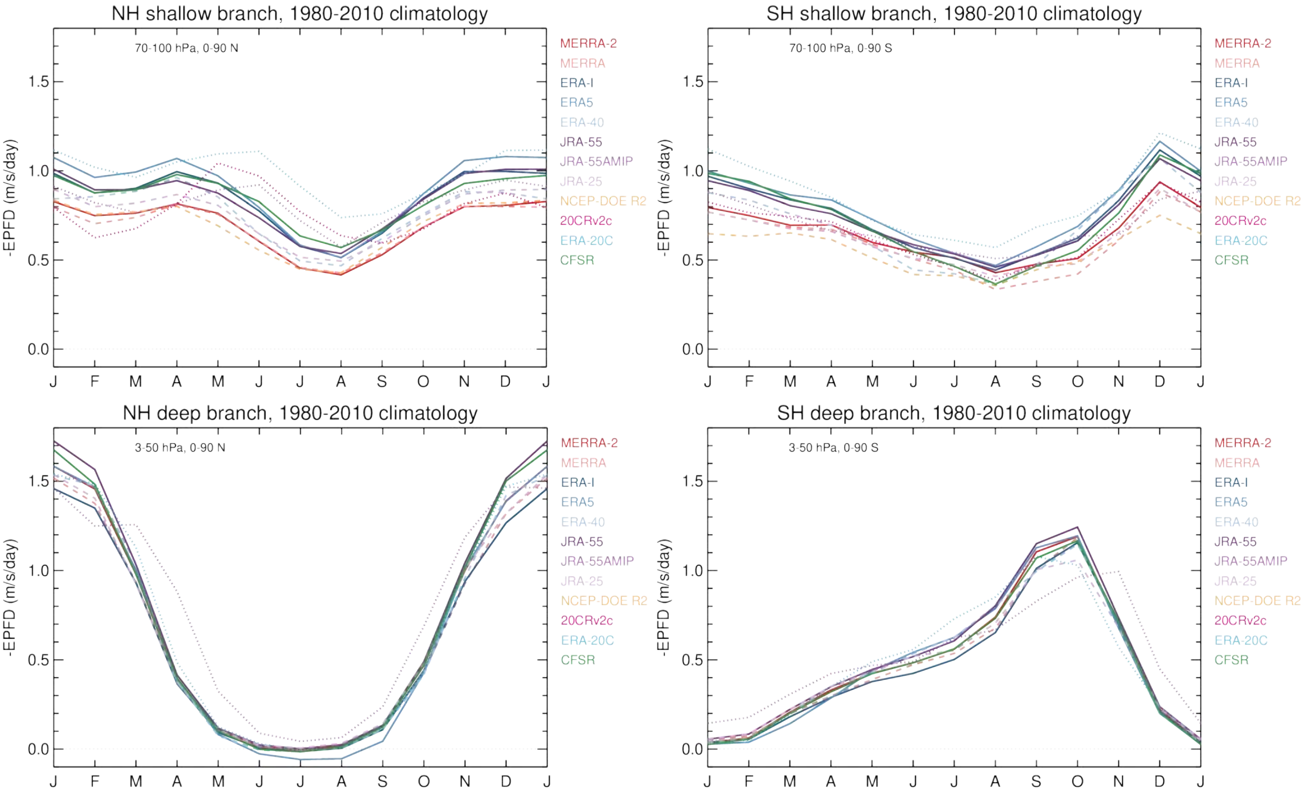


Figure 5.4: *Climatological (1980 - 2010) seasonal evolutions for each reanalysis of the EP-flux divergence for the shallow and deep BDC branches and for each hemisphere separately. Full lines refer to more recent reanalysis products, dashed lines to older reanalysis products, and dotted lines to other products.*

ERA-Interim, JRA-55, and CFSR agree very well except for the deep branch during NH winter, whereas MERRA-2 shows overall less wave forcing. Interestingly, ERA5 shows consistently stronger wave driving of the shallow branch throughout the year in both hemispheres, strongest in NH winter. This could perhaps be due to contributions from partially resolved gravity waves in this much higher resolution product. JRA-55AMIP (the freely evolving atmosphere model version of JRA-55) shows a persistent bias in its seasonality, with a delayed drop-off in wave forcing during spring for the NH deep branch and a delayed peak in wave forcing during spring for the SH deep branch.

The mass overturning at 70 hPa shows a considerable degree of uncertainty between the different products: although the qualitative structure with extratropical downwelling and tropical upwelling is consistent, structural aspects of the upwelling vary strongly in some cases (**Figure 5.5**). For example, the local minimum in tropical upwelling near the equator (*Ming et al.*, 2016) is very pronounced for MERRA-2 and CFSR (minimum roughly zero), but is only weakly present in JRA-55. Of the older products JRA-25 does not exhibit a local minimum, ERA-40 shows a noisy and too narrow upwelling structure, and MERRA shows local downwelling over the equator. The peak in SH tropical upwelling strength during JJA is about one order of magnitude smaller in JRA-55 compared to CFSR (not shown).

Tropical upwelling at 70 hPa is known to exhibit strong seasonality (*e.g.*, *Rosenlof* 1995). However, this arises primarily due to its contribution in the SH where upwelling is much stronger during the NH cold season (**Figure 5.6**). Here, SH upwelling is simply calculated via the streamfunction difference between the equator and its SH minimum. Likewise, NH upwelling is based on the streamfunction difference between its NH maximum and the equator. NH tropical upwelling shows a weak seasonal cycle with stronger upwelling during the SH cold season in some products (most pronounced in JRA-55 and MERRA-2, similar to the climate models), but seasonality is generally inconsistent between products.

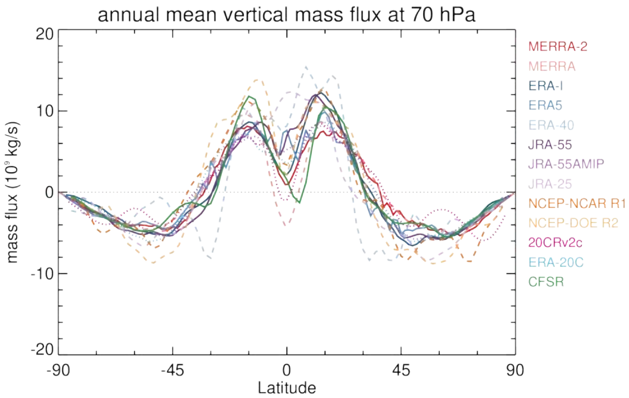
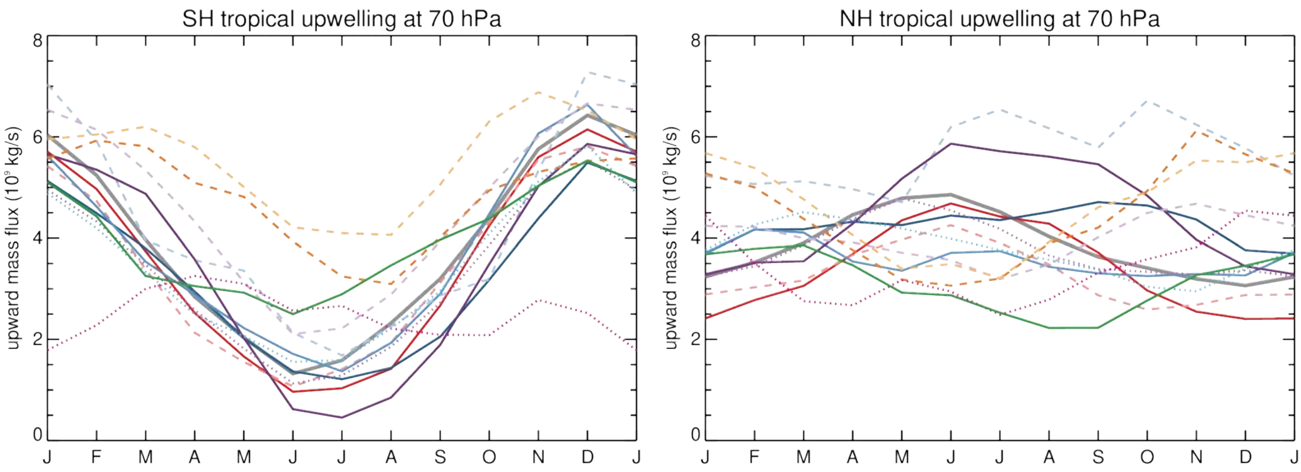


Figure 5.5: *Climatological (1980 - 2010) vertical mass flux at 70 hPa as a function of latitude for the annual mean. Line styles as in **Figure 5.4**.*



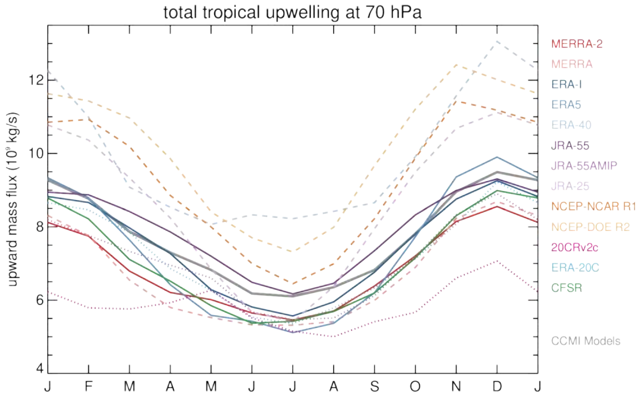
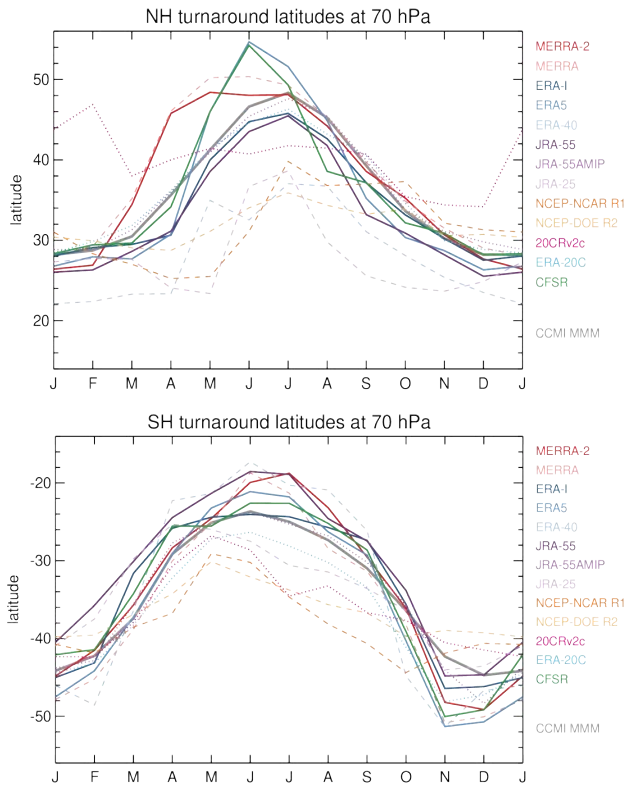


Figure 5.6: *Climatological (1980-2010) tropical upwelling characteristics for each individual product at 70 hPa. Top: individual hemispheric contributions. Bottom: total tropical upwelling between turnaround latitudes (i.e., sum of hemispheric contributions shown in the top panels). Gray lines show the multi-model-mean (MMM) of the CCMI models.*



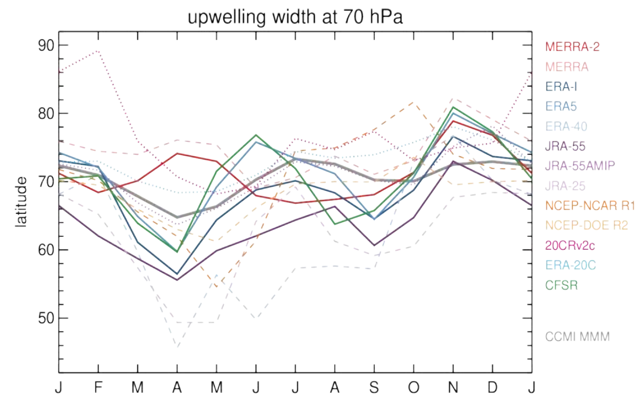


Figure 5.7: *Climatological turnaround latitudes (left) and tropical upwelling width (right) for each individual product at 70 hPa. Line styles as in* ***Figure 5.4****.*

When combined, total tropical upwelling has a consistent upwelling seasonal cycle. Its amplitude is significantly larger in the older products (dashed lines in **Figure 5.6**, bottom), although the newest product (ERA5) also shows a larger amplitude than its predecessor (ERA-Interim), MERRA-2, JRA-55, CFSR, and the chemistry climate models (CCMs). The more recent products show a very similar seasonal cycle to the CCMs. 20CR stands out as essentially completely missing upwelling seasonality and generally showing much too weak upwelling. CFSR shows a consistent seasonality for total tropical upwelling, however, this arises due to compensating biases between the hemispheres during May-September. ERA-20C and JRA-55AMIP agree overall quite well with the newer reanalysis products. Note that seasonality in extratropical downwelling is generally very consistent across products (not shown), including its hemispheric differences.

The annual cycle in total tropical upwelling at 70 hPa is primarily determined by the annual cycle in local upwelling strength, as opposed to the annual cycle in the upwelling width (see **Figure 5.7**). Even though seasonal variations in turnaround latitudes are large in each hemisphere, with turnaround latitudes farthest poleward during summer and closest to the equator in winter, the upwelling width (distance between turnaround latitudes) generally shows much weaker seasonal variations. Furthermore, seasonality of the width is inconsistent between different products and the range between different products is of similar magnitude as seasonal variations. Large disagreement occurs especially during northern spring where NH turnaround latitudes exhibit very different seasonal transitions between different products: *e.g.*, MERRA and MERRA-2 are already close to their maximum poleward position during April, whereas most other products, including the CCMI MMM, only reach these positions during June-July. 20CR represents an outlier in that it lacks the correct NH seasonality.

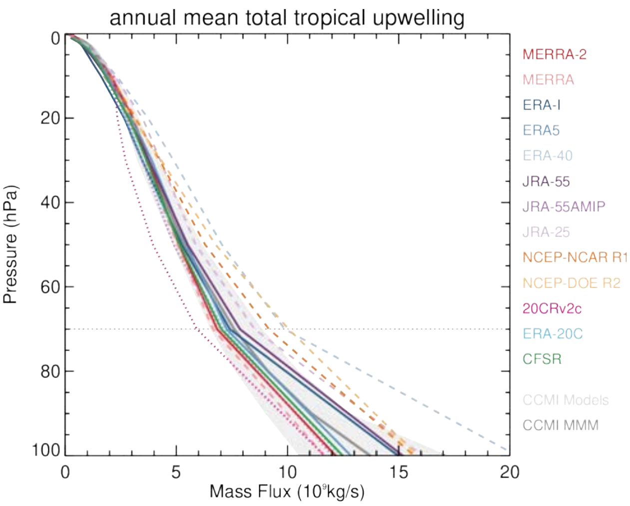


Figure 5.8: *Climatological annual mean total tropical upwelling as a function of pressure (between turnaround latitudes at each level). Line styles as* ***Figure 5.4****. Gray shading shows range of CCMI models with the thick gray line marking the MMM.*

Figure 5.6 showed that the total tropical upwelling at 70 hPa is spread over a fairly wide range between different products, with the older products showing much larger upwelling than the newer products and the climate-like runs (JRA-55AMIP, ERA-20C). This spread is even larger at 100 hPa but tends to decrease at higher altitudes (**Figure 5.8**). The vertical gradient of total tropical upwelling gives the net poleward mass flux from the upwelling region to extratropical latitudes. This gradient is generally much stronger in the older products (dashed lines) above 70 hPa. This means that leakage out of the tropical pipe is much stronger in these products. Between the newer products, JRA-55, ERA-Interim, and CFSR have similar leakage, while MERRA-2 shows a somewhat smaller leakage. The difference in upwelling between 100 - 70 hPa could be interpreted as an estimate for the net shallow branch divergence. However, it is important to note that because of the large gradient in turnaround latitudes between 100 - 70 hPa such an estimate includes both, poleward and downward mass fluxes. A large degree of variability arises due to the downward component, which also partly explains why there is generally a large spread in the mass flux gradient between 100 - 70 hPa in **Figure 5.8**. There is no clear change in this mass flux gradient from the older to the newer products. Part of the discrepancy also results from discrepancies in the turnaround latitudes (upwelling width) and their difference between 100 - 70 hPa. The upwelling width is much smaller at 100 hPa than at 70 hPa for most products. However, some products show only a small difference (*e.g.*, CFSR), whereas others show a very large difference (*e.g.*, ERA-Interim).

The upwelling strength at 100 hPa also shows a wide spread across the CCMs (gray shading in **Figure 5.8**). Presumably, differences in Hadley cell strength and vertical extent also play into the 100 hPa upwelling spread as the spread decreases markedly between 100 hPa and 70 hPa. The model diagnostics also contain the 90 hPa and 80 hPa levels, which demonstrates that the shallow branch divergence is likely weaker than diagnosed based on the 100 hPa and 70 hPa levels. ERA-Interim and JRA-55 are both close to the MMM throughout the lower stratosphere, although these two products show stronger upwelling than other recent reanalysis products (including ERA5) and the MMM at 100 hPa. For ERA-Interim and JRA-55 we had model level output available, which confirms that the shallow branch divergence based on the upper half of the 100 - 70 hPa layer is significantly weaker (about half) than that based on its lower half. Furthermore, interannual variability, similar to the spread across products, is much larger near 100 hPa. MERRA-2 and CFSR are both near the low end of model upwelling strengths. Higher up in the stratosphere (above ~ 10 hPa) the upwelling strength in the models is significantly larger than in the reanalyses, indicating a more isolated tropical pipe in the models.

To further quantify the mass flux within the shallow branch and to avoid the large sensitivity near the 100 hPa level, we consider the poleward residual flow at 70 hPa evaluated at the turnaround latitudes (**Figure 5.9**).

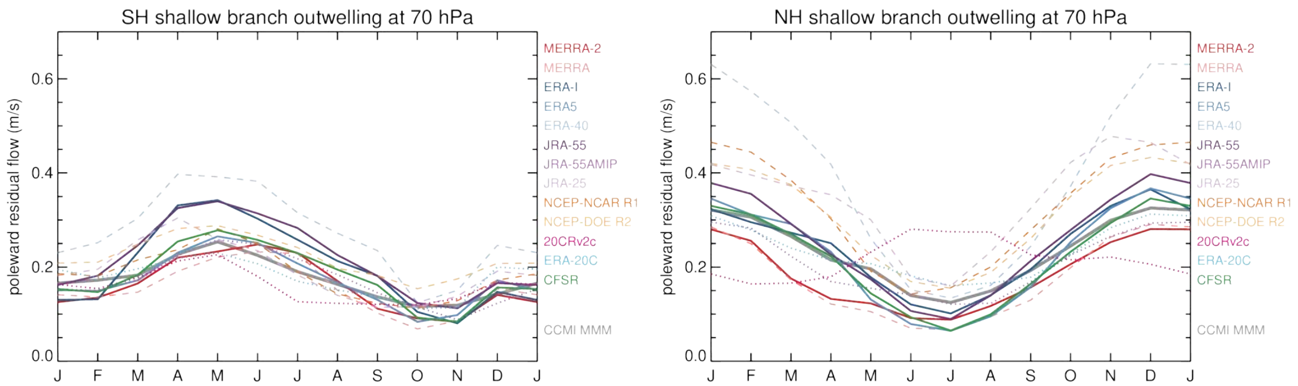


***Figure 5.9:** Climatological seasonal cycles of shallow branch tropical "outwelling" (poleward residual flow through the turnaround latitudes at 70 hPa) for the SH (left) and the NH (right). Line styles as **Figure 5.4**.*

The SH shallow branch poleward mass flux shows a maximum between fall and early winter (strongest in ERA-Interim, JRA-55, ERA-40) and a minimum in late spring. In the NH, outwelling is generally strongest during winter and weakest during summer, except for 20CR, which is the only product to show the opposite seasonality. The older products tend to show much stronger NH outwelling throughout the year compared to the newer products. Both MERRA products show consistently weaker NH outwelling during the cold season compared to other recent products. Taking both SH and NH together the outwelling diagnostics confirms that MERRA and MERRA-2 are at the low end of both lowermost stratospheric upwelling and shallow branch outwelling. Overall, shallow branch wave driving as quantified by EPFD (**Figure 5.4**) only explains part of the spread and variation in outwelling. This is likely because at certain latitudes gravity wave drag becomes important (not diagnosed here), while closer to the equator adjustments in relative vorticity become important (which modifies the relation between $\bar{v}^*$ and EPFD).

So far we have concentrated on diagnostics that directly quantify the (local) strength of the BDC. The RCTT diagnostic provides estimates of the integrated circulation strength. **Figure 5.10** shows the annual mean structure of RCTTs for the climatological REM. Since RCTTs result from backward trajectories with transit times up to several years near the poles, we discard the first few years of the time series. In order to still obtain a 30-year climatology we use the period 1986 - 2016 for the REM. The overall structure of the RCTTs agrees well with that from CCMs (*cf.*, *Birner and Bönisch*, 2011; *Dietmüller et al.*, 2018): a strong vertical gradient in the tropical pipe, which is similar to that of AoA, and a strong meridional gradient with strongly increasing RCTTs toward the poles in both hemispheres. Interestingly, both hemispheres have about equal RCTT structures and if anything the NH shows larger RCTT values near the pole in the lowermost stratosphere, perhaps due to the fact that the NH circulation reaches deeper into the upper stratosphere and the mesosphere (see **Figure 5.2**). The double peak in tropical upwelling shows up as a double peak in RCTTs with smallest values near 20° N/S and a local maximum near the equator (at a given level between the tropopause and ~ 10 hPa).

Figure 5.11 shows climatological annual mean RCTT structures and differences from the REM for individual reanalyses, comparing the more recent products to the older products. Overall, the more recent products tend to be much more consistent compared to the older products. ERA-40 and JRA-25 show much smaller RCTTs compared to the REM, indicating that their BDC is too strong. MERRA's tropical upwelling is biased, primarily because of local downwelling over the equator (see **Figure 5.5**), and this leads to a large positive bias in RCTTs over the equator. Of the more recent products, JRA-55 shows the smallest RCTTs, consistent with strongest tropical upwelling (*cf.*, **Figure 5.8**). JRA-55 is also the only product that does not exhibit the double peak in tropical upwelling, and likewise in RCTTs in the tropics, with a local maximum near the equator. ERA-Interim tends to show the largest RCTTs, except for the NH mid-latitude lower stratosphere (NH shallow branch). ERA-Interim's RCTTs show a pronounced hemispheric asymmetry consistent with a stronger shallow circulation branch in the NH compared to the SH.

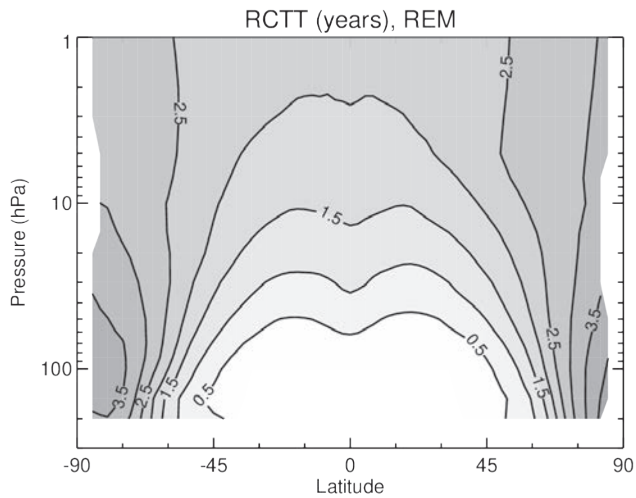
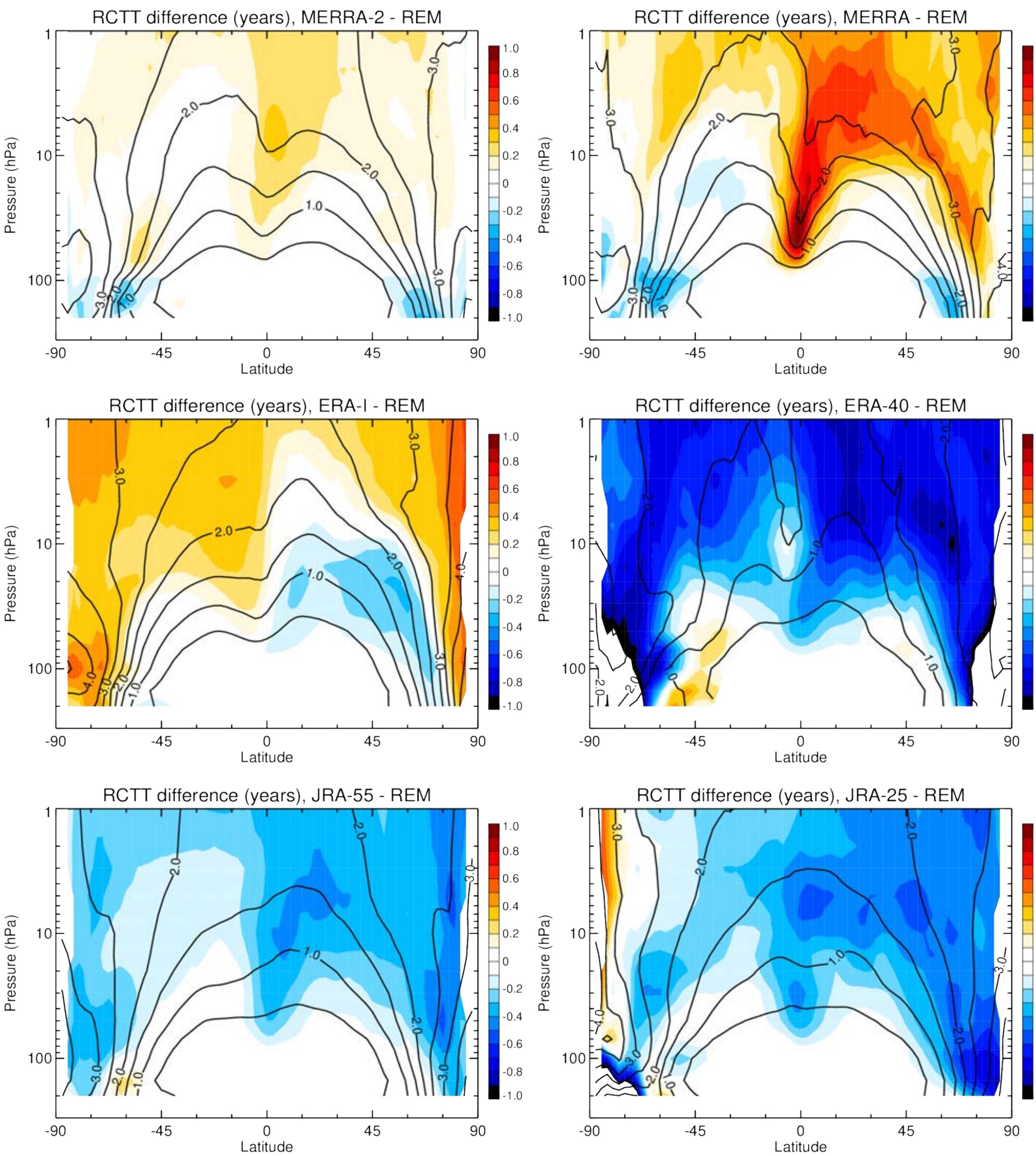


***Figure 5.10:** Climatological (1986 - 2016) annual mean residual circulation transit time (RCTT, in years) distribution for the multi-reanalysis mean (REM).*



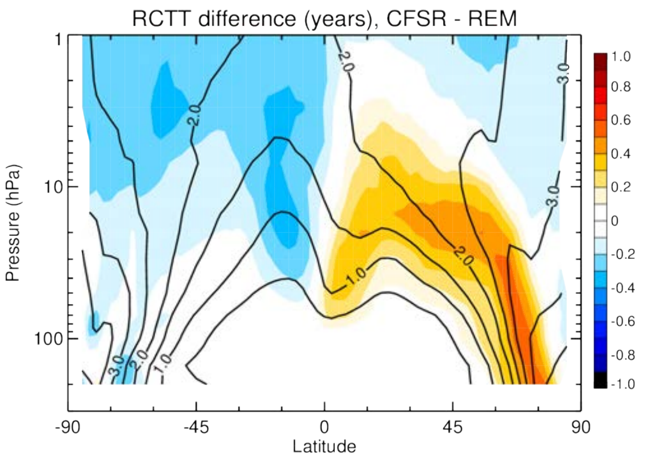


Figure 5.11: *Climatological annual mean residual circulation transit time (RCTT, in years, black contours) distribution for the recent reanalysis products (eft column, period 1986 - 2016) and older products (ight column, end years differ: 2013 for MERRA, 2000 for ERA-40, 2012 for JRA-25). The respective differences from the REM are shown in colors.*

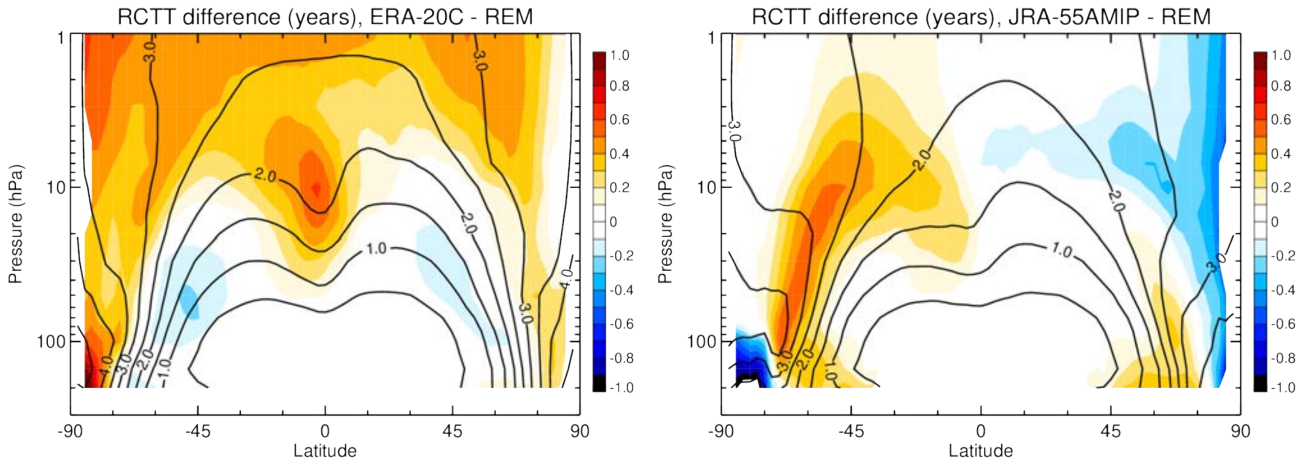


Figure 5.12: *As **Figure 5.11** but for ERA-20C (left, end year is 2010) and JRA-55AMIP (right, end year is 2012).*

This asymmetry disappears in ERA5, which agrees well with the REM overall, although RCTTs are still higher in the upper stratosphere (not shown, but climatological contours are included below in **Figure 5.17**). MERRA-2 shows a similar structure compared to its older version, but with smaller RCTTs, especially over the equator. CFSR shows smaller RCTTs throughout much of the SH, but larger RCTTs in the NH, especially along the deep circulation branch. CFSR exhibits a hemispheric asymmetry that is opposite of that in ERA-Interim.

RCTTs are overall similar between ERA-20C and ERA-Interim (**Figure 5.12**), which are based on a similar underlying model, although the strong hemispheric asymmetry in ERA-I is not present in ERA-20C. The free-running model version of JRA-55 does show an asymmetry with smaller RCTTs (stronger circulation) in the NH. Differences to the REM exist in the SH mid-latitudes and NH high latitudes.

5.5.1.2 Tropical upwelling trends

Even though some observational evidence for a strengthening of the BDC exists, modern reanalyses do not consistently show such strengthening. Specifically, ERA-Interim shows inconsistent trends compared to other reanalysis products, depending on the upwelling measure used (*Abalos et al.*, 2015).

Figure 5.13 shows that interannual variability in tropical upwelling at 70 hPa is large and likely spurious in some of the older reanalysis products, such as ERA-40 (perhaps due to older data assimilation systems). Corresponding trends are therefore not trustworthy. This variability is reduced and more consistent (see below) among the more recent reanalysis products. Furthermore, these more recent products lie within the range of CCMs with JRA-55 closely following the MMM, whereas MERRA-2 and CFSR consistently lie near the lower edge of model time series. ERA-Interim, which is the only recent product that shows a negative trend, initially closely follows the MMM and JRA-55, but from the late 1990's onward more closely follows the other three recent products. ERA5 shows overall similar variability to ERA-Interim but with consistently smaller upwelling values between ~1980 - 2005 transitioning to larger upwelling values from 2006 onward. ERA-20C and JRA -55AMIP also show a similar time series to those models with weaker overall upwelling. 20CR is generally biased low compared to all other products.

Visually, JRA-55, MERRA-2, and CFSR all show positive trends in tropical upwelling indicating a strengthening of the BDC. These trends are quantified in **Table 5.1** and for these three products are in the 2 - 3 %/decade range. ERA-Interim, on the other hand, shows a negative upwelling trend of the same order of magnitude (*cf.*, *Abalos et al.*, 2015), indicating a weakening of the BDC in this product. ERA5 shows a weak negative trend that is, however, not statistically significant.

To compare the reanalysis trends to those from CCMs we also calculated associated upwelling trends (see **Table 5.2**), but in this case for the longer period of 1960 - 2009 (the common period of 1980 - 2009 between the CCMs and the recent reanalyses is marginally short to obtain robust trends, *cf.*, *Hardiman et al.*, 2017).

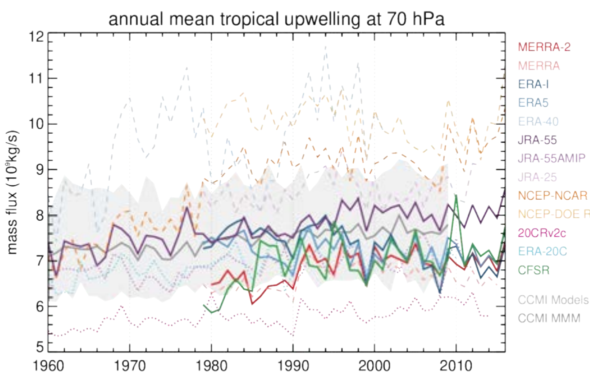


Figure 5.13: *Time series of annual mean tropical upwelling mass flux at 70 hPa (between turnaround latitudes). Line styles as **Figure 5.4**. The gray shading denotes the range of CCMI models with the multi-model mean shown as thick gray line.*

***Table 5.1:** 1980-2016 trends (in %/decade) of total tropical upwelling at 70 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude.*

MERRA-2	+**2.5**±1.3
ERA-I	−**3.4**±1.4
ERA5	−0.7±1.3
JRA-55	+**2.3**±0.9
CFSR	+**3.4**±2.0

The MMM trend is significantly smaller than those based on reanalyses, but some individual models (*e.g.*, WACCM) reach a similar upwelling trend of ~2%/decade. ERA-20C also exhibits a trend of ~2 %/decade over this time period, with JRA-55 (the only product, which has both examined time periods available) producing a similar trend than over the 1980-2016 period (see above). The free-running version of JRA-55 exhibits a trend near the lower end of, but consistent with, CCM trends.

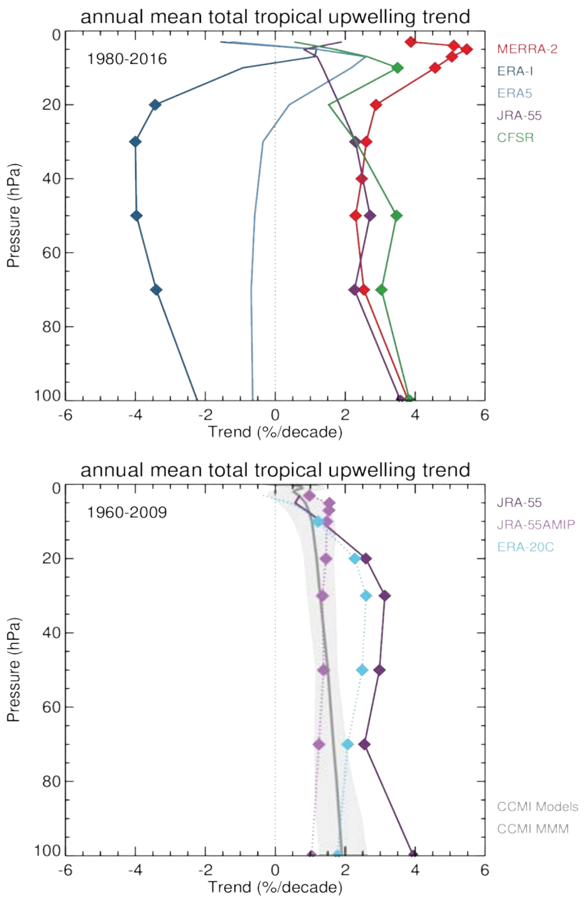


***Figure 5.14:** Trends in annual mean total tropical upwelling as a function of pressure (between turnaround latitudes at each level). Top: for the more recent reanalysis products and the period 1980-2016. Bottom: comparing climate models with reanalyses and other products for the period 1960-2009 (gray shading shows range of CCMI model trends with the thick gray line marking the MMM). Symbols indicate trends that are statistically significantly different from zero (based on 2σ uncertainty). For the MMM line all levels have statistically significant trends.*

***Table 5.2:** 1960-2009 trends (in %/decade) of total tropical upwelling at 70 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude. Note: individual model trends range from +1.1±0.5 (CMAM) to +2.1±0.7 (WACCM).*

CCMI MMM	+**1.7**±0.4
JRA-55AMIP	+**1.2**±0.8
JRA-55	+**2.5**±0.6
ERA-20C	+**2.1**±0.8

The trends in 70 hPa tropical upwelling are overall consistent with those at other stratospheric levels as shown in **Figure 5.14**. Specifically, MERRA-2, JRA-55, and CFSR all show mostly consistent, statistically significant positive upwelling trends for the period 1980-2016 over much of the stratosphere. These trends are somewhat stronger at 100 hPa (between +3.5-4%/decade), vertically coherent in the +2-3%/decade range up to 20 hPa, above which the different reanalysis products disagree about the trend. In contrast, ERA-Interim shows negative upwelling trends that are statistically significant between 70-20 hPa. These results are consistent with, and serve as an update of, the results presented in (*Abalos et al.*, 2015) based on a slightly shorter time period. The new ERA5 product shows small negative trends between 100-30 hPa, although none of them are statistically significant in the analysed pressure range. We again also consider the period 1960-2009 to compare to CCMI results (**Figure 5.14**, bottom). The MMM shows upwelling trends between +1-2%/decade throughout the stratosphere, which slightly decrease with height. JRA-55AMIP's trend, on the other hand, is generally within the range of CCMI trends, except for at 100 hPa. JRA-55's trends are much larger in the lower stratosphere, well outside the range of CCMI trends. ERA-20C lies between the two JRA products.

The disagreement in overall magnitudes and trends between even the recent reanalysis products, raises questions about their ability to capture long-term climate variations. A perhaps less stringent test is to examine the interannual variability of the different products. **Table 5.3** reveals that interannual variability is reasonably well correlated between ERA-Interim, ERA5, JRA-55, and MERRA-2, but not so much between CFSR and these products.

***Table 5.3:** Correlation coefficients for interannual variability of total tropical upwelling at 70 hPa between recent reanalysis products. Time series have been detrended before calculating correlations. Bolded values indicate statistical significance at the 95% confidence interval.*

	ERA-I	MERRA-2	JRA-55	CFSR
ERA-5	**0.75**	**0.66**	**0.65**	0.21
ERA-I	-	**0.67**	**0.70**	**0.50**
MERRA-2	-	-	**0.82**	0.14
JRA-55	-	-	-	0.25

Table 5.4: *1980-2016 trends (in %/decade) of total shallow branch outwelling at 70 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude.*

MERRA-2	+ **7.2** ± 2.3
ERA-I	+ 2.2 ± 2.3
ERA5	−1.3 ± 1.9
JRA-55	+ **3.2** ± 1.4
CFSR	+ 1.1 ± 2.3

5.5.1.3 Tropical outwelling and RCTT trends

Tropical upwelling at 70 hPa and above primarily measures the deep branch of the BDC. As before we use the poleward residual flow through the turnaround latitudes at 70 hPa as a measure of the shallow branch outwelling; its time series of the combined NH+SH outwelling are shown in **Figure 5.15**. Similar to lower stratospheric upwelling strength, net tropical outwelling is significantly weaker in the more recent reanalysis products compared to their predecessors (except for MERRA). The more recent products agree in their overall strength with the CCMI models, as do the other reanalysis-related products (JRA-55AMIP, ERA-20C, 20CR). Visually, MERRA-2 exhibits a strong increasing trend between 1980 - 2000, and JRA-55 exhibits a long-term trend over the entire depicted record. With the exception of ERA5, all recent reanalysis products show positive trends for the period 1980 - 2016 (see **Table 5.4**), although this trend is not statistically significant in ERA-Interim and CFSR. The weak negative trend in ERA5 is likewise not statistically significant.

Over the longer period from 1960 - 2009 JRA-55 exhibits a consistent trend in net shallow branch outwelling with the shorter period (both between 3 - 4 %/decade, *cf.*, **Table 5.5**). While some CCMI models almost reach this strong accelerating trend (WACCM), the MMM trend is somewhat weaker and JRA-55AMIP's trend is at lower end of CCMI model trends. ERA-20C exhibits a positive outwelling trend within the range of CCMI model trends.

Similar to the 70 hPa upwelling time series, we have also analyzed interannual variability in net shallow branch outwelling (see correlation coefficients listed in **Table 5.6**). Co-variability in this case is weak across many recent reanalysis's. The highest correlation coefficient is 0.68 between

Table 5.5: *1960 - 2009 trends (in %/decade) of total shallow branch outwelling at 70 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude. Note: individual model trends range from + 0.6 ± 0.6 (CMAM) to +2.7±1.2 (WACCM).*

CCMI MMM	+ **1.9** ± 0.5
JRA-55AMIP	+ 0.8 ± 0.8
JRA-55	+ **3.7** ± 0.9
ERA-20C	+ **1.2** ± 0.7

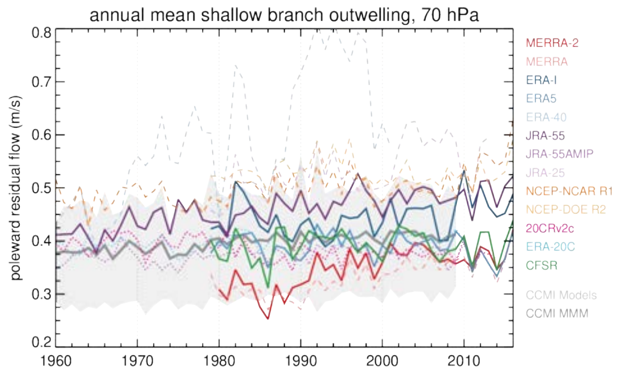


Figure 5.15: *Time series of annual mean tropical outwelling at 70 hPa (total poleward residual flow through the turnaround latitudes). Line styles as* ***Figure 5.4****. The gray shading denotes the range of CCMI models with the multi-model mean shown as thick gray line.*

MERRA-2 and ERA5. The low correlations in this case exist despite coherent variability in EP-flux divergence (not shown), indicating that unresolved processes and/or model biases are primarily responsible for the lack of co-variability in shallow branch outwelling.

We next examine time series of RCTTs. The 50 hPa (~ 20 km altitude) level is often used to compare AoA estimates (see *Section 5.5.2*). **Figure 5.16** shows the time series of annual global mean RCTTs from different products. Consistent with strongest upwelling JRA-55 shows smallest RCTTs that are steadily decreasing over time, consistent with a strengthening of the BDC. CFSR and MERRA-2 both show much larger interannual variations and a large negative trend in the 1980's and 1990's. ERA-Interim is closer to JRA-55 in the beginning of the record, but approaches CFSR and MERRA-2 toward the end of it, in the latter period showing similarly strong interannual variations. ERA5 tends to be more consistent with MERRA-2 and CFSR than with ERA-Interim in this metric, especially from the mid-1990's forward. CCMI models show a wide range of global mean RCTTs at 50 hPa, encompassing essentially all reanalysis products.

RCTT trends are examined for the period 1982 - 2016 (the first few years of the records need to be discarded because of the backward trajectory setup of the calculations), listed in **Table 5.7**.

Table 5.6: *Correlation coefficients for interannual variability of total shallow branch outwelling at 70 hPa between recent reanalysis products. Time series have been detrended before calculating correlations. Bolded values indicate statistical significance at the 95 % confidence interval.*

	ERA-I	MERRA-2	JRA-55	CFSR
ERA-5	0.23	**0.68**	**0.34**	**0.53**
ERA-I	-	**0.29**	0.26	**0.30**
MERRA-2	-	-	**0.43**	**0.55**
JRA-55	-	-	-	**0.52**

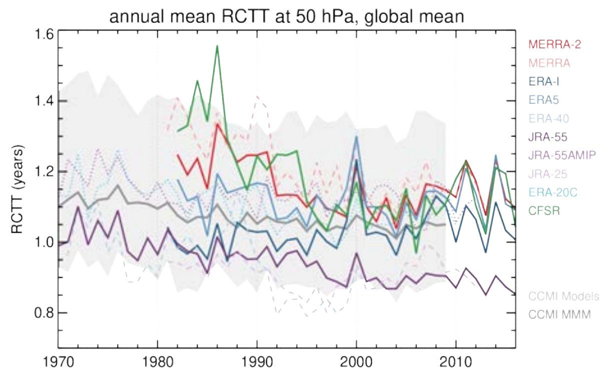


Figure 5.16: *Time series of annual gobal mean RCTTs at 50 hPa. Line styles as **Figure 5.4**. The gray shading denotes the range of CCMI models with the multi-model mean shown as thick gray line.*

Qualitatively, these RCTT trends are consistent with the respective upwelling trends at 70 hPa (*cf.*, **Table 5.1**): we find negative transit time trends indicating a strengthening of the BDC in MERRA-2, JRA-55, and CFSR, a positive trend indicating a weakening of the BDC in ERA-Interim, and a non-significant trend in ERA5. CFSR's trend (not shown), although formally statistically significant, has a large uncertainty due to the questionable interannual and decadal variability in the beginning of the record (see **Figure 5.16**); its trend values are therefore not included here or in other RCTT trend estimates. All recent reanalysis products indicate much weaker trends since the year 2000 (*cf.*, **Figure 5.16**), and all of them show a pronounced maximum in that year, reflecting the weaker upwelling values the year before (presumably due to the strong La Niña event in 1999). The difference in magnitude of BDC trends pre and post 2000 is consistent with recent arguments regarding the role of ozone depletion for BDC trends (*e.g.*, *Abalos et al.*, 2019; *Polvani et al.*, 2019; *Garfinkel et al.*, 2017).

Over the longer period 1970-2009 JRA-55 shows an even stronger negative RCTT trend at 50 hPa (**Table 5.8**). This strong BDC acceleration is not found in the free-running version JRA-55AMIP, although this data set also shows a negative RCTT trend. Moreover, JRA-55AMIP is consistent with the MMM of the CCMI models. ERA-20C's corresponding trend falls somewhere in the middle of those trends.

MERRA-2, ERA-Interim, ERA5, and JRA-55 show reasonably strong interannual co-variability in global mean RCTTs at 50 hPa with correlation coefficients ranging between 0.53 (between MERRA- and ERA-Interim) and 0.85 (between ERA5 and ERA-Interim, *cf.*, **Table 5.9**). CFSR (not shown) variability agrees well with the other products in the latter part of the record (*cf.*, **Figure 5.16**).

Table 5.7: *1982-2016 trends (in %/decade) of 50 hPa global mean RCTTs with their 2σ uncertainties based on recent reanalysis products (see text for details). Bolded values indicate trends exceeding their 2σ uncertainty in magnitude. Note, CFSR is not included here because it shows questionable decadal variability.*

MERRA-2	− **2.9** ± 1.9
ERA-I	+ **2.1** ± 1.8
ERA5	−0.2 ± 1.8
JRA-55	− **3.6** ± 1.0

Table 5.8: *1970-2009 trends (in %/decade) of global mean RCTTs at 50 hPa with their 2σ uncertainties. Bolded values indicate trends exceeding their 2σ uncertainty in magnitude.*

CCMI MMM	− **2.2**±0.5
ERA-20C	− **3.0** ±1.2
JRA-55	−**4.2** ± 1.0
JRA-55AMIP	− **1.7**± 1.1

The latitude-pressure structure of individual products' RCTT trends are shown in **Figure 5.17**. Overall, MERRA-2 and JRA-55 show mainly negative trends, in some cases reaching - 20 %/decade (*e.g.*, MERRA-2 in the NH subtropical lower stratosphere). JRA-55 tends to show smallest climatological RCTTs in both hemispheres (black contours), whereas ERA-Interim shows largest RCTTs for the SH deep branch with MERRA-2 showing largest RCTTs for the NH deep branch. MERRA-2 shows a large negative trend in the first half of the record for the NH deep branch, which does not continue over the latter half of the record (not shown). ERA-Interim shows primarily weakly positive trends, except for in the lowermost mid-latitude stratosphere, consistent with a weakening of its deep branch but a strengthening of its shallow branch. A similar picture emerges with ECMWF's new product, ERA5, with perhaps a wider area of negative trends in the shallow branch. The shallow branch strengthening is fairly consistent across products, except for the free-running models (as above). The strengthening of the shallow branch seen in the RCTT trends appears to be the only robust trend that is consistent across all recent reanalysis products.

This robust strengthening trend of the shallow branch is confirmed for the longer period (1970-2009) in JRA-55 (**Figure 5.17**). In fact, these longer-term trends are generally larger in magnitude for JRA-55. However, a consistent trend across the available products for this longer period only exists for the SH deep branch. Interestingly, the MMM of the CCMI models and JRA-55AMIP show a positive trend in RCTTs for parts of both hemisphere's shallow branches (indicative of weakening), suggesting a robust mismatch between the CCMs and reanalyses.

Table 5.9: *Correlation coefficients for interannual variability between 1982-2016 of global mean RCTTs at 50 hPa between recent reanalysis products. Time series have been detrended before calculating correlations. CFSR is not included (see text for details). Bolded values indicate statistical significance at the 95 % confidence interval.*

	ERA-I	**MERRA-2**	**JRA-55**
ERA-5	**0.85**	**0.77**	**0.70**
ERA-I	-	**0.53**	**0.64**
MERRA-2	-	-	**0.71**

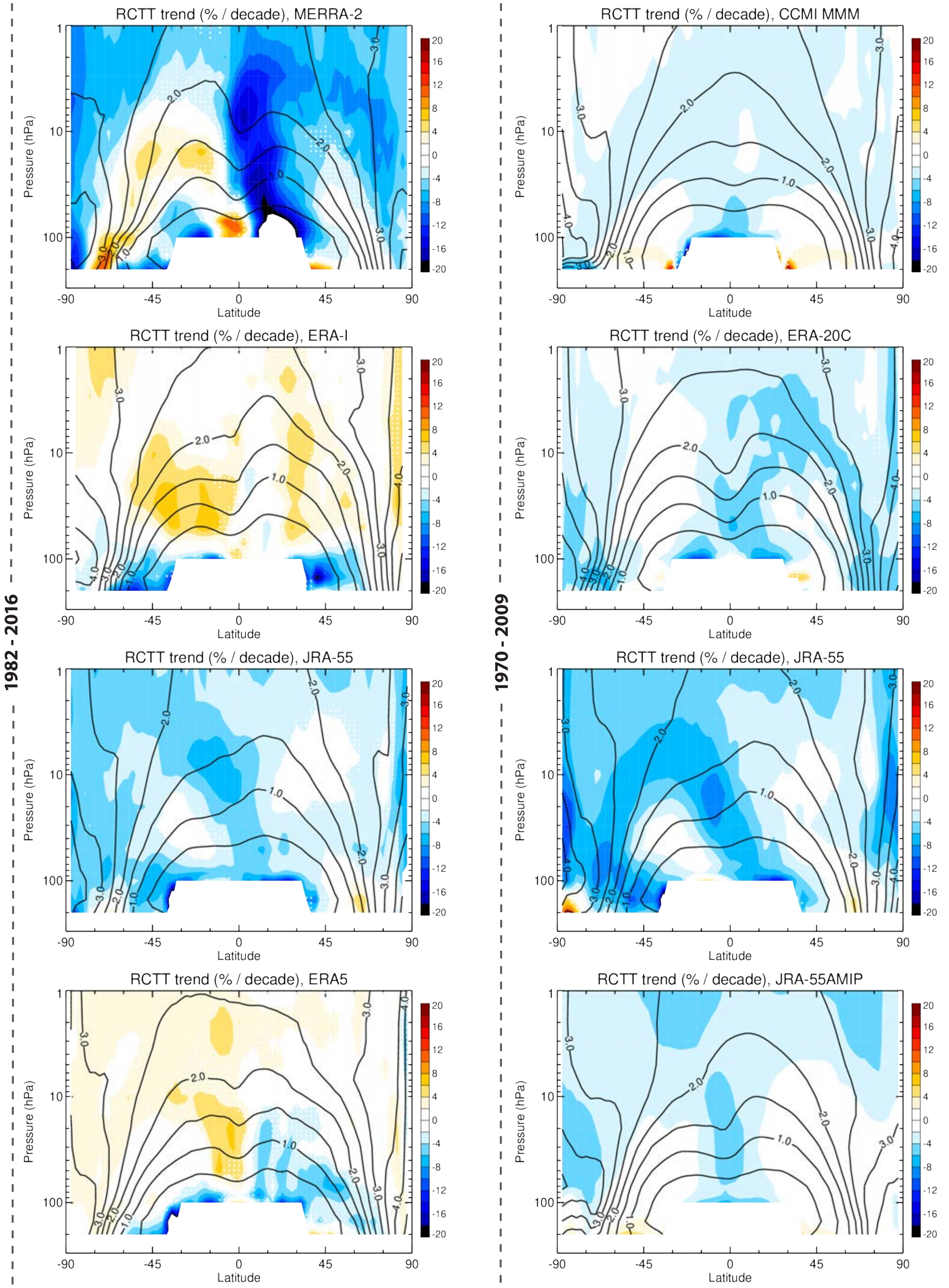


Figure 5.17: *Trends in annual mean RCTT as a function of latitude and pressure (color shading, in %/decade) for the period 1982 - 2016 in the recent reanalysis products (eft column, note that ERA5 is included instead of CFSR, see text), as well as the period 1970 - 2009 in the products shown (ight column). Each product's climatology over the respective period is depicted as black contours.*

For the deep branches the MMM of the CCMI models shows robust negative trends in RCTTs, indicating opposing trends between parts of the shallow and deep branches in those models (at least by this metric of the BDC).

5.5.2 Results from transport tracers simulations

5.5.2.1 Heating rates

Heating rates [3] from reanalysis are not only a stratospheric circulation diagnostic in itself, but they are also one of the fields used to drive some of the offline models employed for our tracer transport simulations. The CLaMS and TRACZILLA offline models use heating rates for their advection schemes. TOMCAT/SLIMCAT also uses diabatic heating rates when run in "SLIMCAT" mode. Therefore, a comparison of diabatic heating rates in the different reanalyses datasets contributes to identify differences in stratospheric transport in the considered simulations.

Figure 5.18 shows the annual cycle of the diabatic heating rate, $\dot{\theta}$ (K/day), in isentropic coordinates, at the tropical UTLS based on daily data covering 1980-2010; data come from the ERA-Interim, JRA-55, MERRA-2 and CFSR reanalyses.

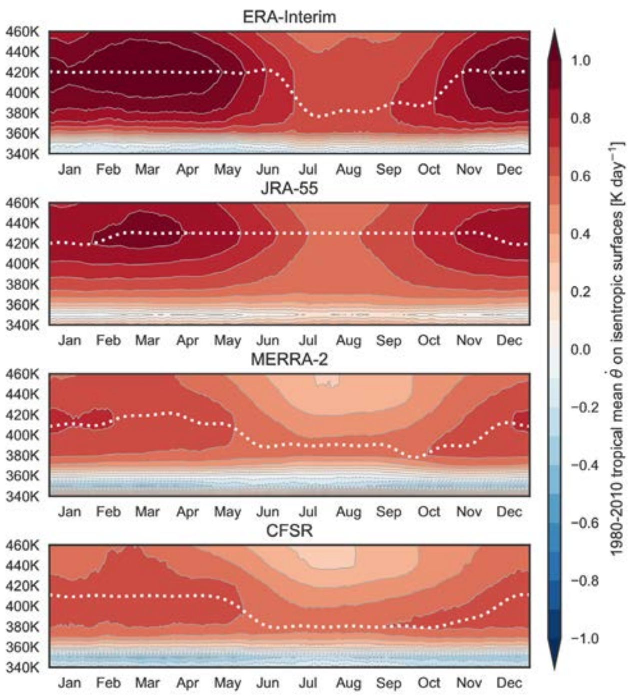


***Figure 5.18:** Annual cycle of the tropical mean (30°S-30°N) of the diabatic heating rate field (K/day) on isentropic surfaces, from 340-460K. The field has been averaged over the period 1980-2010 for the ERA-Interim (top row), JRA-55 (second row), MERRA-2 (third row) and CFSR (bottom panel) reanalyses. The white dotted line in each panel shows the annual cycle of the local maximum within the lower stratosphere for the corresponding reanalysis.*

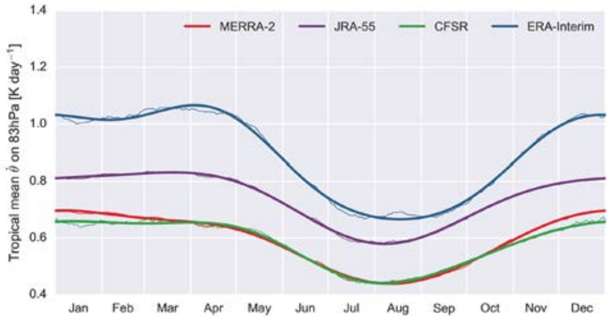


***Figure 5.19:** Annual cycle of the tropical mean (30°S-30°N) of the diabatic heating rate field (K/day) at the 83hPa level from ERA-Interim (blue), JRA-55 (purple), MERRA-2 (red) and CFSR (green). Annual cycles are based on day-of-year means (thin lines) and smoothed using FFT-based low pass filters (thick lines).*

The field has been averaged over the broader tropical region (30°S-30°N). The white dotted line in each panel shows the annual cycle of the local maximum within the lower stratosphere for the corresponding reanalysis. Details on the way these fields have been calculated can be found in *Wright and Fueglistaler* (2013) and *Dessler et al.* (2014), and a detailed discussion on this fields in *Martineau et al.* (2018). The figure shows that the strongest annual cycle and the largest values for this field correspond to the ERA-Interim reanalysis, the structure of the cycle is similar for JRA-55 although with weaker and smaller values, especially over the months with maximum values. MERRA-2 and CFSR, show smaller $\dot{\theta}$ values and a weaker annual cycle than the other two reanalyses for the whole vertical profile. The corresponding annual cycle for the 83hPa level is shown in **Figure 5.19**, showing both the annual cycles based on day-of-year means (thin lines) and those smoothed using FFT-based low pass filters. As in **Figure 5.18**, there are significant differences among reanalyses, with ERA-Interim showing the highest values and the most pronounced seasonal cycle, and CFSR showing the lowest values and least pronounced seasonal cycle. MERRA-2 is surprisingly very similar to CFSR and JRA-55 values are in between the CFSR and the ERA-Interim ones, although the amplitude of the seasonal cycle in JRA-55 is as low as for CFSR and MERRA2.

The way the $\dot{\theta}$ field has evolved with time for the different reanalyses is shown in **Figure 5.20** as the time series of the tropical mean (30°S-30°N) of the diabatic heating rate field (K/day) at the 83hPa level for the diagnosed field (darker solid lines) and for the forecast field (lighter solid lines). The corresponding linear trends (K/day per decade) are also shown in this figure. All reanalyses except ERA-Interim show considerable differences between the diagnosed and the forecast fields, and even for ERA-Interim the corresponding linear trends are different for both sets. Figures shown here evidence the large differences that exist between reanalyses for the annual cycle of the diabatic heating rates. There are also large differences between reanalyses regarding the diurnal cycle of diabatic heating rates in the UTLS in convective regions (*Tegtmeier et al.*, 2020).

[3] See the footnote on diabatic heating rates in reanalyses in *Chapter 12, Section 12.1.3.*

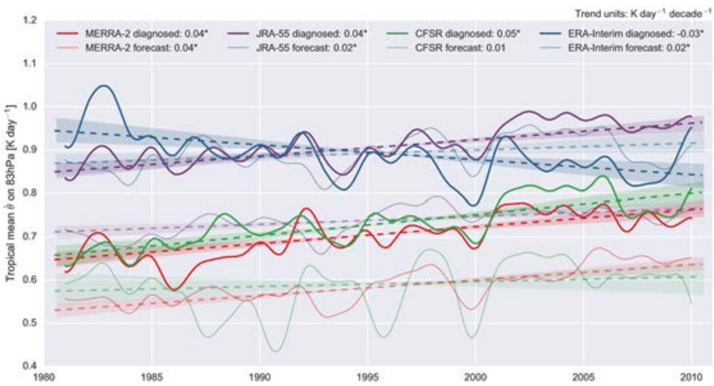


Figure 5.20: *Time series of the tropical mean (30 ° S-30 ° N) of the diabatic heating rate field (K/day) at the 83 hPa level for the diagnosed field (darker solid lines) and for the forecast field (lighter solid lines). The corresponding linear trends (K/day·decade) are also shown (dashed lines). Time series have been low-pass filtered via a 24-month rolling mean using a Hamming window. Trends are calculated from annual mean values via the Theil-Sen estimator (95 % confidence intervals estimated via bootstrapping). The best-estimate trend values are shown in the legend. Reanalyses shown and colour scale as in* ***Figure 5.19****.*

5.5.2.2 Mean age-of-air from observations

In this section we discuss mean AoA results obtained from recent observation based studies and datasets described in *Section 5.4*. We use these in later sections to compare our results from the offline models driven by the different reanalyses.

"Standard" observations for model intercomparisons

Observation-based mean AoA is derived from concentration measurements of long-lived tracers with an approximately linear increase at the surface, such as CO_2 or SF_6. For CO_2 one needs to take into account the surface seasonal cycle, which can still affect derived AoA values in the lower stratosphere, while SF_6 is affected by the mesospheric sink which makes derived values in the upper stratosphere biased towards older values. Multidecadal datasets were compiled from balloon soundings or aircraft flights using both CO_2 and SF_6 measurements (*e.g., Ray et al.*, 2014; *Andrews et al.*, 2001; *Ray et al.*, 1999; *Boering et al.*, 1996; *Elkins et al.*, 1996; *Harnisch et al.*, 1996). These observational datasets have been used for model validation in numerous studies and SPARC model intercomparison activities (*e.g., Ploeger et al.*, 2019; *Chabrillat et al.*, 2018; *Dietmüller et al.*, 2018; *Monge-Sanz et al.*, 2012; 2007; *Eyring et al.*, 2006; *Waugh and Hall*, 2002).

British Antarctic Survey (BAS) polar measurements

Cook and Roscoe (2009, 2012) used stratospheric measurements of polar summer NO_2 to derive trends in the BDC. Stratospheric column of NO_y over the Antarctic station of Rothera (67 ° S) were derived from measurements of NO_2 taken during 1990 - 2007; a photochemical model and observed ozone and temperature profiles were used to determine column values (**Figure 9** in *Cook and Roscoe*, 2009). Years 1991 and 1992 were excluded from their calculations because of the large amounts of volcanic aerosols from the Pinatubo eruption still present in the stratosphere.

A reconstruction from a multiple regression of these NO_y values, in which the solar cycle, the QBO, the ENSO, and a linear term are considered, is shown in **Figure 5.21**. The ratio of NO_y column to the BDC strength can be calculated following the methods in *Cook and Roscoe* (2009); they found a trend value in NO_y of – 1.1 ± 3.5 %/decade which translated into an increase in BDC of 1.4 ± 3.5 %/decade. Therefore, from the studies of *Cook and Roscoe* (2009, 2012), the conclusion was that the BDC exhibited no significant trend over the summer Antarctic for the period considered. However, they also pointed out the existence of an unexplained cycle with an amplitude of at least 15 % and a period longer than 17 years, with a minimum in BDC strength (maximum in NO_y values) around year 2000.

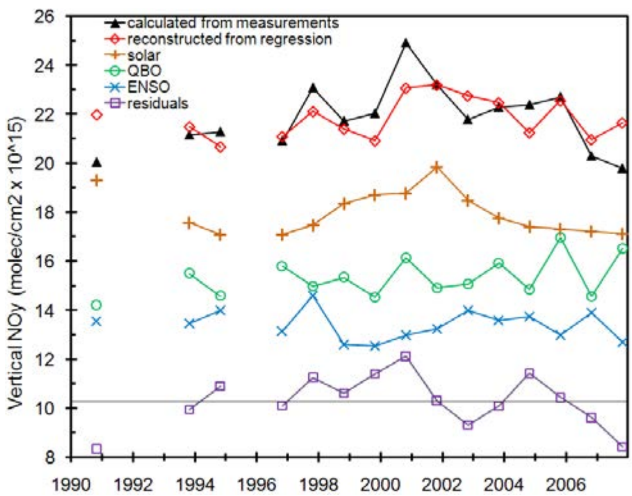


Figure 5.21: *Reconstructions from a multiple regression of the* NO_y *values from* ***Fig. 9*** *in Cook and Roscoe (2009) against solar cycle, QBO, ENSO, and a linear term (as shown in the key legend). The measurements and total reconstruction have the scale shown, but the separate terms in the reconstruction have been arbitrarily displaced for clarity. The linear trend term is the thin black line with the residuals centred on it; note that the residuals have been offset upwards and the trend is negligible. The conclusion was that the trend in speed of BDC was – 1.1 ± 3.5 %/decade. The speed also had a large unexplained cycle of amplitude > 15 % and period > 17 years. See Cook and Roscoe (2009, 2012) and main text for further details.*

Balloon observations over NH midlatitudes

The balloon-borne observations used in *Engel et al.* (2009) were taken in a region between 24 km and 35 km where the vertical gradient in mean age at NH midlatitudes was found to be very small, leading to little variability in this region. The balloon data were limited to a total of 28 flights and showed a positive trend of 0.24 years per decade for this region, which was, however, estimated to be non-significant.

These results have been recently updated by *Engel et al.* 2017), **Figure 5.22**, adding balloon-borne AirCore observations, to extend the previous data series so that it now covers more than 40 years. The corresponding updated trend is smaller than that from *Engel et al.* (2009) with a small positive value of 0.15 ± 0.18 years per decade. Although the trend is statistically non-significant, these observations are still in contrast to the strong negative trends in mean AoA derived from most climate model calculations (*e.g., Waugh*, 2009). The potential of the relatively cheap measurements of the AirCore instrument in *Engel et al.* (2017) makes them a promising way to keep monitoring AoA in the LS regions of interest.

An acknowledged caveat in the observations in the studies by *Engel et al.* (2009; 2017) and *Cook and Roscoe* (2009, 2012) is that the corresponding datasets covered only limited regions (midlatitudes or the Antarctic) and were also sparse in time. Global measurements from satellites are needed to provide a more complete picture of the BDC strength. For the period 2002 - 2012 AoA values derived from the SF_6 measurements taken by the MIPAS instrument on board Envisat are available (see next subsection).

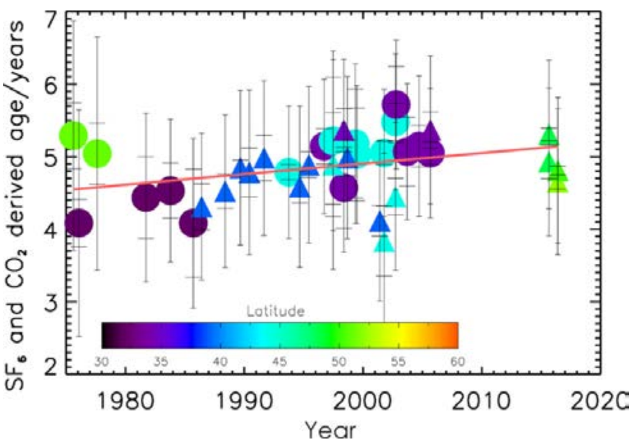


Figure 5.22: *Time series of mean age derived from balloon observations. Data before 2010 are from Engel et al. (2009), while data from 2015 and 2016 are derived from the Air-Core measurements in Engel et al. (2017). Each data point represents the average mean AoA between 5 - 30 hPa. Error bars represent the variability (inner error bars), and the uncertainty (outer error bars) as discussed in Engel et al. (2009). A non-significant trend of 0.15 (± 0.18) years per decade is derived from these observations. Figure from Engel et al. (2017).*

In this Chapter we have extensively used this MIPAS dataset to validate our offline model simulations obtained with the different reanalyses.

MIPAS mean AoA dataset

The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) was an instrument on board of the Envisat satellite, measuring the mid infrared emission of the atmosphere against the space background. The measurements were done in limb scanning mode covering an altitude range of cloud top (or about 6 km in cloud-free cases) to about 72 km. The emission signatures of molecules in the atmosphere were used to retrieve the spatial distribution of up to 30 different trace gases and temperature with good global coverage from pole to pole, also during (polar) night. The mission extended from July 2002 to April 2012.

Information on the stratospheric mean AoA is obtained from the spatio-temporal distribution of the SF_6 tracer, measured by MIPAS with a vertical resolution of 4 km to 6 km and a single profile precision of about 10 - 20 %. Although the single profile precision is rather low, the huge number of profiles measured (more than 2 million profiles over the MIPAS mission lifetime) provided very valuable information on AoA from zonal mean distributions. The SF_6 distributions were retrieved from the upper troposphere up to about 50 km. Above 35 km, the systematic errors become rather large, and the vertical resolution deteriorates; for this reason quantitative analysis of SF_6 and AoA above 35 km is not recommended.

AoA has been derived from SF_6 zonal monthly averages using a surface SF_6 reference curve from the NOAA/GMD network. The combined global mean surface SF_6 from NOAA (**http://www.esrl.noaa.gov/gmd/hats/combined/SF6.html**) was used to calculate the time lag between the time of stratospheric measurement and the time when the same SF_6 amount had been observed on the surface. It was confirmed that the MIPAS-measured SF_6 mixing ratios of the free tropical troposphere and their trends agree well with the surface SF_6 abundances, and a small bias correction was applied before using the surface reference. For a strictly linearly growing tracer, this time lag is identical with the first moment of the AoA spectrum, the mean age of stratospheric air. Since SF_6 is not strictly linearly growing, the AoA calculation was corrected by convolving the SF_6 surface time series with ideal AoA spectra; more details are given in *Stiller et al.* (2012). In this way, monthly-mean zonal-means of mean AoA were provided for 10° latitude bands and 1 - to - 2 km altitude steps, for the period July 2002 to March 2012.

In **Figure 5.23** the mean AoA derived from MIPAS observations at 20 km altitude is compared to earlier airborne AoA observations taken during a large number of aircraft and balloon campaigns (*Waugh and Hall*, 2002; *Andrews et al.*, 2001; *Hall et al.*, 1999; *Ray et al.*, 1999).

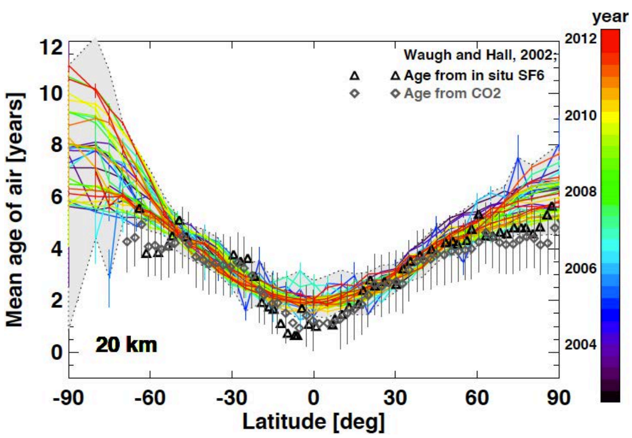


Figure 5.23: *Comparison of MIPAS-derived AoA (coloured lines) as a function of latitude at 20 km of altitude with airborne AoA measurements on basis of SF_6 (triangles) and CO_2 (diamonds). Shown for MIPAS is the monthly zonal mean for every third month between July 2002 and March 2012. The grey shaded area is the range of variability of all MIPAS zonal mean data. The vertical error bars are measurement uncertainties for CO_2-derived airborne AoA measurements. Airborne measurements are shown after Waugh and Hall (2002) and Hall et al. (1999). Figure from Haenel et al. (2015, Supplement).*

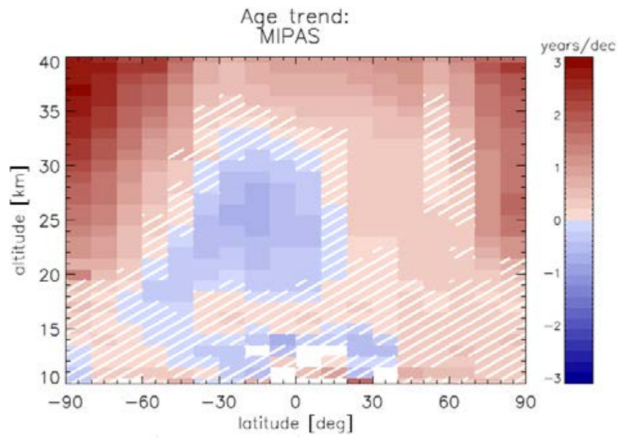


Figure 5.24: *Linear trend of AoA from a multivariate linear regression including seasonal variation and QBO effects for the period 2002 to 2012 from MIPAS data. Hatching indicates bins where the trend is not statistically significant in terms of its 2-sigma uncertainty. Figure from Haenel et al. (2015).*

The MIPAS values as a function of latitude are displayed for every second month, showing the considerable temporal variability of AoA at 20 km. In the midlatitudes, the MIPAS-derived AoA agrees very well with the SF_6-derived AoA from the aircraft campaigns, while it is higher compared to the CO_2-derived AoA (but in the same range as the difference between the SF_6- and CO_2-derived AoA from the aircraft campaigns). At southern polar latitudes, the variation of MIPAS-derived AoA is very large, however, these latitudes were not covered by the aircraft measurements. At northern polar latitudes, the MIPAS derived AoA are older than the aircraft data, although the SF_6-derived aircraft data still fall into the variability range of MIPAS. The most significant difference between MIPAS-derived and aircraft-derived AoA, however, is in the tropics. Here MIPAS-derived AoA is considerably older and this leads to a lower meridional gradient in AoA. Since there are more than ten years in between the measurements of MIPAS and most of the aircraft data (*e.g.*, see **Table 1** in *Waugh and Hall*, 2002), it cannot be determined whether this discrepancy is due to a change in atmospheric behaviour, *e.g.*, stronger inmixing of extratropical air into the tropical pipe, or if it is an artefact in one of the two datasets. The reference use for SF_6 to calculate the AoA values may also play a role: for the aircraft data, SF_6 measurements at the tropical tropopause were used as reference, while for MIPAS the global mean SF_6 time series at the surface provided by NOAA Global Monitoring Laboratory (GML) has been used.

The MIPAS monthly zonal means of AoA have been studied with respect to their short-term variability (seasonal cycle, QBO impact) and their decadal linear trend (*Haenel et al.*, 2015; *Stiller et al.*, 2012). The corresponding MIPAS-derived linear trends for the latitude/altitude bins are provided in **Figure 5.24**. There are wide areas where the AoA trend is significantly different from zero. Interestingly, a positive AoA trend is found all over the stratospheric Northern extratropics. These trends in the Northern midlatitudinal middle stratosphere agree well with the 30-year trends derived by *Engel et al.* (2009) from balloon-borne cryogenic air sampling data (see **Figure 7** in *Haenel et al.*, 2015) and are significantly positive. In the Southern Hemisphere, the Northern tropics, and the UTLS in both hemispheres, negative AoA trends are found, in agreement with most climate model predictions (*e.g.*, *Waugh*, 2009). Both mean AoA values and the trends derived from SF_6 can be affected by the mesospheric SF_6 loss (*e.g.*, *Reddmann et al.*, 2001; *Ravishankara et al.*, 1993). The SF_6-depleted air subsiding from the mesosphere in polar winters is thus misinterpreted as very old air; similarly, as the amount of SF_6 depletion scales with the absolute SF6 abundance that is increasing, the mesospheric loss leads to an apparent positive age trend. Both issues affect not only the polar winter air, but also the rest of the stratosphere, to the extent the previously mesospheric air is mixed into lower latitudes and altitudes after the polar vortex break-down. There have only been a few studies providing some estimation of the impact of the SF_6 mesospheric sink on absolute AoA and AoA trends in the stratosphere; see related discussion with the KASIMA results later in *Section 5.5.2.3*, and also in *Stiller et al.* (2012) and *Kovacs et al.* (2017).

5.5.2.3 Mean AoA from offline models

To assess the reanalyses' ability to reproduce atmospheric tracers distribution and evolution, we have used a set of different offline models (*Section 5.3*). By using several offline models we obtain a certain spread in the performance of the different reanalyses, which helps to overcome the sensitivity that a particular reanalysis may have to a particular offline model configuration.

The use of different types of offline models (*e.g.*, kinematic and diabatic models, in which the vertical motion is derived from the wind velocity or diabatic heating rate fields respectively) also allows us to narrow the source of the transport differences to particular fields in the reanalyses.

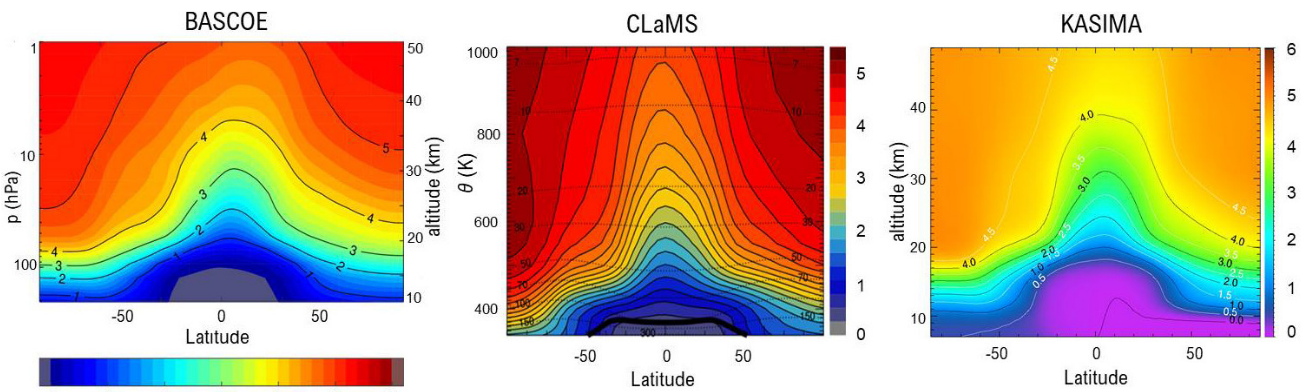
We have not included ERA-40 data (*Uppala et al.*, 2005) in our comparisons as this reanalysis was already shown to provide unrealistically fast stratospheric transport by numerous studies (*e.g.*, *Chipperfield*, 2006; *Scheele et al.*, 2005; *Meijer et al.*, 2004; *van Noije et al.*, 2004). The TOMCAT CTM was the first one to show the improvements obtained in the stratospheric transport representation with the newer generation ERA-Interim reanalysis compared to the ERA-40 reanalysis (*Monge-Sanz et al.*, 2007).

As an overall comparison of the different participating offline models, we examine how they reproduce the mean AoA from simulations driven with the ERA-Interim reanalysis. **Figure 5.25** shows the cross section of mean AoA values obtained from all the offline models driven by meteorological fields from this reanalysis, averaged over the period 1989 - 2010. All models reproduce similar distributions although some differences are also seen: Eulerian kinematic models (BASCOE, KASIMA and TOMCAT) simulate overall younger mean AoA values than the diabatic Lagrangian models (CLaMS and TRACZILLA). The shape of the tracer isolines is narrower over the tropics for the Lagrangian models, and the tropical peak shows a slight tilt towards South, while for the Eulerian models this tilt is less pronounced and goes northwards.

In **Figure 5.25**, the top right panel shows the annual-mean zonal-mean distribution of the mean AoA as simulated by the KASIMA CTM driven by ERA-Interim reanalysis fields for the period 1980 - 2010. It shows the typical bell form with maximum mean age values slightly older than 5 years; and the asymmetry with older mean age in southern polar latitudes, related to the persistent downwelling of old air from the mesosphere during the Antarctic winter polar vortex. This figure shows the AoA distribution obtained with the ideal tracer T1 (see *Section 5.3*) and it can therefore be compared to the analogous distribution obtained with the TOMCAT CTM.

The bottom left panel in **Figure 5.25** shows the mean AoA zonal mean distribution averaged over the period 1989 - 2010 from the ERA-Interim TOMCAT CTM simulation. Maximum values older than 5.0 years are reached above 58 km over the tropics, and above 30 km in the SH high latitudes; younger values are found over the NH high latitudes than over the SH. Thus TOMCAT mean AoA values are in good agreement with KASIMA, although KASIMA yields slightly older values over high latitudes in both hemispheres. TOMCAT AoA values are also similar to those obtained with the TRACZILLA "raw" simulation (without corrections), and with the CLaMS ERA-Interim simulation.

The mean AoA distribution obtained with TOMCAT for the period 2002 - 2007 is in good overall agreement with that from BASCOE (**Figure 5.26**), although BASCOE AoA maximum values are slightly younger (less than 0.5 years) than TOMCAT for all latitudes, except over the SH polar region where they are around 0.5 years older.



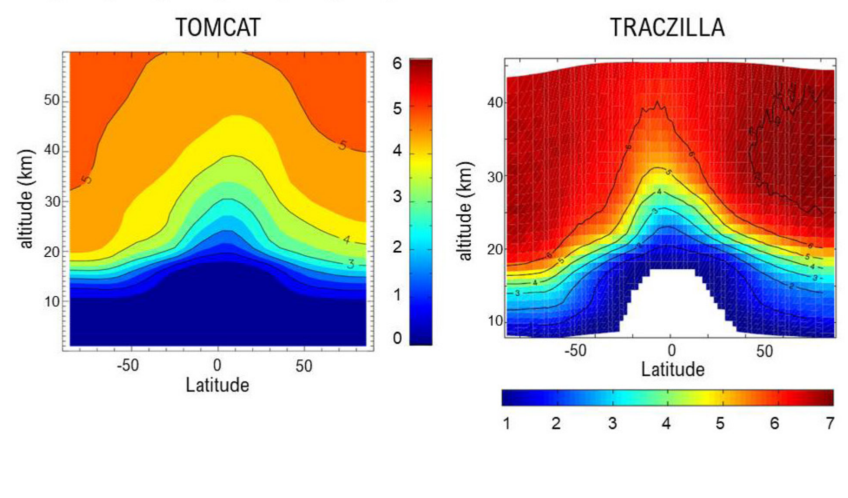


Figure 5.25: *Zonal cross-section of the mean AoA, averaged over 1989 - 2010, as obtained by the different offline models with ERA-Interim reanalysis fields. The offline models used are: (first row left to right) BASCOE, CLaMS, KASIMA, (second row left to right) TOMCAT and TRACZILLA.*

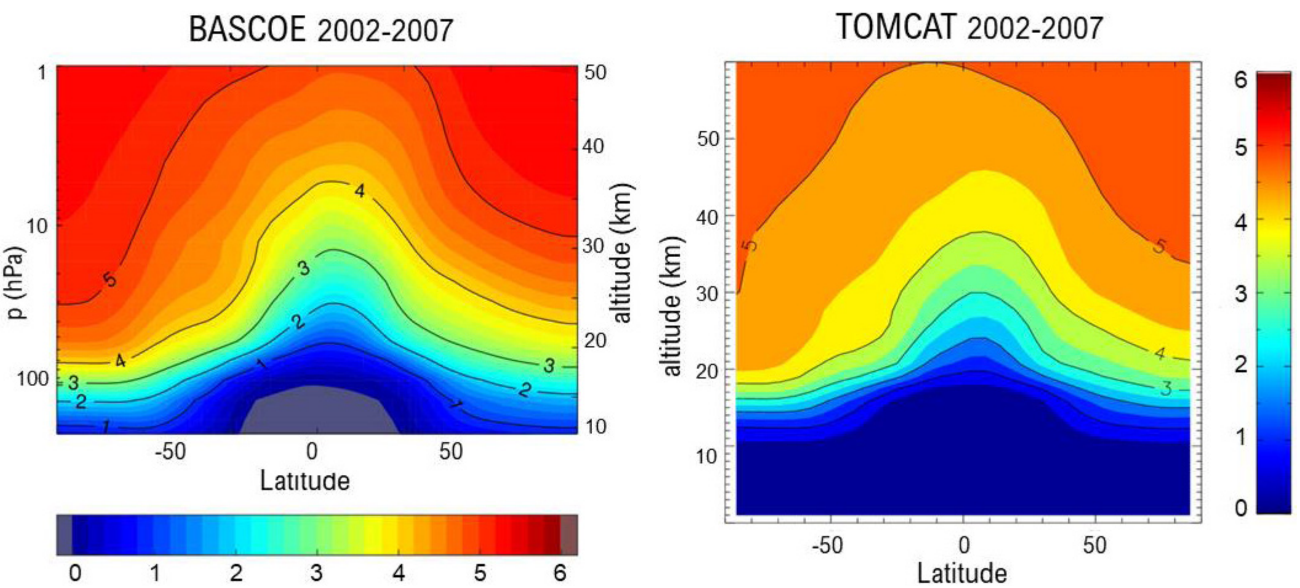


Figure 5.26: *Mean AoA zonal mean averaged over the period 2002 - 2007 from the BASCOE (left) and the TOMCAT (right) simulations driven by ERA-Interim reanalysis.*

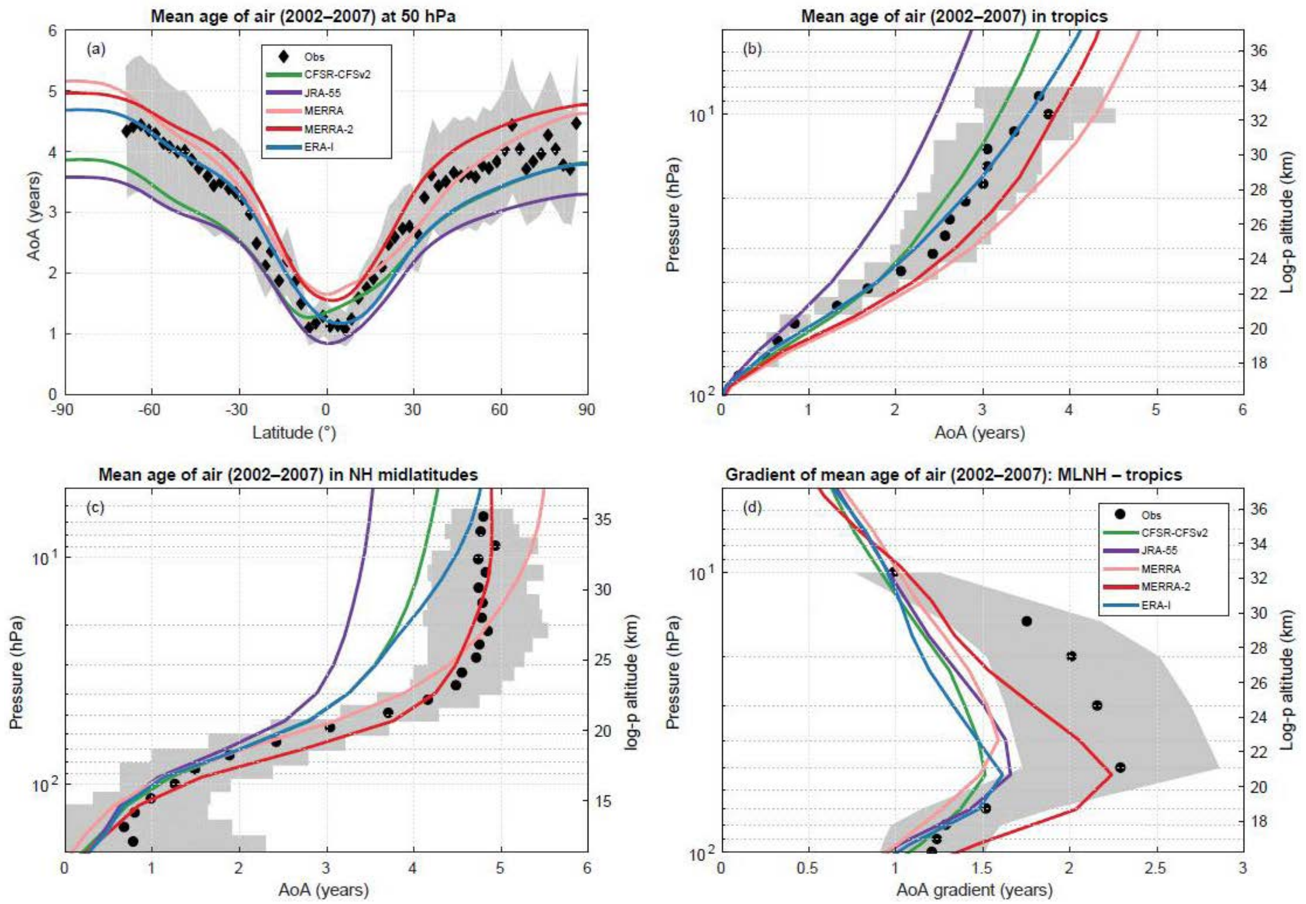


Figure 5.27: *Mean AoA in 2002 - 2007 by the BASCOE TM driven by five reanalyses (colour solid lines) versus in-situ observations (symbols) with their 1σ uncertainties (grey shading). The five reanalyses are ERA-I (blue), MERRA-2 (red), MERRA (pink), JRA-55 (purple) and CFSR (green). The modeled AoA fields are corrected so that mean age = 0 at the tropical tropopause (100 hPa). (a) AoA at 50 hPa with aircraft observations of CO_2 (Andrews et al., 2001); (b) AoA in the tropics(10° N - 10° S) with aircraft observations (Andrews et al., 2001); (c) AoA in the northern mid-latitudes (35° N - 45° N) with balloon observations s (Engel et al., 2009) and (d) AoA gradient between the northern mid-latitudes and tropics (Chipperfield et al., 2014; Neu et al., 2010). Figure from Chabrillat et al. (2018).*

A comparison of five different reanalyses (CFSR, JRA-55, MERRA, MERRA-2 and ERA-Interim) over this shorter period (2002 - 2007) obtained with BASCOE is shown in **Figure 5.27**. The AoA distribution is averaged for this period to remove seasonal and quasi-biennal oscillation signals. The figure shows the intercomparison of AoA zonal means at 50 hPa, vertical profiles over the tropics and over northern midlatitudes, and the gradient profile between these two latitudinal bands.

The intercomparison at 50 hPa (**Figure 5.27.a**) shows large disagreement between the five model simulations. JRA-55 yields the youngest AoA at all latitudes, with values ranging from 0.8 years at the equator to 3.6 years at the South Pole, while MERRA and MERRA-2 give the oldest AoA with 1.6 years at the equator and around 5 years at the South Pole. CFSR and ERA-Interim yield intermediate results with nearly identical values in the northern extratropics but different latitude gradients in the tropics and SH. In the SH, CFSR results in mean AoA values nearly as young as JRA-55 while ERA-Interim reaches larger values much closer to observations. Overall, the spread between the five simulations at 50 hPa is larger than the 1-σ observational uncertainties in the tropics, and nearly as large in the extratropics.

The AoA difference between the tropics and mid-latitudes (**Figure 5.27.d**) is directly related to the inverse of the tropical upwelling velocity and is independent of quasi-horizontal mixing: a smaller AoA latitudinal gradient indicates faster tropical ascent (*Linz et al.*, 2016). The agreement among reanalyses for the considered period is better for the AoA latitudinal gradients than for the AoA profiles. The spread between the four reanalyses (MERRA-2 excluded) reaches a maximum of 0.2 years at 30 hPa. While there is good agreement with the observation-based latitudinal gradient from 10 - 60 hPa, the four reanalyses significantly underestimate the AoA for the pressure range in between those two levels. This indicates an overestimation of the tropical upwelling obtained with ERA-Interim, CFSR, JRA-55 and MERRA in the LS region. MERRA-2 shows an outlying vertical profile of mean AoA for the latitudinal gradient; it underestimates the tropical upwelling in the lowermost stratosphere (100 - 60hPa), agrees well with observations at 50 hPa and joins the results of the four other reanalyses above that level.

The zonal cross-section of mean AoA for the period 2002 - 2007 from the BASCOE simulation driven by ERA-Interim (**Figure 5.26** left panel) shows the expected hemispheric asymmetry with a stronger latitudinal gradient in southern mid-latitudes and polar regions than in the NH. It also shows old air masses reaching lower altitudes over the Antarctic than over the Arctic. The corresponding mean AoA distributions obtained with the other four reanalyses (**Figure 5.28**) are significantly different. JRA-55 and CFSR are the "younger reanalyses" with AoA not exceeding 5 years in the polar upper stratosphere; MERRA is the "oldest reanalysis" with maximum AoA values as large as 6.5 years; ERA-Interim shows intermediate values (5.8 years in the same regions). MERRA-2 shows upper stratospheric values similar to those with ERA-Interim but very different latitudinal gradients. Also the hemispheric asymmetry is more evident with ERA-Interim than with any other reanalysis, *e.g.*, the 3 and 4-year isolines with JRA-55 and CFSR respectively, or the 5-year isoline with MERRA-2 and MERRA, reach nearly the same level above the North Pole than above the South Pole. MERRA-2 stands out in the middle stratosphere with nearly vertical isolines, *i.e.*, very small vertical gradients. Regarding differences in the mean AoA values themselves (bottom row in **Figure 5.28**) the largest relative differences with respect to ERA-Interim are found in the tropical lower stratosphere for all reanalyses, except for JRA-55 which shows the largest differences in the middle and upper stratosphere.

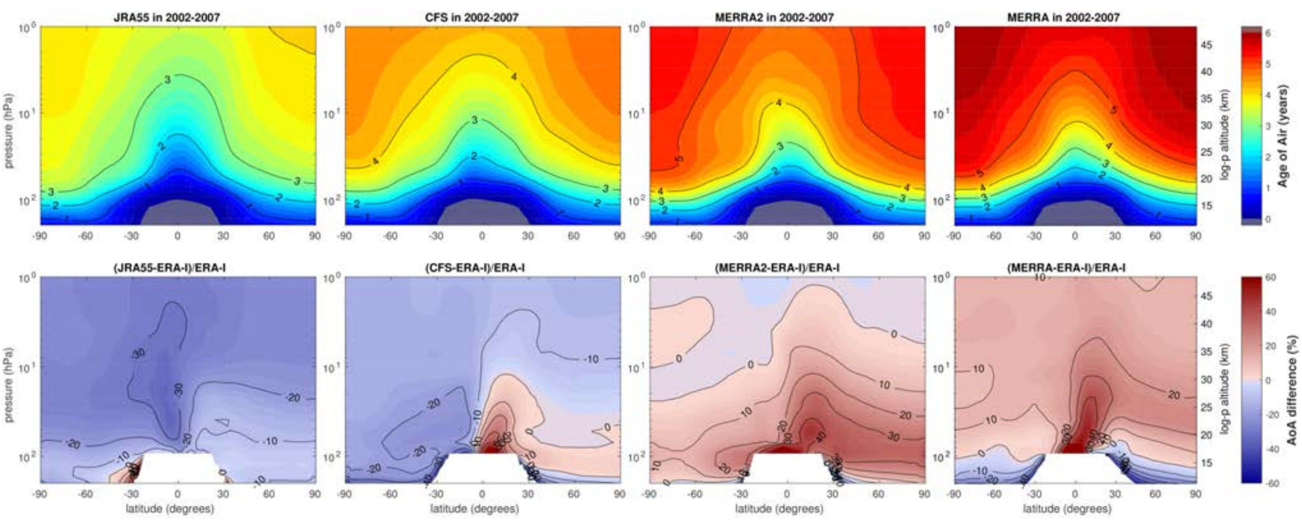


***Figure 5.28:** Latitude-pressure distribution of mean AoA averaged over the period 2002 - 2007 from BASCOE simulations driven by all reanalyses but ERA-Interim (top row). The reanalyses are, from left to right, JRA-55, CFSR, MERRA-2 and MERRA. The bottom row shows corresponding relative differences with respect to the mean AoA from the ERA-Interim-driven simulation in **Figure 5.26**; darker blue indicates more negative differences and darker red more positive differences. Figure from Chabrillat et al. (2018).*

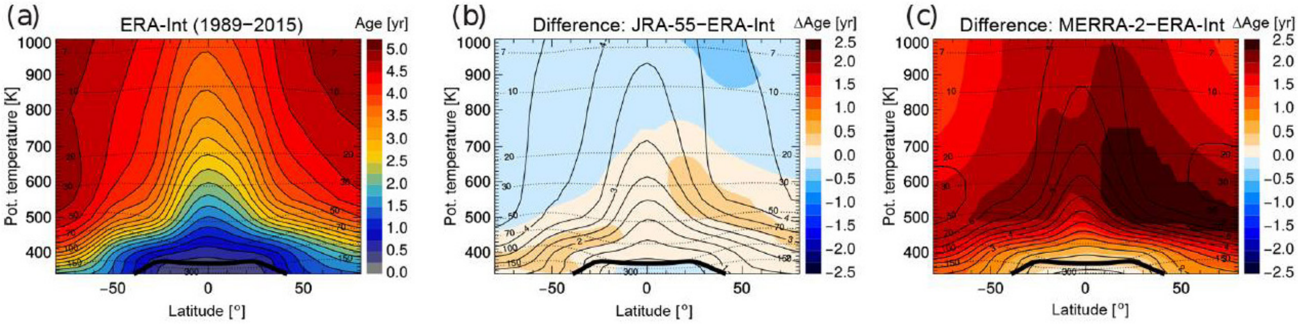


Figure 5.29: *Mean age climatology (1989 - 2015) obtained from the CLaMS simulation with for ERA-Interim (left), and differences for the corresponding simulations with JRA-55 (middle), and MERRA-2 (right). Thin solid black lines highlight particular mean age contours, thin dashed black lines show pressure levels in hPa, and the thick black line is the (lapse rate) tropopause (calculated from each reanalysis following WMO, 1957). Figure from Ploeger et al. (2019).*

JRA-55 is younger than ERA-Interim for all latitudes and altitudes, while MERRA-2 is older than ERA-Interim for all regions. MERRA and MERRA-2 exhibit a similar differences pattern but MERRA show younger values than ERA-Interim in the LS extratropics. The overall classification of mean AOA values from BASCOE simulations for the period 2002 - 2007 also holds for the whole 1989 - 2015 period (as discussed later in *Section 5.5.2.5*): MERRA and MERRA-2 result in the oldest mean AoA, JRA-55 and CFSR the youngest, ERA-Interim lays in between.

Figure 5.29 shows the climatological average for the zonal mean AoA obtained with the CLaMS offline model with ERA-Interim, and the differences for the corresponding simulations with JRA-55 and MERRA-2, for the period 1989 - 2015. For the CLaMS simulations, JRA-55 is older than ERA-Interim in the LS region for all latitudes (as opposed to BASCOE results), and younger above 700 K (as found with BASCOE). The mean age from MERRA-2 is older than ERA-Interim by more than two years throughout most of the stratosphere (similarly to BASCOE results, but with CLaMS we see differences larger than two years).

Figure 5.30 shows the zonally averaged mean AoA for December to February (DJF) and June to August (JJA) seasons as obtained with the CLaMS model; these CLaMS simulations compare results obtained with ERA-Interim, JRA-55 and MERRA-2 and are averaged over the period 1980 - 2015. The global patterns in the mean age distribution are robust for the three reanalyses. However, the exact mean age values are sensitive to the dataset used. Overall, JRA-55 shows the youngest stratospheric mean age, MERRA-2 the oldest mean age, and ERA-Interim is in between. In particular, MERRA-2 shows the largest differences compared to the other two reanalyses, with mean AoA values about two years older in many regions of the stratosphere, consistent with the corresponding differences in the annual climatology (**Figure 5.29**).

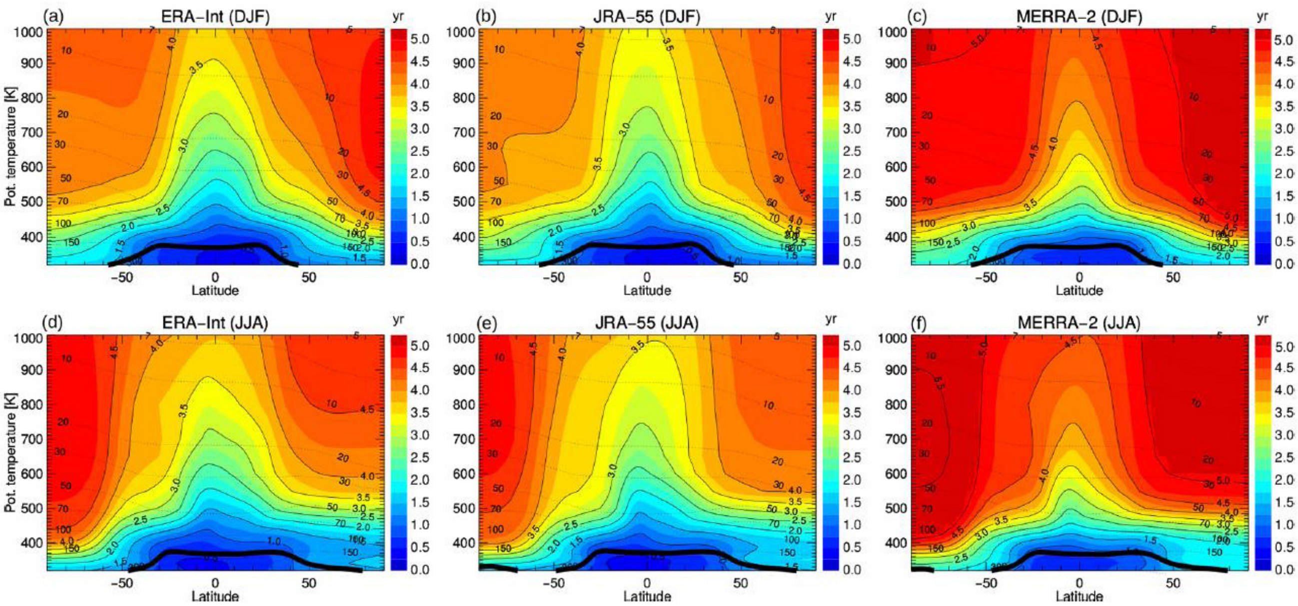


Figure 5.30: *Mean age climatology (1989 - 2015) for DJF (top) and JJA (bottom) for ERA-Interim (left), JRA-55 (middle), and MERRA-2 (right). Thin solid black lines highlight particular mean age contours, thin dashed black lines show pressure levels in hPa, and the thick black line is the (lapse rate) tropopause (calculated from each reanalysis following WMO, 1957). Figure from Ploeger et al. (2019).*

These differences in mean AoA are consistent with differences in the diabatic heating rate fields that drive the vertical transport in CLaMS; heating rates in **Figure 5.20** showed a stronger tropical upwelling for JRA-55 than for ERA-Interim for the period 1989 onwards, and a weaker tropical upwelling for MERRA-2. Despite the different ways in which both transport models operate, the classification of older and younger reanalysis obtained with CLaMS agrees with that obtained from BASCOE, which provides robustness to this classification result. The only exception is the difference between JRA-55 and ERA-Interim in the lower stratosphere.

Simulations with the TRACZILLA Lagrangian model also confirm this overall classification of mean AoA values obtained with these reanalysis datastes (**Figure 5.31**). TRACZILLA, as CLaMS, calculates the mean AoA values from the age spectrum distributions. **Figure 5.31** displays the zonal-mean mean AoA obtained by TRACZILLA when using ERA-Interim, MERRA and JRA-55 reanalyses. The youngest values correspond to JRA-55, MERRA shows the oldest values and ERA-Interim is in between but much closer to JRA-55 than to MERRA. The figure also shows a comparison of the effect of mass correction and of different correction techniques applied to the tail of the age spectrum for the different datasets.

The effect of the mass correction is smaller for ERA-Interim than for the other two reanalysis's, and for this dataset it acts making mean AoA values overall younger; a similar effect is true for JRA-55. However, in the case of the MERRA dataset, applying the mass correction makes mean AoA overall older. The clipping techniques have an effect on the mean AoA which to a large extent depends on the tail of age distribution. The slope of this distribution is much flatter for MERRA than for ERA-Interim and JRA55, which makes MERRA the most sensitive dataset to these corrections.

In general, uncorrected ages not accounting for the tail (first row in **Figure 5.31**) are too young, *e.g.*, for ERA-Interim the tail correction accounts for an increase in AoA of up to 1.8 years in polar regions. Compared to observation-based AoA distributions (*e.g., Section 5.5.2.2* above), the uncorrected and unclipped ages are too young for all reanalysis but the effect of applying tail correction varies according to the chosen clipping in a different reanalyses. The strong interplay between mass correction and clipping in the ERA-Interim suggests that uniform mass correction is probably inappropriate for ERA-Interim.

From these mean AoA distributions, MERRA data seem to provide much weaker tropical upwelling than the other reanalyses used by TRACZILLA. When applying the tail clipping correction techniques, MERRA provides an unrealistically old AoA compared to the other reanalyses.

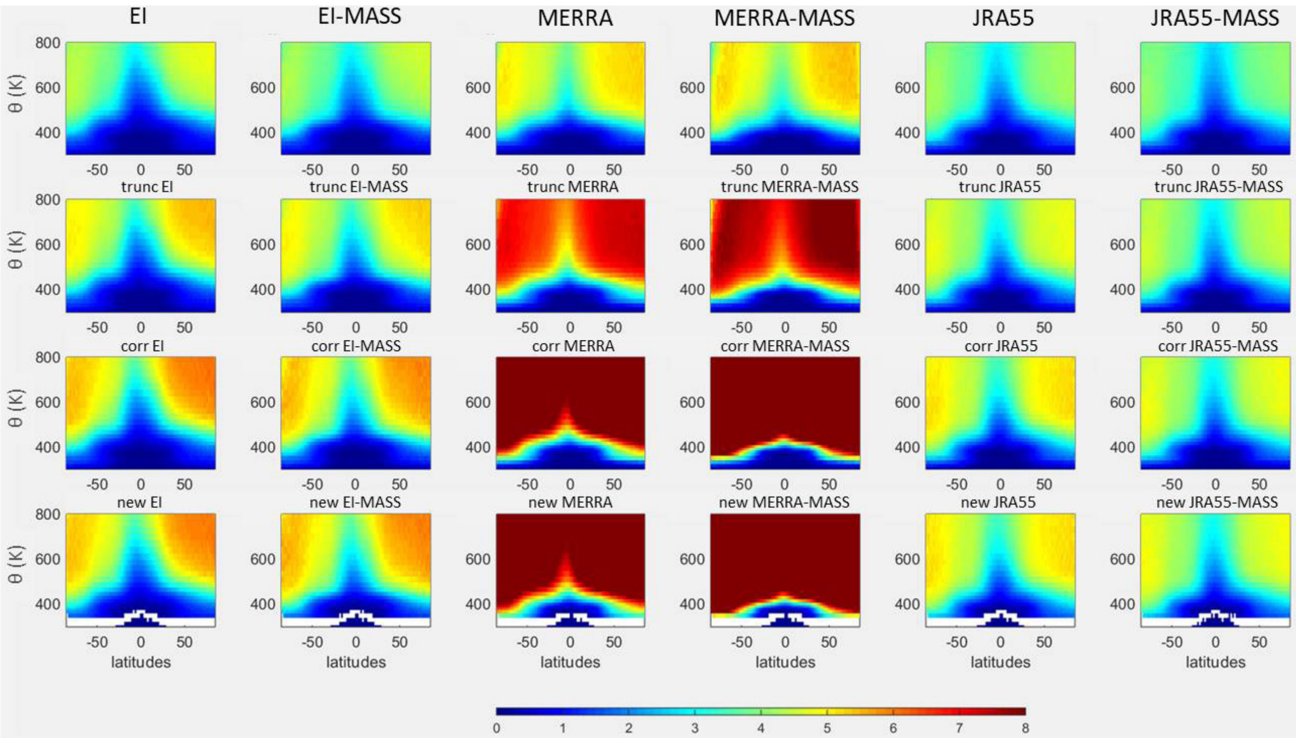


Figure 5.31: *Comparison of the mean AoA for ERA-Interim (two left columns), MERRA (two middle columns) and JRA-55 (two right columns) from TRACZILLA simulations for the period 1989-2010. For each reanalysis the left column shows the case without mass correction, and the right column shows the case with mass correction. The first row is without tail correction, the second row uses a correction by setting all the old parcels to 10 years. The third and the fourth rows are two different versions of the spectrum tail correction applied according to Scheele (2005), more details in the main text.*

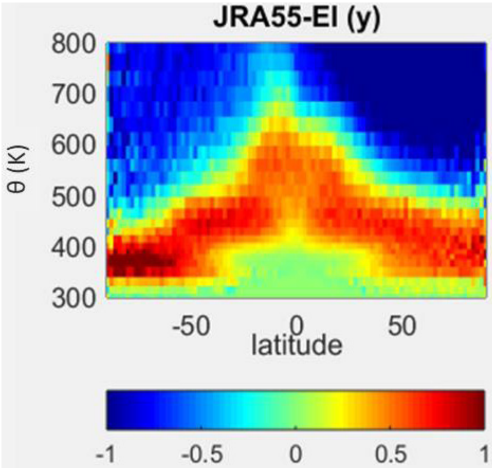


***Figure 5.32:** Difference in mean AoA from the TRACZILLA run with JRA-55 minus the run with ERA-Interim. The configuration is that giving the best choice for ERA-Interim, i.e., with tail correction, clipping at 0.5 hPa, and no mass correction.*

The results from these TRACZILLA simulations clearly evidence that the same correction technique that makes one particular reanalysis dataset perform more realistically may not have the same effect with a different reanalysis and, therefore, such correction techniques need to be applied with caution.

Using the combination of corrections and clipping that gives the best results with ERA-Interim, **Figure 5.32** shows the difference in mean age-of-air from the TRACZILLA runs with JRA-55 and ERA-Interim. In agreement with CLaMS results (**Figure 5.29b**), JRA-55 is older than ERA-Interim in the lower tropical stratosphere and the extratropical lowermost stratosphere while it is younger at higher altitudes, especially in the NH. This suggests that the BDC favours the lower branch in ERA-Interim and the upper branch in JRA-55.

5.5.2.4 Age spectrum

The enormous advantage of the mean AoA diagnostic is the possibility of comparing it with actual tracers observations; however, a more complete picture of stratospheric transport in models can be obtained from the age spectrum diagnostic. Age spectrum distributions have been computed by the CLaMS and the TRACZILLA Lagrangian models with ERA-Interim for the period 2000 - 2010, showing a remarkable agreement of the spectra between the two models (Figure not shown); although the median and mean ages were overall older in TRACZILLA than in CLaMS. This is due to differences in the tail distribution of ages, in agreement with the differences in mean AoA distributions shown earlier for these two models.

CLaMS has performed age spectrum calculations with ERA-Interim, JRA-55 and MERRA-2 for the period 1989 - 2013 (**Figure 5.33**). The 400 K isentrope has been chosen as a representative level for the shallow BDC branch, while the spectra at 600 K represent levels at which the deep BDC branch dominates. The CLaMS age spectra show similar variability between ERA-Interim, JRA-55 and MERRA-2. In particular, multiple peaks in lower stratospheric age spectra are a common and robust feature for the three reanalysis. Effects of mixing as shown by changes in the spectrum tail are more sensitive to the reanalysis data used. For MERRA-2, the transition between tropical and extratropical age spectra is less pronounced, for age values older than 2 years, indicating stronger exchange between tropics and middle latitudes in the LS region. This stronger exchange likely causes a stronger recirculation of extratropical older air masses into the tropics, resulting in the older AoA mean ages values shown in previous sections.

Age spectra results from TRACZILLA simulations, for the period 1979 - 2010, are displayed in **Figure 5.34** for ERA-Interim, JRA-55 and MERRA. The annual modulation shown is due to the more intense BDC in the NH winter. It is visible that the amplitude decays much faster as a function of age in MERRA but has also a flatter tail, similar to what CLaMS has found with MERRA-2 data (**Figure 5.33**). JRA-55 is the reanalysis with the strongest annual modulation of the spectrum in TRACZILLA simulations. With TRACZILLA the three reanalysis, but especially MERRA, display reduced age and reduced modulation of the cycle in years following the Pinatubo eruption in June 1991. This effect appears to propagate across most of the 1990's, although the post Pinatubo transition coincides with the introduction of AMSU satellite observations in the reanalyses, and both effects can be confused. It is worth noting that none of the reanalyses considered explicitly includes the effects of the Pinatubo aerosols injection.

For the same TRACZILLA simulations, **Figure 5.35** shows that the horizontal distribution of the spectrum displays fairly similar patterns in the three reanalyses in the lower stratosphere, but at higher altitudes the two-lobe pattern clearly exhibited by ERA-Interim and JRA-55 is replaced by a one-lobe pattern in MERRA. This is an indication of a more leaky tropical pipe in MERRA, which is consistent with the distribution of young air in the tropical region for this dataset (**Figure 5.31**).

5.5.2.5 Mean AoA time evolution

Time series of mean AoA in the middle stratosphere, averaged between 30 hPa and 5 hPa, are displayed in **Figure 5.36**. These have been obtained with BASCOE with the five reanalyses shown for the SH and the NH. This figure shows the large disagreements among the five reanalyses over the long-term period 1989 - 2015. In the SH, MERRA and MERRA-2 values decrease quickly until 1995 and increase after 2007 while ERA-Interim values follow an opposite pattern.

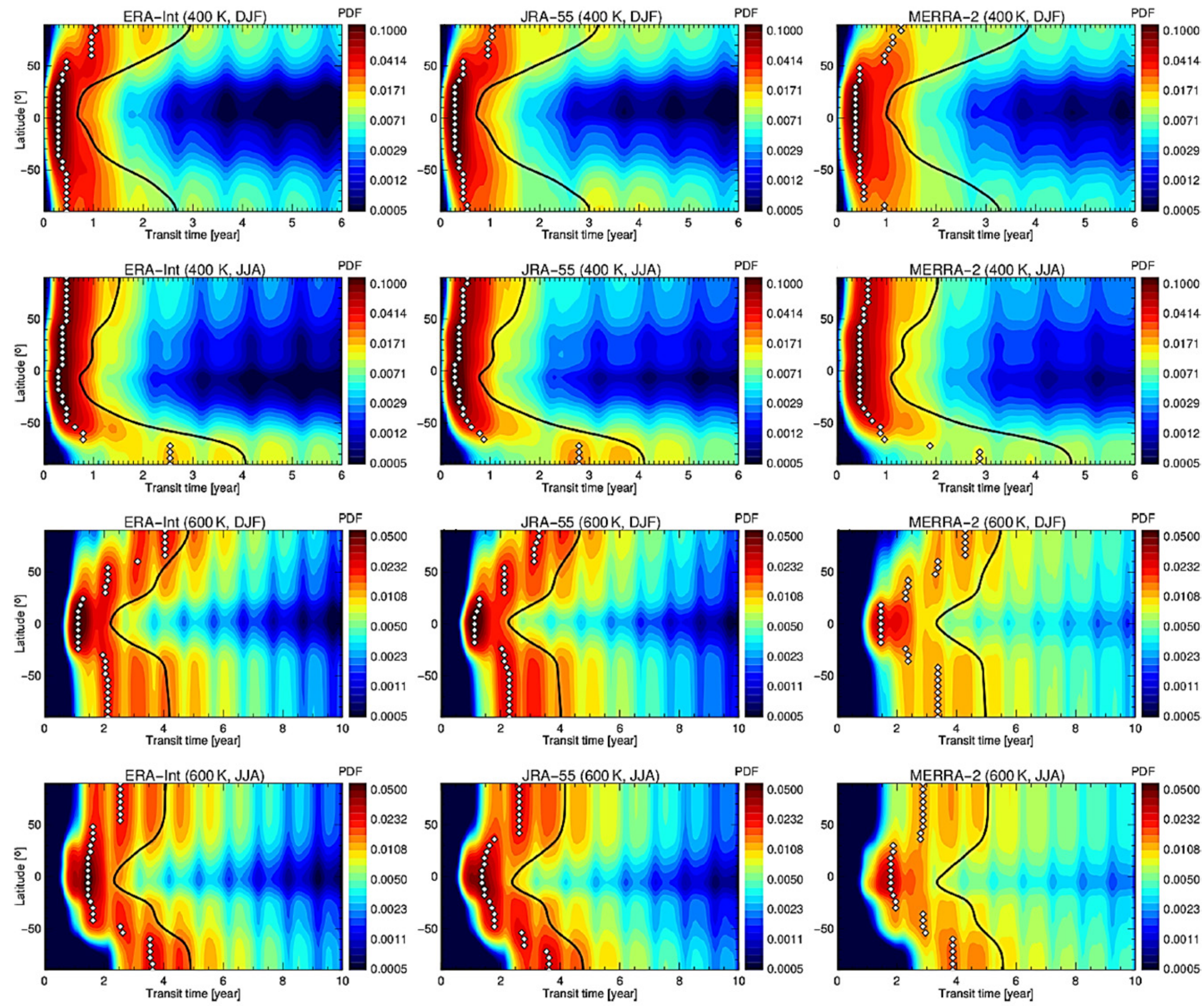


Figure 5.33: *Age spectrum from CLaMS simulations for 1989 - 2013 at 400 K (top two rows) for Dec-Feb (DJF) and Jun - Aug (JJA); and the same at the 600 K potential temperature isentrope (bottom two rows). Results correspond to the simulations using ERA-Interim (left), JRA-55 (middle), and MERRA-2 (right) reanalyses. The black line shows the mean of the AoA spectrum in each case, while the white symbols show the modal age. Figure adapted from Ploeger et al. (2019).*

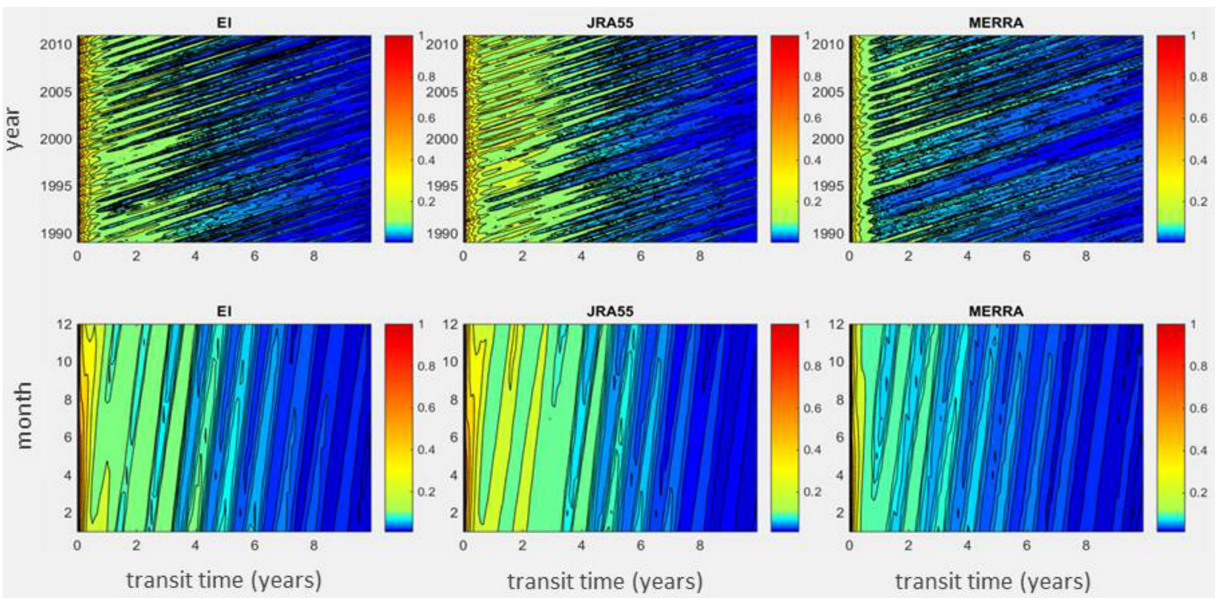


Figure 5.34: *Age spectrum average over the whole stratosphere below 800 K as obtained by TRACZILLA for the period 1989 - 2010 (upper panels) and the age spectrum mean annual cycle in the lower row. The reanalyses used, left to right, are ERA-Interim, JRA-55 and MERRA.*

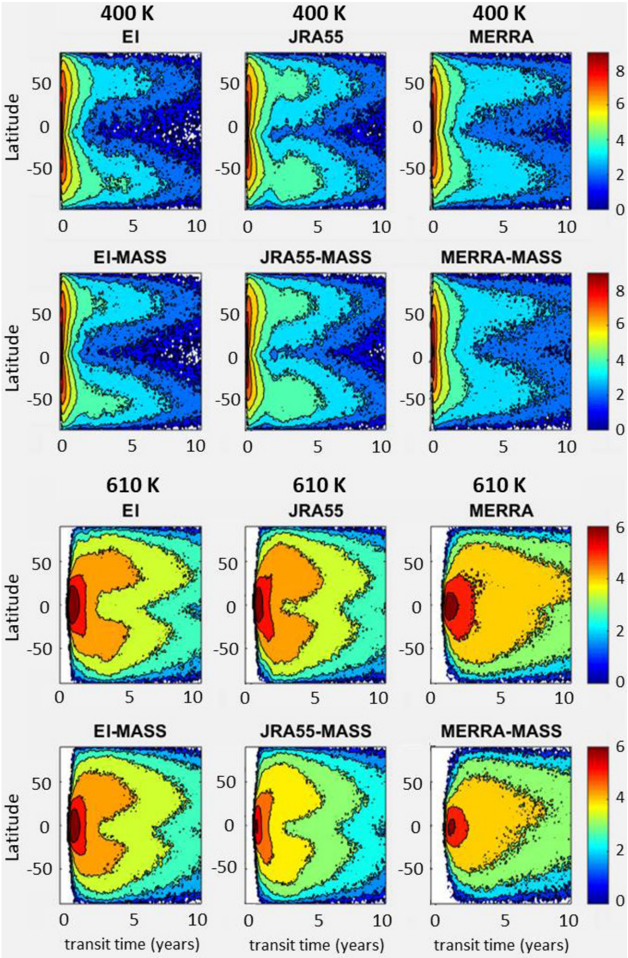


***Figure 5.35:** Horizontal sections of the age spectrum as a function of latitude for two different levels: 400K (top), and 610K (bottom), as obtained from TRACZILLA simulations with ERA-Interim (left), JRA-55 (centre) and MERRA (right), without mass correction (upper row) and with mass correction (lower row).*

The long-term evolution of AoA in this region is very different with JRA-55, which shows a gradual decrease until 2002 followed by a slight recovery and stabilization after 2005, and differs also from CFSR, which shows no trend before 1997 and a rapid increase during 1997 - 2003.

Thin lines in **Figure 5.36** allow a qualitative comparison of faster variations in the five time series. The seasonal signal dominates in all cases, and all reanalyses show similar phases: AoA is older in autumn and younger in spring. The seasonal amplitudes in the SH are very dependent on the particular year but also on the considered reanalysis. It can be seen that some reanalyses, in particular MERRA and ERA-Interim, exhibit a stronger modulation of the seasonal cycle by the QBO than the others; for these two reanalyses the seasonal amplitude during easterly QBO years (*e.g.*, 2006, 2008) is half of that during westerly QBO years (*e.g.*, 2005, 2009). For the NH, **Figure 5.36** (right panel) compares the BASCOE model results with the balloon observations from *Engel et al.* (2009; 2017). The spread between the five simulations is as large as the observational uncertainties, highlighting again the magnitude of the disagreements between the five reanalyses. ERA-Interim delivers a small positive trend over the period 1989 - 2015, in agreement with the balloon observations.

5.5.2.6 Mean AoA trends

In the late 2000s, *Engel et al.* (2009), based on CO_2 and SF_6 observations, suggested that the widespread result from climate models predicting increasing strength of the BDC (younger mean AoA values) was not holding over the NH midlatitude stratosphere for recent past decades. **Figure 5.22** (updated from **Figure 3** in *Engel et al.*, 2009) shows the time evolution of mean AoA between 24 - 35 km altitude from SF_6 and CO_2 in-situ measurements from aircrafts and balloons taken from 1975. The *Engel et al.* (2009) study was based on sparse mean AoA observation-based values, and the trend obtained was not statistically significant compared to the observations' uncertainties.

But in 2012 new published studies gave robustness to this apparent discrepancy between climate models and observations. Based on MIPAS global satellite observations, *Stiller et al.* (2012) found a region in the middle stratosphere over NH midlatitudes where mean AoA trends were positive during the MIPAS period (2002 - 2012); this region coincided with the one considered in *Engel et al.* (2009). At the same time, *Monge-Sanz et al.* (2012) was the first model study to show the dipole structure in the mean AoA trend, using offline simulations of the Eulerian TOMCAT CTM driven by ERA-Interim reanalyses covering a 20-year period (1990 - 2009). This model study found a statistically significant (at 95 % confidence level) positive trend in the mean AoA between 25 - 40 km altitude over the NH, in overall agreement with the results derived from MIPAS observations by *Stiller et al.* (2012). A parallel study using the Lagrangian transport model TRACZILLA (*Diallo et al.*, 2012) also showed an heterogeneous structure in the mean AoA trend using ERA-Interim meteorological fields.

These early studies with ERA-Interim prompted an active research debate on the causes for discrepancies between observations and what climate models had been predicting. Increasing our knowledge on this issue has been one of the scientific objectives of the work done by the different model scientists involved in this SRIP Chapter. This section summarises results we have found when computing AoA trends with the different reanalyses.

Figure 5.37 is an updated version of **Figure 3** in *Monge-Sanz et al.* (2012), showing the zonal cross-section of the linear trend in the mean AoA from the offline TOMCAT simulations driven by ERA-Interim for the period 1990 - 2013. The dipole structure in the mean AoA displays maximum positive values over the NH middle stratosphere midlatitudes of up to 0.24 years/decade, and minimum values of up to - 0.14 years/decade. The figure shows that this trend over the NH

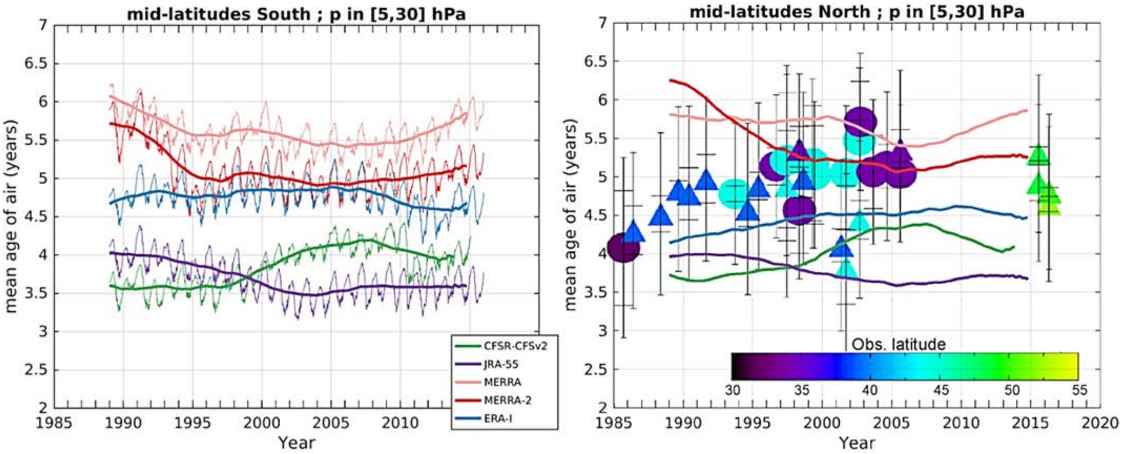


Figure 5.36: *Time evolution of AoA averaged from 30 hPa to 5 hPa (approximately 24 km to 36 km) in the southern (50° S - 40° S, left) and northern mid-latitudes (40° N - 50° N, right). Solid lines show model output with color codes according to the legend shown in the left panel. Thin lines (left panel only; omitted from right panel for clarity) show instantaneous model output every 5 days while thick lines are smoothed with a one-year running mean. Northern mid-latitude symbols (right panel) represent values derived from balloon observations of SF_6 (circles) and CO_2 (triangles) with color code showing the latitude of the measurements and outer error bars including sampling uncertainties (Engel et al., 2017). Adapted from Chabrillat et al. (2018).*

and SH middle stratosphere is statistically significant. The equivalent figure for the TOMCAT trend over the MIPAS period is displayed in **Figure 5.38**, which shows an intensification of the dipole with maximum values of up to + 0.50 years/decade over the NH and - 0.50 years/decade over the SH. This intensification in the AoA trend with ERA-Interim is consistent with the hypothesis in *Miyazaki et al.* (2016). They suggested that the increased eddy transport in the subtropics, and the weakened mean poleward motion in the middle stratosphere found with ERA-Interim during the period 2000 - 2012, would translate into larger increasing trends in the NH compared to the previous 20 years (1979 - 2000).

Figure 5.39 compares the latitude-pressure distributions of AoA trends across five reanalyses for the early (1989 - 2001), recent (2002 - 2015) and overall (1989 - 2015) periods as obtained from BASCOE simulations. It is important to note that the trends over the early and overall periods should be considered with more caution because of the beneficial impact of assimilation of new datasets in later years (*e.g.*, the AMSU dataset from 1998).

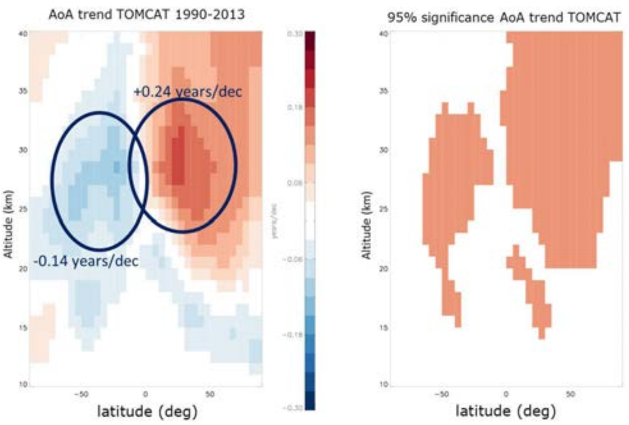


Figure 5.37: *Cross-section of the linear trend (years per decade) of the mean AoA for the period 1990 - 2013 from the TOMCAT simulation with ERAInterim fields (left); red colours indicate positive trends and blue colours negative trends. Regions where the trend is significant at least to the 95% confidence level are shown by the shaded areas in the right panel. The dipole structure in the mean AoA displays maximum positive values over the NH middle stratosphere midlatitudes of up to 0.24 years/decade, and minimum values of up to -0.14 years/decade. Updated from Monge-Sanz et al. (2012).*

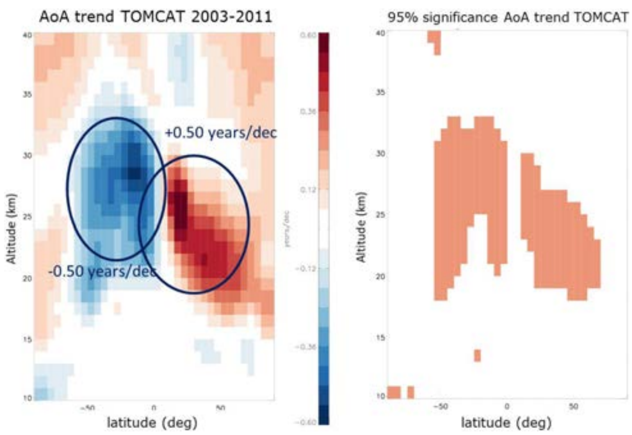


Figure 5.38: *Cross-section of the linear trend (years per decade) of the mean AoA for the period 2003 - 2011 from the TOMCAT simulation with ERA-Interim (left); red colours indicate positive trends and blue colours negative trends. Regions where the trend is significant at least to the 95 % confidence level are shown by the shaded areas in the right panel. The dipole structure in the mean AoA displays maximum positive values over the NH middle stratosphere mid-latitudes of up to 0.50 years/decade, and minimum values of up to - 0.50 years/decade. (From Monge-Sanz et al., in prep).*

The AoA trends derived from ERA-Interim wind fields during the early period (upper left) show unexpected growth in both hemispheres, except in the northern lowermost stratosphere. During the recent period, the dipole structure derived from ERA-Interim (**Figure 5.39** upper middle) is similar to, but less clear than, over the shorter period 2002 - 2012 (**Figure 11** in *Chabrillat et al.*, 2018), with weaker increases in the NH which remain significant only in the polar lower stratosphere. The trend for the overall period 1989 - 2010 (**Figure 5.39** upper right) does not show a dipole structure but positive trends in the middle stratosphere, which are statistically significant over the NH region with positive trends during the 1989 - 2001 period, and significantly negative trends in the lowermost stratosphere at all latitudes (except the SH polar latitudes).

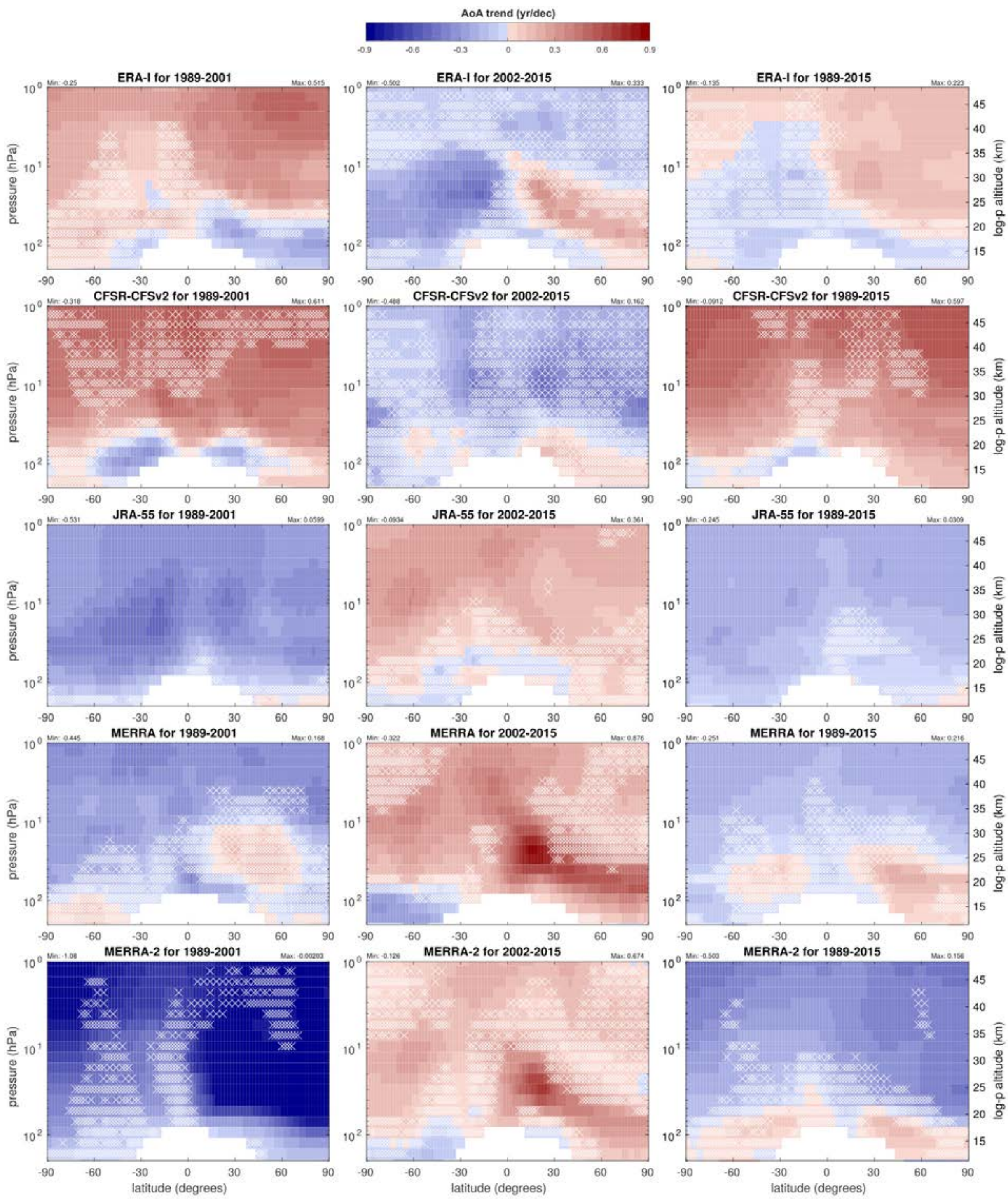


Figure 5.39: *Latitude-pressure distributions of AoA trends (years/decade) over 1989 - 2001 (left column), 2002 - 2015 (middle column) and 1989 - 2015 (right column) using the five reanalyses (from top to bottom: ERA-I, CFSR, JRA-55, MERRA, MERRA-2). White crosses indicate where the sign of the trend is not significant at the 95 % confidence level. Darker blues indicate more negative trends and darker reds more positive trends. Figure from Chabrillat et al. (2018).*

Diallo et al. (2012), using the diabatic Lagrangian transport model TRACZILLA driven by ERA-Interim for the period 1989 - 2010, found negative AoA trends in the lower stratosphere and positive trends in the mid-stratosphere, suggesting that the shallow and deep BDCs may be evolving in opposite ways. *Monge-Sanz et al.* (2012) with the Eulerian TOMCAT model showed significant positive trends over the NH middle stratosphere and negative trends in practically all other regions (see also **Figure 5.37**), although the negative trends were significant only in the LS region and the SH middle stratosphere. The BASCOE transport model simulations, using only wind fields and surface pressure from ERA-Interim, show a similar finding to the previous studies with TOMCAT and TRACZILLA for a similar period.

Comparing the BASCOE trend results obtained with ERA-Interim with those from other reanalyses, there is general agreement between ERA-Interim and CFSR (**Figure 5.39**, first and second rows) while JRA-55, MERRA and MERRA-2 (third to fifth rows in **Figure 5.39**) exhibit overall opposite trends for all periods. A remarkable result in **Figure 5.39** is the overall reversal of trends between the early (1989 - 2001) and recent (2002 - 2015) periods. This reversal is found for all five reanalyses in all regions of the stratosphere (first and second columns in **Figure 5.39**). This period separation for the AoA trend is in agreement with the findings of *Cook and Roscoe* (2009; 2012) for BDC trends over the Antarctic based on polar observations of NO_2.

For the early period, there is very good agreement between ERA-Interim and CFSR (**Figure 5.39**, first and second row) while MERRA shows almost exactly opposite trends, except in the LS where MERRA agrees with CFSR and ERA-Interim. Both JRA-55 and MERRA-2 show negative trends in the whole stratosphere for this period. During the recent period MERRA and MERRA-2 show good agreement. Therefore, the sign of the trend and their statistical significance strongly depends on the input reanalysis. ERA-Interim stands out as the only reanalysis showing a dipole structure in the mean AoA trend for the period 2002 - 2015, in overall agreement with trend values derived from observations. **Figure 5.39** also shows the strong dependence of the trend on the particular period considered, with values above 10 hPa varying between approximately -0.4 and 0.4 years per decade for the same reanalysis, within the same range of values of the interannual variability exhibited by the curves in **Figure 5.36**.

Figure 5.40 shows the linear trend of mean AoA derived from the KASIMA Eulerian model simulations with ERA-Interim for two periods, the overall period 1979 - 2012 (left panel) and the MIPAS period 2002 - 2012 (right panel). The linear trend has been obtained from an idealized linear tracer (T1) with a multi-linear regression analysis including additional annual and semi-annual harmonics and the two QBO indices (*Reddmann et al.*, 2001). The results for the overall period show a positive trend over the NH middle stratosphere of up to 0.3 years/decade, and no significant trend elsewhere. For the MIPAS period, the dipole structure emerges, with more confined positive trend values over the NH low and middle stratosphere between 20 - 30 km of altitude (up to 0.10 years/year) and a negative trend region over the SH low and middle stratosphere (up to - 0.10 years/year).

With the KASIMA simulations we can assess the impact of the mesospheric sink of SF_6 on mean AoA trends. The KASIMA model has used an additional SF_6 tracer (T3) that includes the effects of chemical loss as described in *Reddmann et al.* (2001). **Figure 5.41** shows the cross section of the mean AoA trend with ERA-Interim, for the overlapping MIPAS period 2002 - 2012, when including mesospheric SF6 chemical loss. The general pattern in the low to mid latitude stratosphere is preserved showing the dipole structure in the trend, between 20 - 30 km for both hemispheres, but especially in the SH high latitudes the trend is clearly affected by the chemical loss of SF_6.

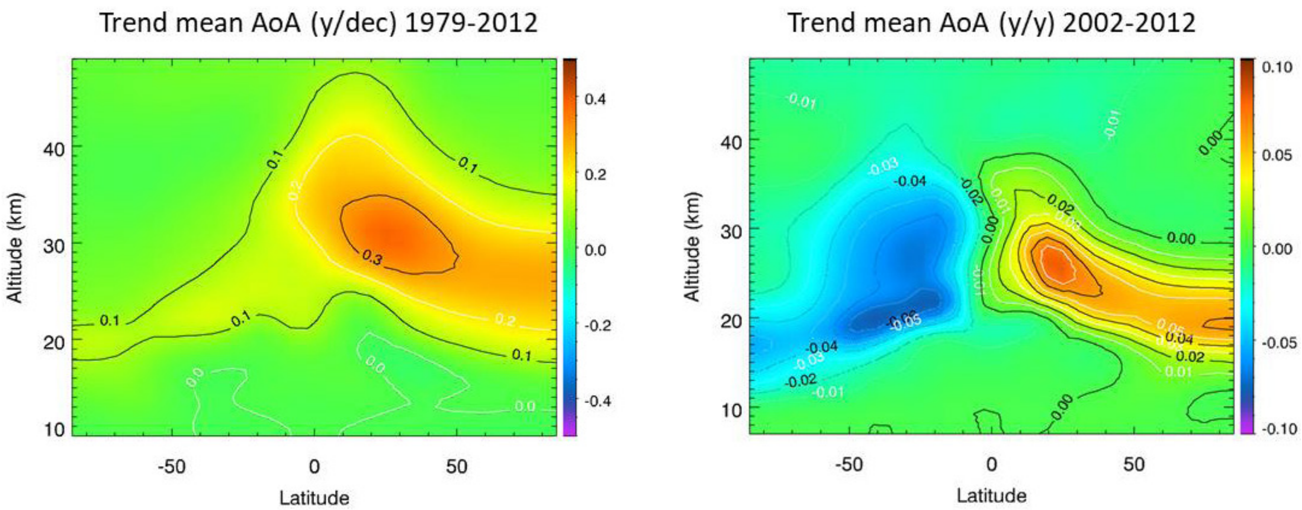


Figure 5.40: *Cross-section of the linear trend in the mean AoA from the linear tracer T1 from the KASIMA model simulation with ERA-Interim. Two different periods are shown: 1979 - 2012 (left) and the MIPAS period 2002 - 2012 (right). Note the different colour scales in both panels.*

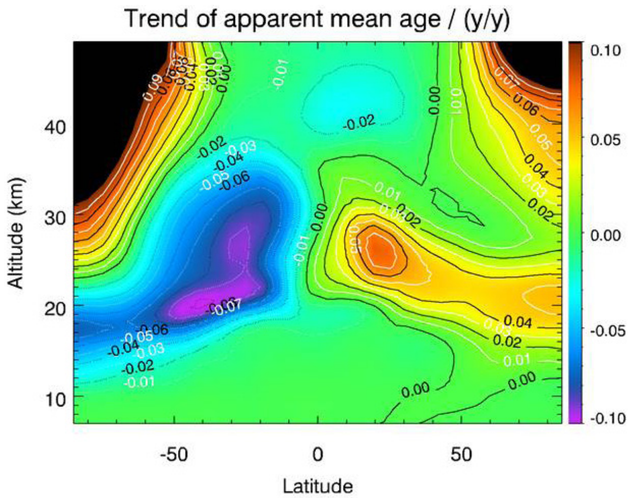


Figure 5.41: *Trend of the apparent mean AoA (expressed as the lag time) of the SF_6 tracer T3 in the KASIMA simulation with ERA-Interim (2002 - 2012), when including mesospheric loss for SF_6.*

The derived trend pattern agrees well with the results of the SF_6 trend features from MIPAS observations in the upper stratosphere (*Haenel et al.*, 2015; *Stiller et al., 2012*). Whereas tracer T3 provides the most realistic results from KASIMA's simulations compared with SF_6 observations, one needs to be cautious as the loss mechanism of SF_6 is subject to significant uncertainties.

Figure 5.42 shows the effect of mass-correction in the mean AoA trend values obtained with the TRACZILLA model driven by ERA-Interim, JRA-55 and MERRA renalayses. The priod 1989 - 2010 has been used in this simulations. There are large differences between reanalyses: for the non-corrected reanalyses fields, both ERA-Interim and JRA-55 show a decrease of mean AoA in the lower stratosphere, but ERA-Interim shows a positive trend in the higher levels in the extratropics, while JRA-55 shows a general negative trend in the middle and upper stratosphere. MERRA shows a similar pattern to JRA-55 but negative values are larger and an area of significant positive trends appears centred over 50 °N in the LS. This agrees with the overall trend structure found with BASCOE (**Figure 5.39** right column panels) when comparing these three reanalyses, however mean AoA trends from CLaMS simulations only agree with BASCOE and TRACZILLA trends for the ERA-Interim reanalysis (**Figure 7** in *Ploeger et al.*, 2019).

When using the mass-corrected TRACZILLA simulations, **Figure 5.42** shows a similar overall structure as with the non-corrected fields but differences are also evident: i) positive trends in ERA-Interim become stronger while negative trends in JRA-55 become weaker, and much weaker for MERRA; ii) for the three reanalyses results become non statistically significant in a larger area of the LS and the tropical pipe.

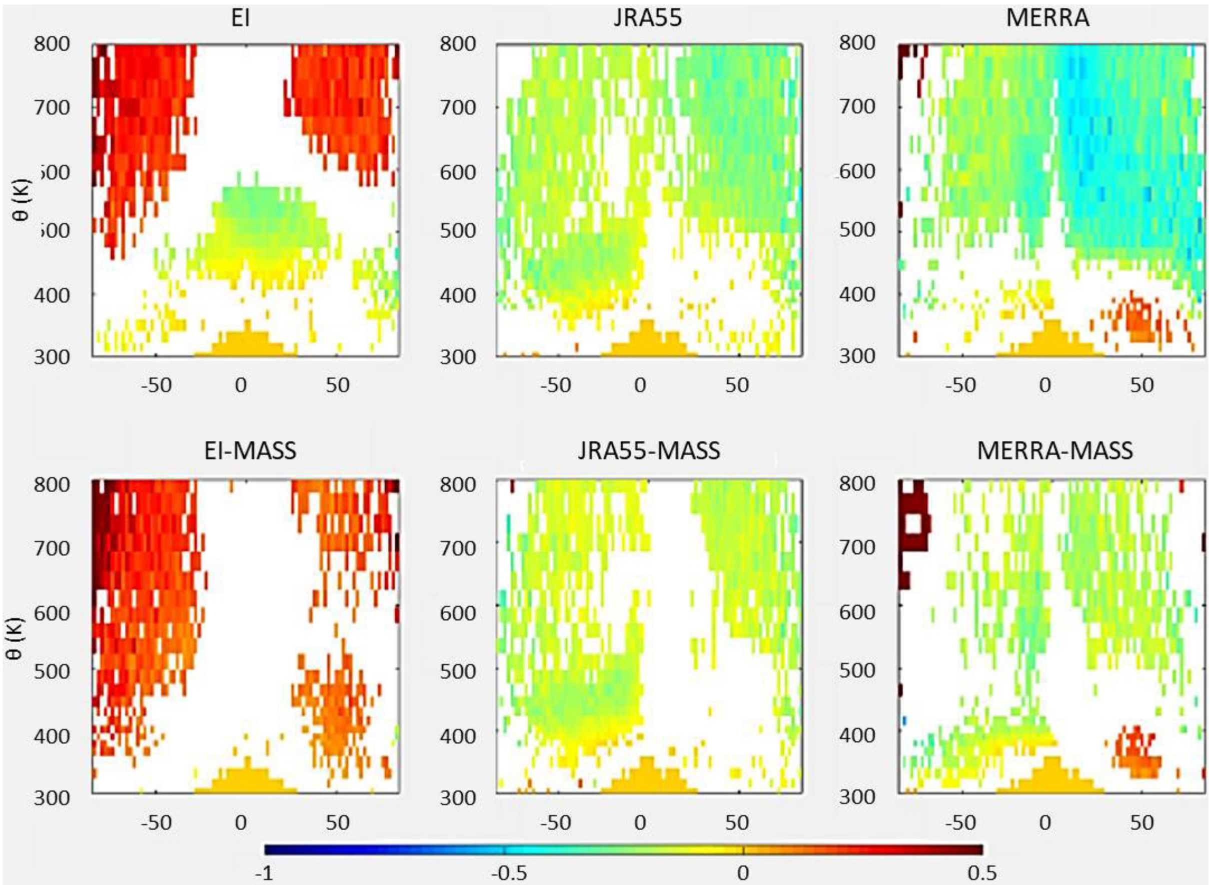


Figure 5.42: *Trends (years/decade) from the TRACZILLA simulations with ERA-Interim (left), JRA-55 (middle) and MERRA (right), with mass correction (bottom) and without (top) mass correction. They have been obtained over the period 1989-2010. Green and blue colours show negative trend values, orange and red colours show positive trends. Non-significant areas are white.*

In addition, an area of positive trend values appears for ERA-Interim around the 400 K isentropic level for NH mid-latitudes, similar to the one featured with MERRA. These differences show the strong effect the mass correction can have on the mean AoA diagnostic, and therefore on chemical tracers distributions obtained by CTMs driven by these reanalyses.

In this section we have shown that mean AoA trends are dependent, not only on the reanalysis used, but also on the exact dates used to calculate such trends; this is also in agreement with recent CCM model studies (*e.g.*, *Garfinkel et al.*, 2017; *Hardiman et al.*, 2017). However, a robust feature emerging from the previous trend distributions is also that, during the period covered by MIPAS observations, ERA-Interim simulations are in significantly better agreement with observations than simulations driven by the other reanalyses; and the trend observed during this period contributes to explain other observed trends in atmospheric tracers (*e.g.*, *Mahieu et al.*, 2014), which adds robustness to this feature.

In the offline tracer simulations that we have examined, as in the real atmosphere, the trends in mean AoA are due to the combined changes in mean-meridional circulation (MMC) and eddy mixing processes. A few recent studies have dealt with ways to quantify the separate contribution of both effects to mean AoA model distributions: *Garny et al.* (2014) quantified the effect of age by mixing in a climate model as the difference between the mean AoA distribution and the corresponding RCTT distribution; *Ploeger et al.,* (2015) do the same with CLaMS ERA-Interim simulations to quantify the two contributions (residual circulation and mixing) to the AoA trend for the MIPAS period (2002 - 2012); and *Miyazaki et al.,* (2016) performed a thorough comparison of MMC and eddy mixing in six reanalyses (ERA-Interim, JRA-55, CFSR, and their predecessor versions ERA-40, JRA-25 and NCEP) and discussed this comparison results also in the context of expected impacts on corresponding AoA distributions. Overall, for the periods and reanalyses they considered, *Miyazaki et al.* (2016) found more consistency among reanalyses regarding mixing processes than MMC.

5.5.2.7 Impact of other processes on the AoA

Quasi-Biennial Oscillation

A point on which the reanalysis strongly disagree is the amplitude and pattern of the correlation of AoA with the quasi-biennial oscillation (QBO).

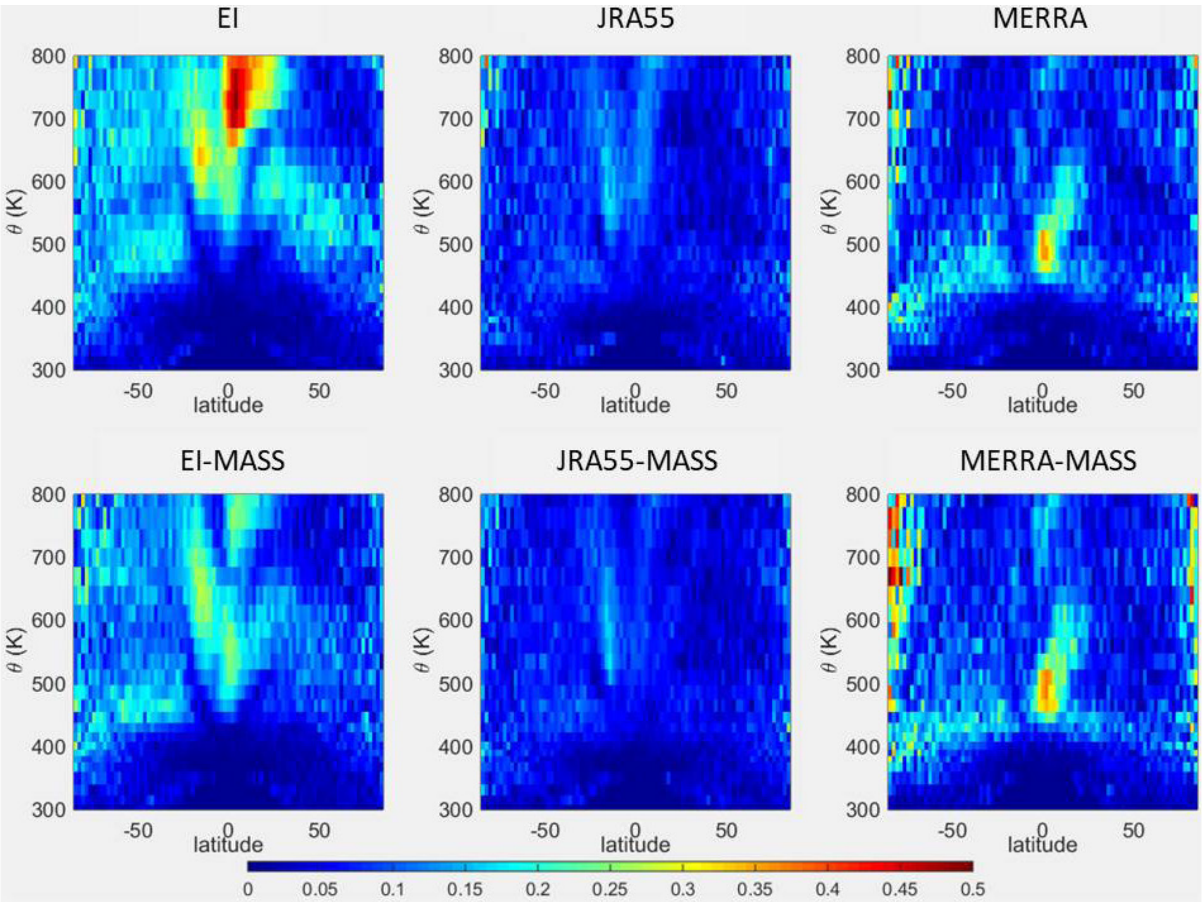


Figure 5.43: *Cross-sections of the amplitude of the correlation of the mean AoA with the QBO signal (defined at the 30 hPa level) in the TRACZILLA model for the ERA-Interim (left), JRA-55 (middle) and MERRA (right) reanalysis without (top) and with (bottom) mass correction. This simualtions cover the period 1989 - 2010.*

Figure 5.43 shows the correlation between the mean AoA and the QBO signal for the TRACZILLA simulations without mass correction (upper rows) and with mass correction (lower rows) for ERA-Interim, JRA-55 and MERRA. ERA-Interim and JRA-55 display approximately the same pattern; however, the amplitude is much stronger for ERA-Interim, reaching 0.5 correlation values over the tropical high stratosphere, while for JRA-55 correlation values stay between 0.0 - 0.2 for all locations. The QBO influence is stronger above the 600 K isentrope, especially over the NH. MERRA also shows a distinct tropical maximum for the correlation, stronger than for JRA-55 and weaker than for ERA-Interim, but the tropical maximum of MERRA is located between 450 K and 500 K at a much lower altitude than for the two other reanalysis. Applying the mass correction has little influence on the pattern of the correlation but reduces its amplitude, especially for ERA-Interim. It is worth noting that mass correction is more important for ERA-Interim than it is for MERRA or JRA-55 which are better balanced.

Figure 5.44 shows averaged timeseries of these stratospheric AOD satellite observations, deseasonalised mean AoA timeseries from CLaMS using ERA-Interim and JRA-55, and residual of the multiple linear regression with and without removal of the AOD signal. It can be seen that for both reanalyses there is a strong positive signal in the mean AoA following the Pinatubo eruption for both reanalyses. For the more recent extratropical volcanic eruptions after 2008, the signal is much smaller and the time lag from the eruption is longer. Therefore, a substantial contribution to decadal variability in the stratospheric circulation, as represented by variability in mean age of air, is caused by volcanic aerosol injections. As shown by *Diallo et al.* (2017), this mean AoA increase after a major volcanic eruption is significantly affected by corresponding induced mixing effects after the eruption. This increase we see in mean AoA is linked on the one hand to an increase in mixing, and on the other hand to a change in the upwelling strength at different levels. *Diallo et al.* (2017) also show that part of the mean AoA positive trend found over the NH for the recent past can be attributed to the minor volcanic eruptions that have taken place after 2008.

Results in **Figure 5.43** are not only due to the differences in AoA but they also point towards differences in the representation of the QBO signal among the different reanalyses. These results agree with other AoA studies looking into QBO effects on AoA, *e.g., Diallo et al.* (2012) or *Chabrillat et al.* (2018, **Figure 10** in their paper, and **Figure 5.36** in this Chapter). A full assessment of the QBO representation in all reanalyses can be found in *Chapter 9* of this Report.

Volcanic aerosols effects

The effects of increases in the stratospheric aerosol loading due to volcanic eruptions on the BDC has been estimated by CLaMS using modelled mean AoA and trends. For this estimation a multiple regression technique accounting for observed stratospheric aerosol has been used (calculation details in *Diallo et al.*, 2017). We have used observed stratospheric aerosol optic depth (AOD) timeseries averaged from 50°S - 50°N over the 1989 - 2012 time period for merged satellites datasets GISS, and SAGE II + GOMOS(525-nm) + CALIPSO (532-nm).

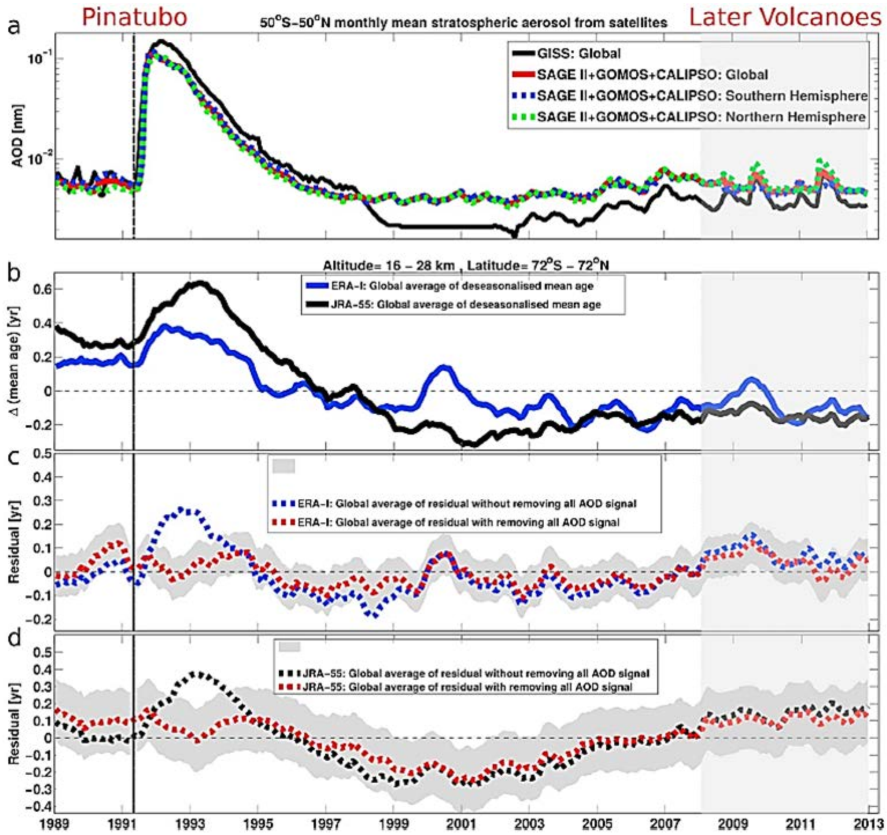


Figure 5.44: *Globally averaged timeseries of the stratospheric AOD, deseasonalised mean AoA and residual of the multiple linear regression with and without removal of all AOD signal. (a) Stratospheric AOD timeseries is averaged from 50°S - 50°N over the 1989 - 2012 time period and is shown for merged satellites datasets (GISS: black and SAGE II+GOMOS(525-nm)+CALIPSO (532-nm): red, blue and green). (b) The deseasonalised mean AoA driven by ERA-Interim and JRA-55 reanalyses is globally averaged between 72°S - 72°N and 16 - 28 km. (c, d) The residual of the multiple linear regression with (red-dashed line) and without (black-dashed line) removing the AOD signal from the deseasonalised mean age (b). The gray shading area indicates the standard deviation. Figure from Diallo et al. (2017). ©American Geophysical Union. Used with permission.*

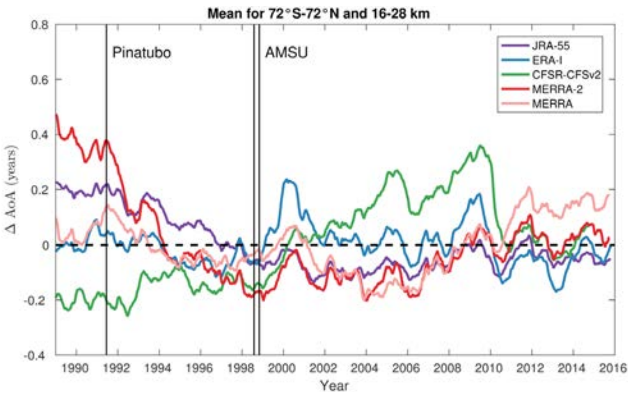


Figure 5.45: *Time evolution of the globally averaged (72° S - 72° N) anomalies of AoA with respect to their mean (1989 - 2015) annual cycles, between 16 km and 28 km, using the five reanalyses with same colour codes as in* ***Figure 5.36****. The black vertical lines highlight the start of the Pinatubo eruption and the first assimilation of AMSU data. From Chabrillat et al. (2018).*

Therefore, the representation of volcanic aerosols is an important element for reanalyses to correctly capture the time evolution of the stratospheric circulation.

Figure 5.45 shows the deseasonalized time series of mean AoA in the extra-polar LS, between 72° S-72° N and 16 - 28 km of altitude. The impact of the Pinatubo eruption is not evident in these BASCOE simulations, while *Diallo et al.* (2017) showed a very clear Pinatubo signal in the AoA time series from CLaMS simulations with ERA-Interim and JRA-55 (**Figure 5.44**). These differences between models can be partly explained by the fact that BASCOE is a kinematic model while CLaMS is a diabatic model. In BASCOE the vertical motion comes from the wind velocity fields while in CLaMS it comes from the diabatic heating rates. Since BASCOE was run in a purely advective mode, it did not take any temperature information from the reanalysis fields. Therefore, this comparison between BASCOE and CLaMS puts into evidence that for transport models to capture the signal from volcanic aerosols using reanalyses fields, radiative or temperature information is explicitely required from the reanalyses, as such signal is not fully present in the reanalyses wind fields. The comparison of results in **Figure 5.44** and **Figure 5.45** therefore shows that ERA-Interim and JRA-55 reanalyses include some volcanic aerosols information in the temperature field, but that wind fields do not contain sufficient information on volcanic signals. Also worth noting that future further investigation comparing volcanic responses in CCMs and CTMs will be needed, as some studies (*e.g.*, *Pitari et al.*, 2016; *Garfinkel et al.*, 2017) have shown different BDC volcanic response in CCM simulations compared to the offline simulations driven by the reanalyses we have considered. Future comparison assessments including CTM results with ERA5 will also be able to provide further information on the impacts of including volcanic aerosol forcing in the model used to produce the reanalysis.

ENSO signal effects

Using a multiple regression method applied to Aura MLS observations and CLaMS model simulations driven by ERA-Interim and JRA-55 reanalysis, we analyse the impact that the El Niño Southern Oscillation (ENSO) signal has on the BDC. **Figure 5.46** shows the zonal mean distribution of the ENSO impact on monthly-mean young and old air mass fractions from CLaMS simulations.

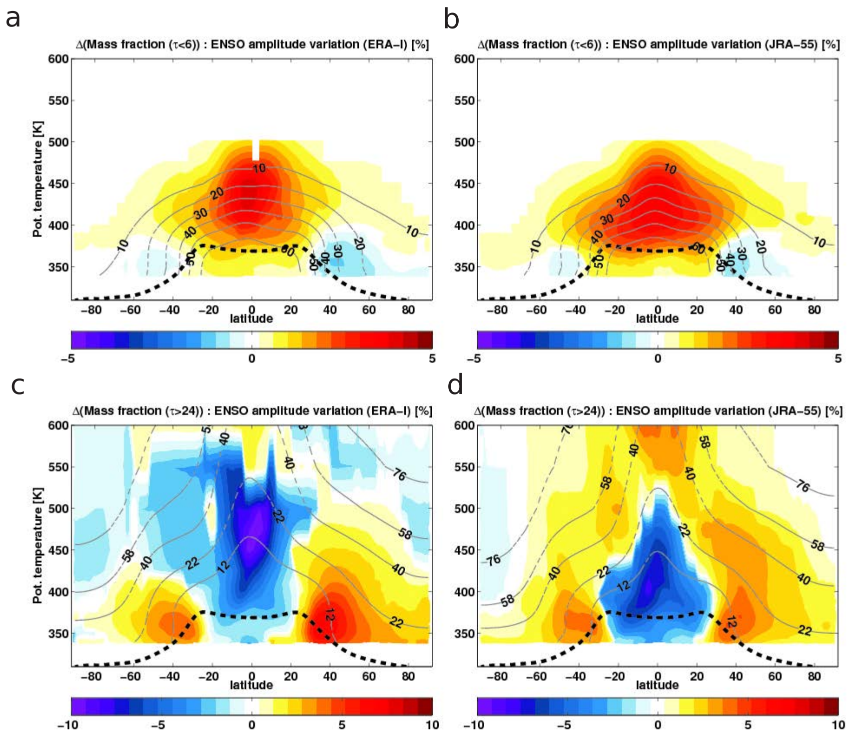


Figure 5.46: *Zonal mean distribution of the ENSO impact on monthly-mean young and old air mass fraction from CLaMS simulations driven by ERA-Interim (left column) and JRA-55 (right column) reanalyses. The amplitude of the air mass fraction variations attributed to ENSO is calculated by using MEI index from the multiple regression fit for the 1981 - 2013 period. (a, b) show the ENSO amplitude variation of the young air mass fraction with transit times shorter than 6 months. (c, d) show ENSO amplitude variation of the old air mass fraction with transit time longer than 24 months. Contours are the climatology values over the 1981 - 2013 period. The black dashed line indicates the tropopause location from reanalyses. Figure from Diallo et al. (2019).*

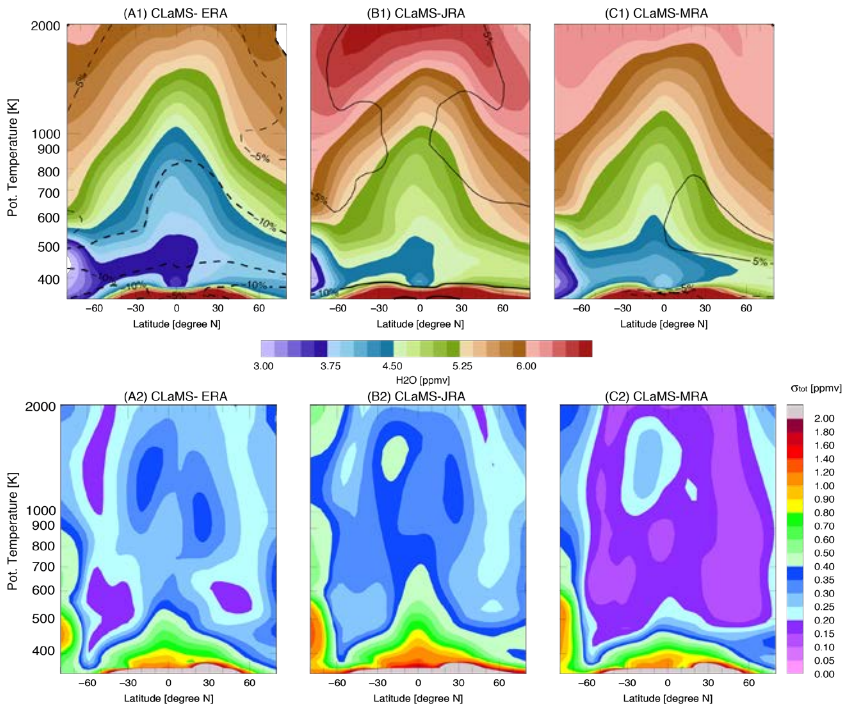


***Figure 5.47:** The zonal and annual mean of water vapor (ppmv) from reanalysis-driven CLaMS simulations, averaged over the period 1980 - 2013 (top panel). In the bottom panel, the total variances (relative to the climatology) of respective monthly means are shown. The black contours show the differences of each CLaMS run relative to the means of (A1), (B1) and (C1). The reanalyses used are ERA-Interim (left), JRA-55 (middle) and MERRA-2 (right).*

The amplitude of the air mass fraction variations attributed to ENSO is calculated by using the Multivariate ENSO index (MEI) from the multiple regression fit for the 1981 - 2013 period (*Wolter et al.*, 1998). The young air mass fraction is defined as that with transit times shorter than 6 months, while the old air mass fraction corresponds to transit times longer than 24 months. Looking into these two fractions gives information on the separate effect the ENSO has on the shallow and the deep branches of the BDC.

During El Niño conditions, the mass fraction of young air increases over the tropical lower stratosphere (up to 4 % increase) while there is a smaller decrease over the extratropical LS region. The structure and amplitude of these changes are in good agreement for ERA-Interim and JRA-55. The changes in the old air mass fraction (lower panels in **Figure 5.46**) show a strong decrease over the tropical tropopause, with a maximum decrease of up to 10 %, located between 450 K - 500 K for ERA-Interim, and of up to 7.5 % for JRA-55 located at lower altitude right above the tropopause at 400 K. The decrease region is much more confined for JRA-55 than for ERA-Interim. In the ECMWF reanalysis the effect of the ENSO signal makes the old-air mass fraction decrease also over middle and high latitudes above 450 K, while in JRA-55 the old-air mass fraction increases everywhere, except for the polar latitudes and the tropics below 500 K. In the extratropical LS region both reanalyses agree, showing regions where the mass fraction of old-air increases, especially over the NH subtropics and midlatitudes.

The ENSO influence on the BDC for ERA-Interim and JRA-55 is more evident for the LS region, below 600 K (~ 24 km), thus it affects the transition and shallow circulation branches of the BDC. During El Niño, the transition branch weakens, while the shallow branch strengthens. Opposite changes occur during La Niña (not shown here). A detailed discussion of these ENSO effects can be found in *Diallo et al.* (2019). Similar patterns are found for ERA-Interim and JRA-55 but the intensity of the effects is different for each reanalysis.

5.5.2.8 Stratospheric water vapour tracer

The zonal annual mean of stratospheric water vapour (SWV) is shown in **Figure 5.47** for CLaMS simulations driven by ERA-Interim, JRA-55 and MERRA-2. These distributions of SWV have been obtained by averaging the model results over the period 1980 - 2013, the total variances with respect to the climatology are also shown in the figure. The overall structure of the climatological annual mean is well captured by the three reanalyses, however there are also several differences among the three simulations. The driest stratosphere corresponds to the simulation with ERA-Interim, the moistest one to JRA-55 and MERRA-2 shows values in between the other two reanalyses: overall, ERA-Interim is 0.75 ppmv drier than JRA-55 and 0.5 ppmv drier than MERRA-2 for all locations. The corresponding total variance distributions for the three simulations show a similar pattern structure, but the magnitude of the variance differs among reanalyses. JRA-55 shows the largest variances, MERRA-2 the lowest ones and ERA-Interim shows in between values more similar to JRA-55 in the NH and to MERRA-2 in the SH. The differences in SWV concentrations are not only due to differences in the stratospheric circulation but also to the entry rates through the TTL, hence to differences in TTL temperatures and mixing processes. From **Figure 5.47** one can see that ERA-Interim already shows the lowest SWV at the tropical tropopause.

Figure 5.48 shows the stratospheric tape-recorder signal based on SWOOSH SWV observations (top panel) and SWV values from the three CLaMS runs, averaged over 20°N-20°S for the period 1980 - 2013.

Upward propagation of the tape-recorder signal between 450 K and 600 K is 0.5 - 1.5 months faster in the ERA-Interim and JRA-55 simulations compared to SWOOSH, and the MERRA-2 simulation is 1 - 1.5 months slower than in SWOOSH. Similarly, the amplitude of the tape-recorder signal is systematically stronger than SWOOSH in the ERA-Interim and JRA-55 simulations, but weaker above 450 K in the one with MERRA-2. These differences are partly attributable to the slower upwelling in MERRA-2 (weaker heating rates as shown in **Figures 5.18** and **5.19**). Slower upwelling not only delays the propagation of the signal but also allows more time for horizontal advection and mixing of middle latitude air into the tropics, which tend to damp the signal.

Tao et al. (2019) also show the strong contribution of CH_4 oxidation in the CLaMS MERRA-2 run, indicated by the blue and red contour lines in **Figure 5.48**. This contribution to the tape-recorder signal is substantially larger than in the other two runs. This feature is a secondary effect of the slow tropical upwelling (in addition to more in-mixing from the extratropics), resulting in a relatively pronounced seasonal cycle in H_2O/CH_4 in CLaMS driven by MERRA-2 with a maximum amplitude of 0.05 ppmv near the 450 K isentrope. The amplitude of H_2OCH_4 in the MERRA-2 run is twice as large as that in the JRA-55 one. The run with ERA-Interim on the other hand, shows virtually no anomalies in H_2O/CH_4 at these levels due to relatively rapid rates of ascent in the lower branch of the BDC.

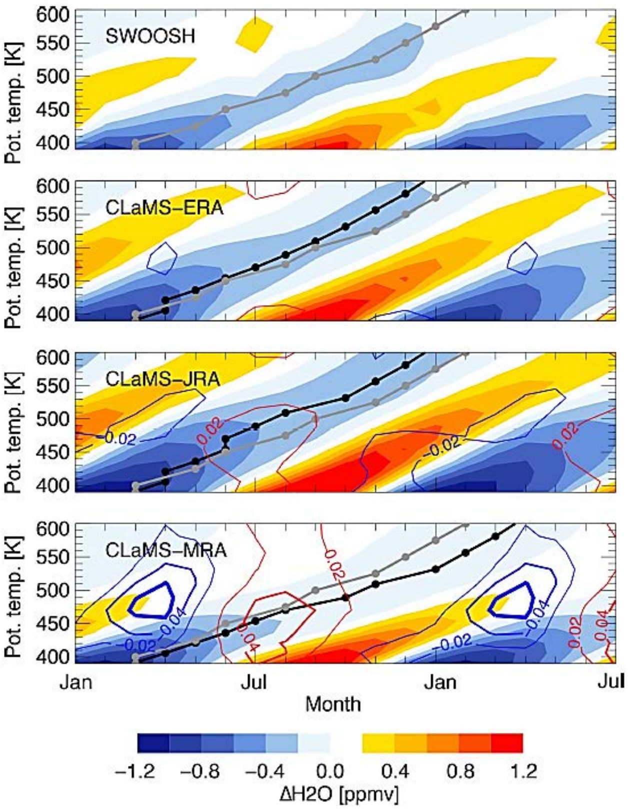


Figure 5.48: *Structure of the stratospheric tape-recorder signal based on SWOOSH observations (top panel) and the three CLaMS runs, averaged over the period 1980 - 2013. The tape-recorder is defined as anomalies in tropical (20° S-20° N) mean H_2O relative to the climatological mean at each level (color shading). The phase of upward propagation (solid black line and circles) is defined by the largest correlation with the layer below. For convenience, propagation based on SWOOSH is included in each panel (grey line). Red and blue contours indicate positive and negative contributions of CH_4 to H_2O anomalies (in units of ppmv, at intervals of 0.02 ppmv). Figure from Tao et al. (2019).*

Figure 5.49 shows the timeseries of the tropical anomalies (averaged between 10° N - 10° S) for water vapour at the 400 K level. Timeseries have been obtained from CLaMS simulations driven by ERA-Interim and JRA-55, and for the overlapping periods are also compared to satellite observations from Halogen Occultation Experiment (HALOE, *Harries et al.*, 1996) and from the Microwave Limb Sounder (MLS, *Waters et al.*, 2006). Both reanalysis products resolve well the subseasonal variability of H_2O fluctuations at the tropical tropopause. The variability on a time scale of 1-3 years (QBO; shaded regions in **Figure 5.49** correspond to easterly QBO phases), as well as on a time scale of 4 - 8 years (ENSO), is better represented with ERA-Interim, especially during the HALOE period (see also *Tao et al.*, 2015).

The lower panel in **Figure 5.49** shows the corresponding mean AoA anomalies in the CLaMS simulations. The decadal variability shows larger differences between ERA-Interim and JRA-55, both for water vapour and mean AoA; JRA-55 shows no trend along the 1979 - 2013 period, while ERA-Interim shows a negative trend for this tropical tropopause region. ENSO and stratospheric volcanic aerosols have been shown to modulate both the tropical ascending branch of the BDC (*e.g.*, **Figure 5.44**; *Diallo et al.*, 2017, 2019) and tropical tropopause temperatures (*e.g.*, *Holton and Gettelman*, 2001; *Mitchell et al.*, 2015), consequently affecting the distribution and evolution of SWV concentrations in the stratosphere.

5.6 Discussion

We have examined how well five modern reanalyses represent the stratospheric Brewer-Dobson circulation (BDC). For this, we have looked into dynamics diagnostics from the reanalyses data and into transport tracers from offline simulations driven by the reanalyses data. Results from both dynamics diagnostics and offline tracers show significant improvements in modern reanalyses compared to previous reanalysis products. This significant improvement in the representation of the BDC in recent reanalysis products reflects the fact that the corresponding agencies have been paying more continuous attention to improve the representation of stratospheric processes (*Fujiwara et al.*, 2017). Our results also show room for future improvement and need for further attention as we discuss later in this section.

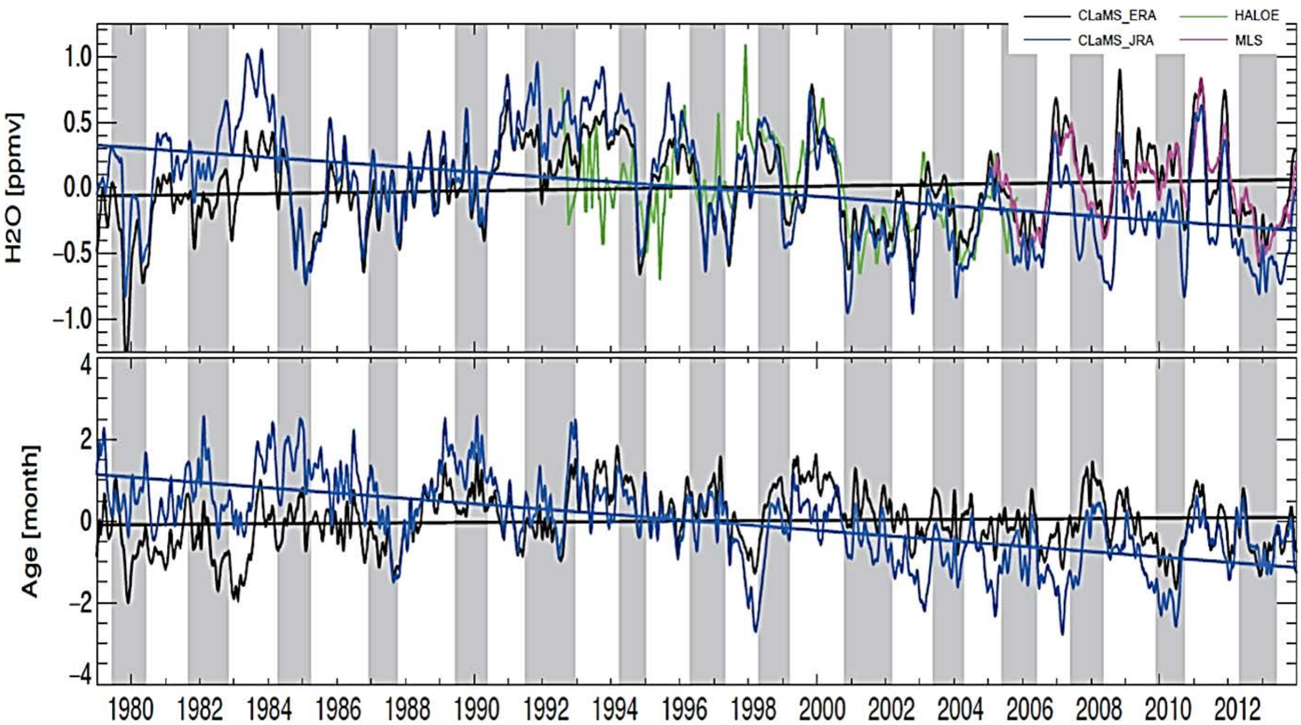


Figure 5.49: *Water vapour (ppmv) tropical anomalies timeseries at 400 K (upper panel) and mean AoA tropical anomalies at 400 K (lower panel) for the period 1979 - 2013 derived from CLaMS simulations driven by ERA-Interim (blue) and JRA-55 (black). Anomalies have been deseasonalised with respect to the 1979 - 2013 climatology and averaged over the tropics (10° N - 10° S) at the 400 K level. Satellite observations are also shown for HALOE (green curve) and MLS (magenta). The corresponding linear trends for the model results are also plotted (straight lines). Grey shading corersponds to the easterly phases of the QBO.*

Our dynamics diagnostics have shown close agreement in terms of climatologies for many derived metrics, such as total tropical upwelling (**Figure 5.8**), although some metrics still show strong disagreement even amongst the most recent products (*e.g.*, upwelling at the equator, **Figure 5.5**). Long-term trends in conventional metrics of BDC strength, such as tropical upwelling, still show disagreement across even the most modern products (**Figure 5.14**). reanalysis products tend to be best constrained in regions and for diagnostics that rely on fundamental balance relations, such as geostrophically balanced flow that couples wind and temperature fields. The mean meridional overturning circulation by definition uses the ageostrophic components of the flow and may therefore be viewed as more prone to uncertainties. In addition, mass conservation is not necessarily strictly fulfilled in reanalysis products due to data assimilation. Our results indicate that the more sophisticated data assimilation schemes employed by modern reanalysis products are less prone to such issues. Nevertheless, most aspects of climatological wave driving, as well as climatological circulation strength and structure are in close agreement (*e.g.*, **Figures 5.4**, **5.11**), especially among the most recent reanalysis products, for which older products showed larger spreads.

An important practical issue for end users of reanalysis products is the vertical resolution of the standard output in the region of the shallow BDC branch. In particular, at least one more output level between 100 hPa and 70 hPa, *i.e.*, the region of strongest vertical gradients in circulation strength, would be necessary to derive more meaningful diagnostics of the shallow BDC branch (such as "outwelling" strength). Note that all modern reanalysis products have at least one model level between 100 - 70 hPa.

Our results from offline simulations have shown that modern reanalyses produce mean AoA in much better agreement with observations than the previous generation of reanalyses (*e.g.*, ERA40). There are however remaining significant discrepancies among reanalyses, and differences with existing observations that imply there is still room for significant improvement in the way reanalyses represent the stratospheric BDC. This means that reanalyses have advanced significantly in the last decades and can still do so in coming ones.

In this Section we discuss possible causes for such discrepancies and point to aspects that need further attention in reanalyses to achieve further improvements in the representation of the BDC. To the extent possible, in the case of diagnostics obtained with CTMs, we also point to CTM model differences that can be causing differences in the results, but since this is not the scope of this Report we do this briefly and refer the interested reader to a more in-depth study we are conducting on this topic (*Monge-Sanz et al.*, in prep.).

All our offline model simulations show decreasing AoA values (strengthening BDC) in the LS region, in agreement with climate models.

However, our offline simulations depict a complex heterogeneous AoA trend in the stratosphere, in agreement with observations and not with most previous climate models studies. There is very good overall agreement between ERA-Interim and JRA-55 but they also show differences, especially in the representation of long-term trends. MERRA and MERRA-2 exhibit too slow vertical transport over the tropics, *e.g.*, as already reflected by the tropical upwelling diagnostic (**Figure 5.8**) and the diabatic heating rates (**Figures 5.18 - 5.20**). This is further shown by the tracer simulations with the offline models, both diabatic and kinematic ones (*e.g.*, **Figures 5.28** and **5.29**), which indicates that the slow BDC bias in the MERRA system is not only related to the radiation budget. The RCTT diagnostic also shows longer residence times for the MERRA datasets but to a much lower extent than the AoA differences, which means that aging by mixing also plays a significant role (*e.g.*, see **Figure 13** in *Ploeger et al.*, 2019). This fact points towards differences among reanalyses in mixing processes across latitudinal barriers (*Stiller et al.*, 2017; *Ploeger et al.*, 2015; *Garny et al.*, 2014).

The best overall agreement with mean AoA observation-based values, both for the climatological value and for trends, is shown using ERA-Interim (*e.g.*, **Figures 5.27**, **5.36)**. This reanalysis dataset is also the only one showing a dipole structure in the mean AoA trend obtained with offline simulations for the MIPAS period (*e.g.*, **Figure 5.38**, **5.40**). This dipole structure is in agreement with the MIPAS satellite observations we have used (*Haenel et al.*, 2015; *Stiller et al.*, 2012), and consistent with some studies explaining other observed tracers' recent past trends (*e.g., Mahieu et al.*, 2014). However we have also shown that AoA trends are very sensitive to the exact period considered and, therefore, future long-term global observations like MIPAS will be essential to understand the evolution of the BDC.

The volcanic signal is not equally present in all reanalyses, and in all simulations. In particular the comparison we have done between BASCOE and CLaMS simulations (**Figures 5.44**, **5.45**) highlights the fact that the volcanic information in the reanalyses is mainly contained in the temperature field, and not in the wind fields, which creates an unrealistic dynamical mismatch among different fields in one same reanalysis dataset. This result points towards the need of a more interactive representation of volcanic aerosols in the reanalyses. ERA5 includes a more realistic treatment of volcanic aerosols than previous reanalyses and it will be necessary to compare the results from offline simulations included in this Chapter with equivalent ones driven by ERA5 fields, to assess associated improvements in the BDC representation.

Here we summarise several possible causes for the discrepancies we have found among reanalyses, and therefore aspects that require further attention in future reanalyses:

Clouds and convection:

The different ways in which reanalyses include the radiative effects of clouds and the parameterisation of convection has also an impact on the tropical entry rates and tropical upwelling of the BDC. MERRA and MERRA-2 have strong cooling during summer in the TTL that tends to block transport, while in ERA-Interim diabatic motion is too fast due to the heating effect of cirrus clouds (*Tegtmeier et al.*, 2019). Deep convection also impacts the tropical UTLS wave activity and therefore the modelled BDC. A detailed comparison of clouds and convection treatment in all major reanalyses and their impact on the TTL is included in *Chapter 8* of this Report. Also a relevant study was conducted with several ECMWF reanalyses and operational analyses (*Feng et al.*, 2011) and should be further investigated with reanalyses from other Centres regarding their impact on wave activity and the BDC.

Gravity wave drag:

ERA-Interim, JRA-55 and CFSR all neglect non-orographic gravity wave drag (except for CFSv2, *i.e.*, CFSR after 2010) and each one uses its own parameterisation of orographic gravity wave drag. MERRA and MERRA-2 use the same parameterisation for orographic gravity wave drag and both take non-orographic gravity wave drag into account. In all the CTM studies we have shown here, MERRA and MERRA-2 provide significantly older AoA than the three other reanalyses. Different parameterisations of gravity wave drag are therefore a possible modelling cause for the disagreements in the stratospheric circulation diagnostics (*e.g.*, *Dharmalingam et al.*, 2019; *Podglajen et al.*, 2016). Since the recent ERA5 reanalysis includes non-orographic gravity wave drag, future comparisons using ERA5 driven simulations will provide further insight on related impacts on the representation of the BDC.

Heat budgets and radiation schemes:

Differences in heat budgets in the tropical region have substantial implications for the representation of transport and mixing in the LS region (*e.g.*, *Wright and Fueglistaler*, 2013). *Abalos et al.* (2015) evaluated the vertical component of the advective BDC in ERA-Interim, MERRA and JRA-55 and found large differences between direct (*i.e.*, kinematic) estimates and indirect estimates derived from the thermodynamic balance (*i.e.,* using diabatic heating rates). TRACZILLA and CLaMS simulations shown in this Chapter have used the reanalyses diabatic heating rates, and their differences in mean AoA are consistent with the differences in the diabatic heating rates fields. Younger AoA values are linked to larger diabatic heating rate values, and viceversa, and also the differences in the amplitude of the annual cycle in AoA follow the differences in the diabatic heating rates annual cycle shown in **Figure 5.18**.

However, the differences among reanalyses are also clearly displayed by offline simulations with kinematic transport models (*e.g.*, BASCOE), indicating that differences are not only coming from differences in the heating rates field. Different radiation schemes and treatment of stratospheric radiative species, as well as differences in the assimilated observations, produce differences in the reanalyses temperature field. Differences in temperature distribution and latitudinal gradients result in differences in the stratospheric wind fields. This will affect offline simulations of the BDC even for simulations that do not use the temperature field from reanalyses, *e.g.,* BASCOE kinematic simulations.

Ozone and Water vapour:

One reason why the temperature field differs among different reanalysis models and radiation codes is the different treatment of stratospheric ozone and water vapour. *Fueglistaler et al.* (2009) already showed that unrealistic or oversimplified ozone descriptions in the reanalysis systems lead to unrealistic radiative heating rates. *Chapter 4* in this Report and *Davis et al.* (2017) provide a thorough comparison of the ozone and water vapour distributions provided by the different reanalyses and gives an overview of the way these two components are treated in the different reanalyses radiation codes. ERA-Interim uses an ozone climatology, JRA-55 uses time-varying ozone fields from an external CCM and MERRA-2 uses interactive ozone. We recommend an assessment of the impacts that different ozone and water vapour modelling approaches in the reanalysis systems have on the representation of the stratospheric circulation. A study looking into how different treatments of stratospheric ozone impact stratosphere-troposphere processes in the ECMWF system has been recently carried out (*Monge-Sanz et al.*, 2020); extending this type of study to other major reanalysis systems would provide useful information.

Resolution and resolved mixing, and top of the model:

For differences in the results between reanalyses, we also need to keep in mind that the original grids of the reanalyses are different, and that interpolating to the CTMs' resolution has different numerical effects for each reanalysis. This will also affect mixing processes and their impact on mean AoA values differently for each reanalysis. Additionally, the altitude of the top of the model and the treatment of the top boundary sponge layer is different among reanalysis systems; this also has an effect on the BDC and on offline simulations for the stratosphere. Extending the altitude of the top of the model and including mesospheric processes into the reanalysis systems would improve the representation of the BDC. We also note that different top boundary conditions imposed in the offline models can be partly causing differences in the age-of-air values obtained with CLaMS (which imposes a top boundary condition to match MIPAS AoA values in the top level) and TRACZILLA (which uses removal of trajectories above a certain potential temperature level or an age limit to trajectories).

QBO representation:

The Quasi-Biennial Oscillation (QBO) signal is not equally captured and represented in the different reanalyses. Therefore the way the QBO links with, and influences the meridional circulation is different in each dataset. In our TRACZILLA simulations we have seen that the QBO correlation with the age-of-air diagnostic largely differs among datasets. These differences may well be linked to differences in the parameterization of non-orographic gravity wave drag. This deserves further investigation, especially in the case of MERRA-2, which shows difficulties representing correctly the QBO before 1995. A comprehensive analysis of the QBO representation in the different reanalyses is found in *Chapter 9* of this Report.

Volcanic influence:

How the different reanalyses capture the influence of large volcanic eruptions is linked to the different representation of aerosols, and to what information goes into the assimilated fields. In ERA-Interim and JRA-55 the effects of stratospheric volcanic aerosol are only included by the assimilation of observed temperature and wind data, as discussed in more detail by *Diallo et al.* (2017), whereas MERRA-2 additionally assimilates aerosol optical depth (*Fujiwara et al.*, 2017). Our offline simulations have shown that the analysed temperature field contains information on the volcanic signal, but that wind fields do not carry enough information about this signal (*Chabrillat et al.*, 2018; *Diallo et al.*, 2017). This fact points to potential dynamical mismatches between temperature and winds in the reanalyses, probably due to high assimilation increment values associated to the volcanic eruption effects. In addition, persistent imbalance will generate spurious gravity waves that artificially strengthen the BDC in the models.

To quantify how much each of these differences contributes to the discrepancy among reanalyses, and how much it contributes to disagreement with observations, tailored experimental datasets from Reanalyses Centres would be needed that do not exist at present. For the ECMWF reanalysis system, one study was conducted using tailored datasets to evaluate different aspects of the Data Assimilation system (assimilation window length, assimilation technique) and the model resolution (*Monge-Sanz et al.*, 2012).

Apart from the processes we have discussed above, there are of course other major processes, in different parts of the Earth System, that influence the BDC, including the ENSO signal or the stratospheric polar dehydration (*Chapter 10* in this Report). And we need to keep in mind that all the mentioned processes actually interact with each other, some of the interaction mechanisms are known while others are still a matter of international investigation efforts. In order to achieve a BDC representation that is more realistic, Reanalysis Centres and models will need to continue to move to a representation of the Earth System that is more complete and more coupled in coming years.

5.7 Conclusions and recommendations

In this Chapter we have analysed different diagnostics for the Brewer-Dobson circulation (BDC) for major reanalyses participating in S-RIP.

We have performed a direct comparison of dynamical diagnostics from the reanalyses datasets, including EP-flux divergence, tropical upwelling and outwelling, and residual circulation trajectories (RCTTs). We have also performed transport tracers simulations with different offline chemistry-transport models (CTMs) driven by the reanalyses, and assessed distribution of several tracers, mean age-of-air (AoA) and age spectrum diagnostics.

5.7.1 Conclusions from dynamics diagnostics

The dynamical diagnostics indicate that the BDC is much more consistent in the more recent reanalysis products, with much reduced spread in the respective climatologies compared to the older products. Furthermore, the BDC is generally less strong in more recent products compared to their older versions. However, even these recent products show significant differences in basic climatological diagnostics in some fields (*e.g.*, shallow branch wave driving, tropical upwelling structure and seasonality, upwelling strength below 70 hPa). Nevertheless, for the dynamical diagnostics analysed here the reanalysis products also show overall remarkable agreement with current chemistry-climate models (CCMs).

Time series of annual mean tropical upwelling mass flux at 70 hPa, a common measure of BDC strength used in many modelling studies (*e.g.*, Butchart et al. 2010), show a fairly strong degree of co-variability amongst the recent products (correlation coefficients between 0.65 - 0.82), except for CFSR. This and time series of other dynamical diagnostics suggests spurious fluctuations in CFSR; this product should therefore not be used for long-term trend or interannual variability analyses (consistent with the transport diagnostics in *Section 5.7.2*, see below).

Although MERRA-2, JRA-55, ERA-Interim and ERA5 agree with regards to co-variability on interannual time scales, there is inconsistency with regards to their long-term trend estimates of tropical upwelling at 70 hPa. MERRA-2 and JRA-55 show acceleration, while ERA-Interim shows deceleration, and ERA5 does not show a statistically significant trend. This also holds true at other pressure levels throughout the tropical lower stratosphere. A similar picture emerges for the poleward mass transport through the turnaround latitudes ("tropical outwelling"), although ERA-Interim in this case does not show a statistically significant opposing trend to MERRA-2 and JRA-55 (which both show a long-term strengthening of the circulation). However, the co-variability on interannual time scales is strongly reduced for this metric compared to upwelling, with correlation coefficients only in the range 0.23 - 0.68 (ERA-Interim among the lowest values). This is perhaps due to large sensitivity to structural differences (including those due to GWD) and suggests that the shallow branch of the BDC is not well constrained, even in modern products.

The RCTT diagnostic offers an integrated view of the circulation strength, possibly more robust to inconsistencies and uncertainties amongst products. The global mean RCTT at 50 hPa, a common reference level used for AoA comparisons, does show a high degree of co-variability among modern products (correlation coefficients between 0.53 - 0.85), but also shows large offsets in total values especially in the 1980's. Long-term trend values in this metric qualitatively agree with those obtained from tropical upwelling, including the disagreement between ERA-Interim and MERRA-2/JRA-55. An inspection of the latitudinally and vertically resolved RCTT trends shows that, by and large, RCTTs decrease (consistent with acceleration of the BDC), except for some regions/data sets. The main exception to this general behaviour are the RCTT trends corresponding to the deep branch of the BDC in both ECMWF reanalyses (ERA-Interim, ERA5). However, even these ECMWF products show primarily negative RCTT trends in the lowermost stratosphere, consistent with a strengthening of the shallow branch of the BDC.

5.7.2 Conclusions from transport tracers simulations

Although the dynamical diagnostics allow a clear comparison among reanalyses, they cannot be compared against observed quantities. We have also performed transport tracers simulations with different offline chemistry-transport models (CTMs) driven by the reanalyses. These sets of simulations have allowed us to compare results against observation-based data for the mean age-of-air (AoA) and stratospheric water vapour (SWV). For these diagnostics we have compared mean distributions as well as time series and evolution of trends for the different reanalysis products.

Our comparison results have shown that recent reanalyses produce mean AoA in much better agreement with observations than the previous generation of reanalysis (*e.g.*, ERA-Interim v. ERA-40), showing the improvement achieved by the reanalysis systems in the representation of the BDC.

However significant discrepancies in AoA and tracers distribution among reanalyses still remain. The spread of AoA obtained with different reanalyses can be as large as among different CCMs (*e.g.*, *Orbe et al.*, 2020).

We have shown that differences in the heating rates field are evident among the reanalyses we have considered, with MERRA reanalyses particularly differing from the rest. Heating rates differences are a major factor affecting the offline simulations of stratospheric tracers with diabatic models. MERRA and MERRA-2 exhibit too slow vertical transport over the tropics, in agreement with the lower values they show for diabatic heating rates compared to the other reanalyses. But the slow tropical transport is shown both by diabatic and kinematic offline simulations, which indicates that the slow BDC bias in the MERRA system is not only related to the radiation budget. The RCTT diagnostic also shows longer residence times for the MERRA datasets.

We have devoted a significant part of the Chapter to quantify mean AoA trends in the stratosphere, to better understand to what extent reanalyses can be used to study changes in the BDC structure and strength. For the overall period (1989 - 2010) our offline results show large spread in values and sign of AoA trends, depending on the reanalysis and on the region of the stratosphere. For the MIPAS period (2002 - 2012) only ERA-Interim is in good agreement with the observed trends, independently of the offline model used. The positive trend in the mean AoA in the NH is a robust feature in our studies and is in agreement with other observed phenomena like HCl observed trends (*Mahieu et al.*, 2014).

Here we need to note that much investigation is still needed on BDC trends, and that trends should be interpreted with caution as many factors affect them, including natural variability and changes in the observation system of assimilated data that make them so sensitive to the particular period chosen (*e.g.*, *Chabrillat et al.*, 2018).

The large spread in AoA results among reanalyses indicate two main aspects: *i)* important differences among the underlying models in the different reanalyses systems, and *ii)* that assimilated observations are not providing a strong constraint for stratospheric transport in reanalyses. As we indicate below in *Section 5.7.3,* we strongly recommend reanalyses centres to invest in model development in order to further improve the representation of the BDC.

We have also discussed in Section *5.5.2.7* how the AoA diagnostic is affected by other Earth System phenomena, not only in the stratosphere like the QBO signal, but also the ENSO and the volcanic signals. This shows the need to include as many Earth System processes as possible in a realistic way to achieve a more accurate BDC representation in future reanalyses.

With one of our offline CTMs (CLaMS) we performed a comparison of SWV distribution using the water model tracer. In this case, the distributions obtained with the different reanalyses showed good overall consistency for climatology and variability in the CTM, but differences against independent observations.

5.7.3 Recommendations to reanalyses users

A summary of the usability of major reanalyses in terms of their representation of the BDC can be found in **Figure 5.50**, where we classify the performance of each reanalysis for the diagnostics we have considered in this Chapter, based on the results and discussions we have included in the above sections.

Although not all the diagnostics we have used can be evaluated against observations, we have decided to assign an evaluation score to all of them. Such value, for those that cannot be compared to observations, reflects their consistency with other processes and our current understanding of the BDC.

In the majority of cases our evaluation is that reanalyses are "suitable with limitations". Such limitations depend on the particular time periods, atmospheric regions and applications. For instance, MERRA-2 is likely not to be a good option for years before 1995. MERRA-2, compared to ERA-Interim and JRA-55, shows difficulties in representing the QBO before 1995 (*Chabrillat et al.*, 2018 and references therein). *Gelaro et al.* (2017) also describe several features in MERRA-2, not present in the other two reanalyses, that can affect stratospheric dynamics, and therefore BDC diagnostics, including the assimilation of Microwave Limb Sounder on the Aura satellite (Aura-MLS) temperatures, from 2004 onwards and above 5 hPa.

Among the recent reanalyses, only in the case of CFSR we have identified several issues that indicate that their use may be problematic for stratospheric BDC studies, especially related to interannular variability and long-term trends. For older reanalyses like ERA-40 and NCEP reanalyses, it had already been shown in numerous published studies that their representation of the BDC, and other stratospheric processes, is unrealistic and, therefore, we also discourage their use for stratospheric studies.

Whenever possible we generally recommend users not to restrict themselves to only one product when it comes to BDC studies. In particular for the period after 2000 a comparison between MERRA-2, JRA-55, and ERA-Interim, together with new products such as ERA5 and JRA-3Q, can help to distinguish robust from non-robust diagnostics results. We also recommend working with reanalyses data on model levels, not only for offline simulations, but also for diagnostics related to the shallow BDC branch as usually no pressure levels are provided between 100 - 70 hPa.

5.7.4 Recommendations to reanalyses centres

From the results and experiences built along this study, this is our list of main recommendations for the development and data release of future reanalysis.

Regarding data availability:

- Provide variables' uncertainty information.
- Provide variables at higher vertical resolution, especially around the UTLS region.
- Provide pressure level data above 1 hPa (important for RCTT calculations).
- Archive data at higher frequencies.
- Archive additional relevant variables by default (e.g. heating rates).

The recently released ERA5 includes most of the above features, although the resolution around the UTLS is still lower than desired.

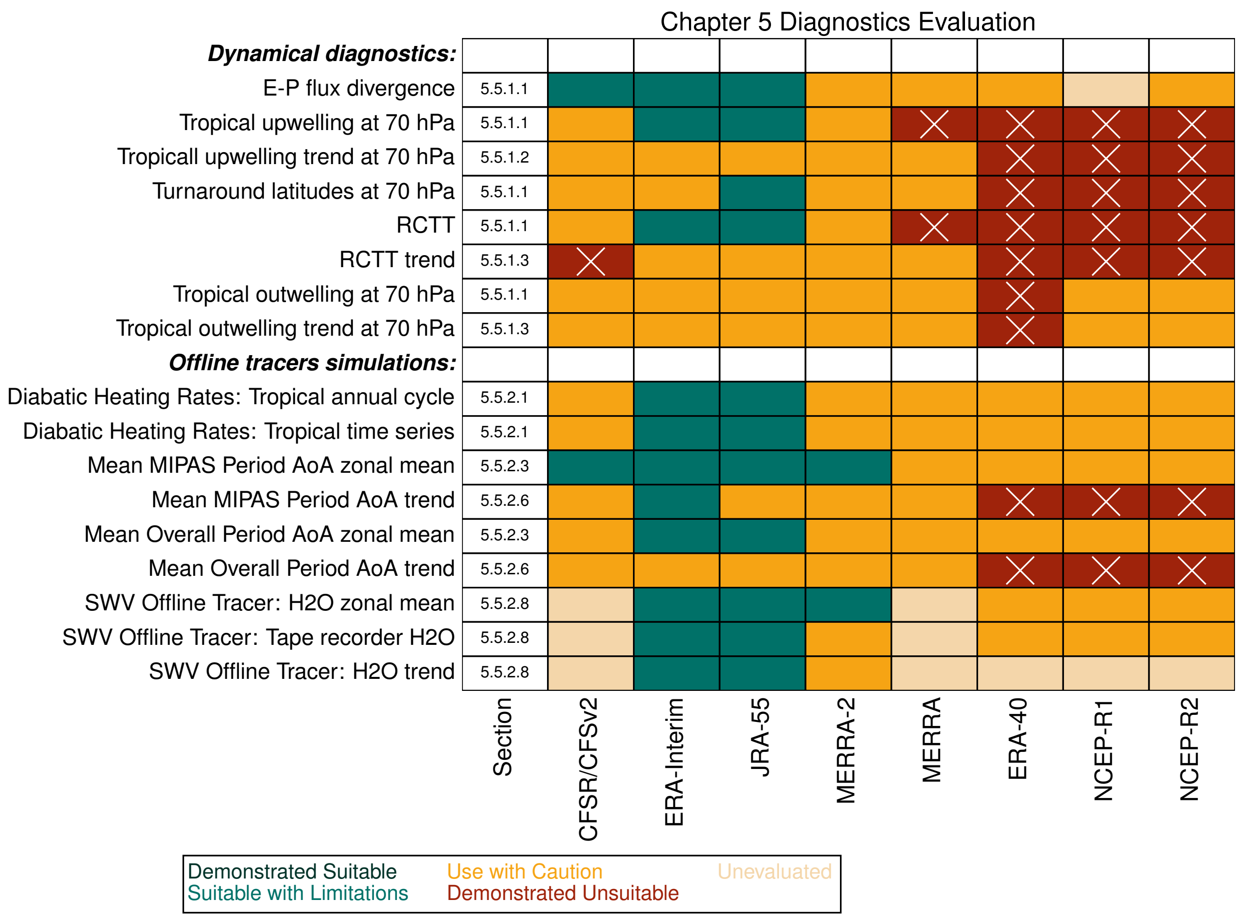


Figure 5.50: *Summary of the BDC diagnostics evaluation. Note that the score corresponding to "demonstrated suitable" was not assigned to any of the diagnostics listed here, so the darkest green colour does not appear in this table.*

From early experiences with ERA5, dealing with its huge volume of data requires improved postprocessing strategies and/or more computing/storage power. Interactive communication channels between reanalyses users and producers to improve sustainable solutions will likely become more important in the future as more high volume data products will be available.

Besides continuous assimilation of stratospheric winds as suitable datasets become available (*e.g.,* from the ESA's recent AEOLUS mission), model development stands out from our study as a major recommendation among the actions required to improve the representation of the stratospheric BDC in future reanalyses. Main model aspects that require attention are:

- Gravity wave drag parameterisations
- Representation of radiative gases and aerosols in the stratosphere
- Clouds and convection parameterisations
- Increase of the model resolution in the UTLS
- Extension of the vertical range to incorporate mesospheric processes.

Last but not least, sustained long-term relevant observations platforms are required to monitor any changes in the strength and the structure of the BDC and, therefore, to keep evaluating how well future reanalyses represent stratospheric major circulation patterns. We strongly advocate for the creation of such observation platforms and the necessity to keep them operative for long enough time periods to cover the relevant time scales to validate BDC evolution and trends.

Code & Data availability

Reanalysis data used in this chapter can be obtained from the corresponding reanalyses centres. Observations datasets and offline model data are available upon request and via the referenced publications. The dynamics diagnostics shown are based on the zonal mean dataset produced by *Martineau et al.* (2018) as referenced in the text; code is available upon request.

Acknowledgements

We thank the reviewers of this chapter, whose comments helped to improve the final revised version. BMS acknowledges partial funding from the UK Natural Environment Research Council (NERC) through the ACSIS project (North Atlantic Climate System Integrated Study) led by the National Centre for Atmospheric Science (NCAS); TB acknowledges funding by the National Science Foundation (grant no. AGS-1643167).

Figure 5.21 is adapted from Cook and Roscoe (2009). **Figure 5.22** is adapted from Engel et al. (2017). **Figures 5.23** and **5.24** are adapted from *Haenel et al.* (2015). F**igures 5.27, 5.28, 5.36, 5.39** and **5.45** are adapted from *Chabrillat et al.* (2018). *Figures 5.29, 5.30* and **5.33** are adapted from *Ploeger et al.* (2019). **Figure 5.46** is adapted from *Diallo et al.* (2019). **Figure 5.48** is adapted from *Tao et al.* (2019). All these reproductions are made under a creative commons attribution 3.0 or 4.0 license (**https://creativecommons.org/licenses/by/3.0/** or **https://crertivecommons.org/licenses/by/4.0/**, respectively).

References

Abalos, M., B. Legras, F. Ploeger, and W.J. Randel, 2015: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979–2012. *J. Geophys. Res.*, **120**, 7534 - 7554, doi: 10.1002/2015JD023182.

Abalos M., *et al.*, 2019: New insights on the impact of ozone depleting substances on the Brewer-Dobson circulation. *J. Geophys. Res: Atmospheres*, **124**, 2435 - 2451, doi: 10.1029/2018JD029301.

Andrews, D.G., J.R. Holton, and C.B. Leovy, 1987: Middle Atmosphere Dynamics, Academic Press, doi: 10.1017/S0016756800014333.

Andrews, A.E., *et al.*, 2001: Mean ages of stratospheric air derived from in situ observations of CO_2, CH_4, and N_2O. *J. Geophys. Res.*, **106** (D23), 32 295 - 32 314, doi: 10.1029/2001JD000465.

Birner, T. and H. Bönisch, 2011: Residual circulation trajectories and transit times into the extratropical lowermost stratosphere. *Atmos. Chem. Phys.*, **11**, 817 - 827, doi: 10.5194/acp–11–817–2011.

Boering, K.A., *et al.*, 1996: Stratospheric mean ages and transport rates from observations of carbon dioxide and nitrous oxide. *Science* **274**, 1340 - 1343 doi: 10.1126/science.274.5291.1340.

Brewer, A.W., 1949: Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Q. J. R. Meteorol. Soc.* **75**, 351 - 363, doi: 10.1002/qj.49707532603.

Butchart, N., *et al.*, 2006: Simulations of anthropogenic change in the strength of the Brewer–Dobson circulation. *Clim. Dyn.*, **27**, 727 - 741, doi: doi.org/10.1007/s00382-006-0162-4.

Butchart, N., *et al.*, 2010: Chemistry–Climate Model Simulations of Twenty-First Century Stratospheric Climate and Circulation Changes. *J. Climate*, **23**, 5349 - 5374,doi: 10.1175/2010JCLI3404.1

Butchart, N., 2014: The Brewer-Dobson circulation. *Rev. Geophys.*, **52**, 157 -184, doi: 10.1002/2013RG000448.

Chabrillat, S., *et al.*, 2018: Comparison of mean age of air in five reanalyses using the BASCOE transport model. *Atmos. Chem. Phys.*, **18**, 14715 - 14735, doi: 10.5194/acp-18-14715-2018.

Chipperfield, M.P., W. Feng, and M. Rex, M., 2005: Arctic ozone loss and climate sensitivity: Updated three-dimensional model study. *Geophys. Res. Lett.*, **32**, L11813, doi: 10.1029/2005GL022674.

Chipperfield, M.P., 2006: New version of the TOMCAT/SLIMCAT off-line chemical transport model. *Q. J. R. Meteorol. Soc.* **132**, 1179 - 1203, doi: 10.1256/qj.05.51.

Chipperfield, M.P., *et al.*, 2014: Multimodel estimates of atmospheric lifetimes of long-lived ozone-depleting substances: Present and future. *J. Geophys. Res. Atmos.*, **119**, 2555 - 2573, doi: 10.1002/2013JD021097.

Cook, P.A. and H.K. Roscoe, 2009: Variability and trends in stratospheric NO_2 in Antarctic summer, and implications for stratospheric NO_Y. *Atmos. Chem. Phys.* **9**, 3601 - 3612, doi: 10.5194/acp-9-3601-2009.

Cook, P.A. and H.K. Roscoe, 2012: Changes in reactive stratospheric gases due to a change in Brewer-Dobson circulation: results from a simple model. *Atmos. Sci. Lett.* **13**, 49 - 54, doi: 10.1002/asl.362, 2012.

Chrysanthou, A., *et al.*, 2019: The effect of atmospheric nudging on the stratospheric residual circulation in chemistry–climate models. *Atmos. Chem. Phys.*, **19**, 11559 - 11586, doi: 10.5194/acp-19-11559-2019.

Davis S.M., *et al.*, 2016: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: a long-term database for climate studies. *Earth Syst Sci Data*. **8**(2):461 - 490. doi: 10.5194/essd-8-461-2016.

Davis, S.M., *et al.*, 2017: Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP. *Atmos. Chem. Phys.*, **17**, 12 743 - 12 778, doi: 10.5194/acp-17-12743-2017, 2017.

Dee, D.P., *et al.*, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, **137**, 553 - 597, doi: 10.1002/qj.828

Dessler, A.E., *et al.*, 2014: Variations of stratospheric water vapor over the past three decades. *J. Geophys. Res. Atmos.*, **119**, 12,588 - 12,598, doi: 10.1002/2014JD021712.

Dharmalingam, S. *et al.*, 2019: Accuracy of Balloon Trajectory Forecasts in the Lower Stratosphere. *Atmosphere,* **10**, 102, doi: 10.3390/atmos10020102.

Diallo, M., B. Legras, and A. Chédin, 2012: Age of stratospheric air in the ERA-Interim. *Atmos. Chem. Phys.*, **12**, 12133 - 12154, doi: 10.5194/acp-12-12133-2012.

Diallo, M., *et al.*, 2017: Significant Contributions of Volcanic Aerosols to Decadal Changes in the Stratospheric Circulation. *Geophys. Res. Lett.*, **44**, 10,780 - 10,791. doi: 10.1002/2017GL074662.

Diallo, M., *et al.*, 2019: Structural changes in the shallow and transition branch of the Brewer–Dobson circulation induced by El Niño. *Atmos. Chem. Phys.*, **19**, 425 - 446, doi: 10.5194/acp-19-425-2019.

Dietmüller, S., 2018: Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCMI-1 models. *Atmos. Chem. Phys.*, **18**, 6699 - 6720, doi: 10.5194/acp-18-6699-2018.

Dobson, G.M.B., 1956: Origin and distribution of the polyatomic molecules in the atmosphere. *Proceedings of the Royal Society of London* **A236**, 187 - 193, doi: 10.1098/rspa.1956.0127.

Elkins, J.W., *et al.*, 1996: Airborne gas chromatograph for in situ measurements of longlived species in the upper troposphere and lower stratosphere. *Geophys. Res. Lett.* **23**, 347 - 350, doi: 10.1029/96GL00244.

Engel, A., *et al.*, 2009: Age of stratospheric air unchanged within uncertainties over the past 30 years. *Nature Geosci.*, **2**, 28 - 31, doi: 10.1038/ngeo388.

Engel, A., *et al.*, 2017: Mean age of stratospheric air derived from AirCore observations. *Atmos. Chem. Phys.*, **17**, 6825 - 6838, doi: 10.5194/acp-17-6825-2017.

Eyring, V., *et al.*, 2006: Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *J. Geophys. Res.*, **111**, D22308, doi: 10.1029/2006JD007327.

Feng, W., *et al.*, 2007: Midlatitude ozone changes: studies with a 3-D CTM forced by ERA-40 analyses. *Atmos. Chem. Phys.*, **7**, 2357 - 2369, doi: 10.5194/acp-7-2357-2007.

Feng, W., *et al.*, 2011: Evaluation of cloud convection and tracer transport in a three-dimensional chemical transport model, *Atmos. Chem. Phys.*, **11**, 5783 - 5803, doi: 10.5194/acp-11-5783-2011.

Fisher, M., A. O'Neill, and R. Sutton, 1993: Rapid descent of mesopheric air into the stratospheric polar vortex. *Geophys. Res. Lett.*, **20**, 1267 - 1270, doi: 10.1029/93GL01104.

Fischer, H., *et al.*, 2008: MIPAS: an instrument for atmospheric and climate research. *Atmos. Chem. Phys.*, **8**, 2151 - 2188, doi: 10.5194/acp-8-2151-2008.

Fu, Q., P. Lin, S. Solomon, and D.L. Hartmann, 2015: Observational evidence of strengthening of the Brewer-Dobson circulation since 1980. *J. Geophys. Res. Atmos.*, **120**, 10,214 - 10,228, doi: 10.1002/2015JD023657.

Fueglistaler, S. , *et al.*, 2009: The diabatic heat budget of the upper troposphere and lower/mid stratosphere in ECMWF reanalyses. *Q.J.R. Meteorol. Soc.*, **135**: 21 - 37. doi: 10.1002/qj.361.

Fujiwara, M., et al., 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp-17-1417-2017.

Garfinkel, C.I., V. Aquila, D.W. Waugh, and L.D. Oman, 2017: Time-varying changes in the simulated structure of the Brewer–Dobson Circulation. *Atmos. Chem. Phys.*, **17**, 1313 - 1327, doi: 10.5194/acp-17-1313-2017.

Garny, H., T. Birner, H. Bönisch, and F. Bunzel, 2014: The effects of mixing on age of air. *J. Geophys. Res. Atmos.*, **119**, 7015 - 7034, doi: 10.1002/2013JD021417.

Gelaro, R., *et al.*, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate*, **30**, 5419 - 5454, doi: 10.1175/JCLI-D-16-0758.1.

Hall, T.M. and R.A. Plumb, 1994: Age as a diagnostic of stratospheric transport. *J. Geophys. Res.-Atmos.*, **99**, 1059 - 1070, doi: 10.1029/93JD03192.

Hall, T.M., D:W. Waugh, K.A. Boering, and R.A. Plumb, 1999: Evaluation of transport in stratospheric models. *J. Geophys. Res.*, **104**, 18815 - 18839, doi: 10.1029/1999JD900226.

Haenel, F.J., et al., 2015: Reassessment of MIPAS age of air trends and variability. *Atmos. Chem. Phys.*, **15**, 13 161 - 13 176, doi: 10.5194/acp-15-13161-2015.

Hardiman, S.C., *et al.*, 2017: The influence of dynamical variability on the observed Brewer-Dobson circulation trend. *Geophys. Res. Lett.*, **44**, 2885 - 2892, doi: 10.1002/2017GL072706.

Harnisch, J., R. Borchers, P. Fabian, and M. Maiss, 1996: Tropospheric trends for CF_4 and C_2F_6 since 1982 derived from SF_6 dated stratospheric air. *Geophys. Res. Lett.* **23**, 1099 - 1102, doi: 10.1029/96GL01198.

Harries, J.E., *et al.*, 1996: Validation of measurements of water vapor from the Halogen Occultation Experiment (HALOE). *J. Geophys. Res.*, **101**, 10205 - 10216, doi: 10.1029/95JD02933.

Holton, J.R., 1982: The Role of Gravity Wave Induced Drag and Diffusion in the Momentum Budget of the Mesosphere. *J. Atmos. Sci.*, **39**, 791 - 799, doi: 10.1175/1520-0469(1982)039<0791:TROGWI>2.0.CO;2

Holton, J.R. and A. Gettelman, 2001: Horizontal transport and the dehydration of the stratosphere. *Geophys. Res. Lett.*, **28**, 2799 - 2802, doi: 10.1029/2001GL013148.

Iwasaki, T., H. Hamada, and K. Miyazaki, 2009: Comparisons of Brewer-Dobson Circulations Diagnosed from Reanalyses. *J. Meteor. Soc. Japan*, **87**, 997 - 1006, doi: 10.2151/jmsj.87.997.

Kida, H., 1983: General circulation of air parcels and transport characteristics derived from a hemispheric GCM, Part 2, Very long-term motions of air parcels in the troposphere and stratosphere. *J. Meteorol. Soc. Jpn.* **61**, 510 - 522, doi: 10.2151/jmsj1965.61.2_171.

Kobayashi, S., *et al.*, 2015: The JRA–55 Reanalysis: General specifications and basic characteristics. *J. Meteorol. Soc. Jpn.*, **93**, 5 - 48, doi: 10.2151/jmsj.2015-001.

Konopka, P., *et al.*, 2004: Mixing and Ozone Loss in the 1999–2000 Arctic Vortex: Simulations with the 3-dimensional Chemical Lagrangian Model of the Stratosphere (CLaMS). *J. Geophys. Res.*, **109**, D02315, doi: 10.1029/2003JD003792.

Konopka, P., *et al.*, 2015: Hemispheric asymmetries and seasonality of mean age of air in the lower stratosphere: Deep versus shallow branch of the Brewer-Dobson circulation. *J. Geophys. Res. Atmos.*, **120**, 2053 - 2066. doi: 10.1002/2014JD022429.

Kouker, W., I. Langbein, T. Reddmann, and R. Ruhnke, 1999: The Karlsruhe Simulation Model of The Middle Atmosphere Version 2. *Wiss. Ber.* **FZKA 6278**, Forsch. Karlsruhe, Karlsruhe, Germany, doi: 10.5445/IR/270045162.

Kovács, T., *et al.*, 2017: Determination of the atmospheric lifetime and global warming potential of sulfur hexafluoride using a three-dimensional model. *Atmos. Chem. Phys.*, **17**, 883 - 898, doi: 10.5194/acp-17-883-2017.

Lin, S.-J. and R.B. Rood, 1996: Multidimensional Flux-Form Semi-Lagrangian Transport Schemes. *Mon. Weather Rev.*, **124**, 2046 - 2070, doi: 10.1175/1520-0493(1996)124<2046:mffslt>2.0.co;2.

Linz, M., R.A. Plumb, E.P. Gerber, and A. Sheshadri, 2016: The Relationship between Age of Air and the Diabatic Circulation of the Stratosphere. *J. Atmos. Sci.*, **73**, 4507 - 4518, doi: 10.1175/JAS-D-16-0125.1.

Mahieu, E., *et al.*, 2014: Recent Northern Hemisphere stratospheric HCl increase due to atmospheric circulation changes. *Nature*, **515**, 104 - 107, doi: 10.1038/nature13857.

Martineau, P., J.S. Wright, N. Zhu, and M. Fujiwara, 2018: Zonal-mean data set of global atmospheric reanalyses on pressure levels. *Earth Syst. Sci. Data*, **10**, 1925 - 1941, doi: 10.5194/essd-10-1925-2018.

McKenna, D.S., et al., 2002: A new Chemical Lagrangian Model of the Stratosphere (CLaMS): 1. Formulation of advection and mixing. *J. Geophys. Res.*, **107**, 4309, doi: 10.1029/2000JD000114.

Meijer, E.W., B. Bregman, A. Segers, and P.F.J. van Velthoven, 2004: The influence of data assimilation on the age of air calculated with a global chemistry-transport model using ECMWF wind fields. *Geophys. Res. Lett.*, **31**, L23114, doi: 10.1029/2004GL021158.

Ming, A., P. Hitchcock, and P. Haynes, 2016: The Double Peak in Upwelling and Heating in the Tropical Lower Stratosphere. *J. Atmos. Sci.*, **73**, 1889 - 1901, doi: 10.1175/JAS-D-15-0293.1.

Mitchell, D., *et al.*, 2015: Signatures of naturally induced variability in the atmosphere using multiple reanalysis datasets. *Q. J. R. Meteorol. Soc.*, **141**, 2011 - 2031.

Miyazaki, K., *et al.*, 2016: Inter-comparison of stratospheric mean-meridional circulation and eddy mixing among six reanalysis data sets. *Atmos. Chem. Phys.*, **16**, 6131 - 6152, doi: 10.5194/acp-16-6131-2016.

Monge-Sanz, B.M. and M.P. Chipperfield, 2006: Chemical Transport Modelling. *Proceedings of the ECMWF/GEO Workshop on Atmospheric Reanalyses*, UK, 19 - 22 June 2006, ECMWF, pp. 39-44.

Monge-Sanz, B.M., M.P. Chipperfield, A.J. Simmons, and S.M. Uppala, 2007: Mean age of air and transport in a CTM: Comparison of different ECMWF analyses. *Geophys. Res. Lett.*, **34**, L04801, soi: 10.1029/2006GL028515.

Monge-Sanz, B.M., *et al.*, 2012: Improvements in the stratospheric transport achieved by a chemistry transport model with ECMWF (re)analyses: identifying effects and remaining challenges. *Q. J. Roy. Meteor. Soc.*, **139**, 654 - 673.

Monge-Sanz B.M., *et al.*, 2013: On the uses of a new linear scheme for stratospheric methane in global models: water source, transport tracer and radiative forcing. *Atmos. Chem. Phys.*, **13**, 9641 - 9660, doi: 10.5194/acp-13-9641-2013.

Monge-Sanz, B.M., *et al.*, 2020: A stratospheric prognostic ozone for seamless Earth System Models: performance, impacts and future. *Atmos. Chem. Phys. Discuss.*, doi: 10.5194/acp-2020-1261.

Mote, P.W., *et al.*, 1996: An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor. *J. Geophys. Res.*, **101**, 3989 - 4006, doi: 10.1029/95JD03422.

Mote, P.W., *et al.*, 1998: Vertical velocity, vertical diffusion, and dilution by midlatitude air in the tropical lower stratosphere. *J. Geophys. Res.* **103**, 8651 - 8666, doi: 10.1029/98JD00203.

Orbe, C., *et al.*, 2020: Description and Evaluation of the specified-dynamics experiment in the Chemistry-Climate Model Initiative. *Atmos. Chem. Phys.*, **20**, 3809 - 3840, doi: 10.5194/acp-20-3809-2020.

Pisso, I. and B. Legras, 2008: Turbulent vertical diffusivity in the sub-tropical stratosphere. *Atmos. Chem. Phys.*, **8**, 697 - 707, doi: 10.5194/acp-8-697-2008.

Pitari, G., *et al.*, 2016: Impact of stratospheric volcanic aerosols on age-of-air and transport of long-lived species. A*tmosphere*, **7**(11), 149, doi: 10.3390/atmos7110149.

Ploeger, F., *et al.*, 2010: Impact of the vertical velocity scheme on modeling transport in the tropical tropopause layer, *J. Geophys. Res.*, **115**, D03301, doi: 10.1029/2009JD012023.

Ploeger, F., et al., 2015: Variability of stratospheric mean age of air and of the local effects of residual circulation and eddy mixing. J. *Geophys. Res. Atmos.*, **120**, doi:10.1002/2014JD022468.

Ploeger, F. and T. Birner, 2016: Seasonal and inter-annual variability of lower stratospheric age of air spectra. *Atmos. Chem. Phys.*, **16**, 10195 - 10213, doi: 10.5194/acp-16-10195-2016.

Ploeger, F., *et al.*, 2019: How robust are stratospheric age of air trends from different reanalyses? *Atmos. Chem. Phys.*, **19**, 6085 - 6105, doi: 10.5194/acp-19-6085-2019.

Plumb, R.A., 2002: Stratospheric transport. *J. Meteor. Soc. Japan*, **80**, 793 - 809, doi: 10.2151/jmsj.80.793.

Podglajen, A., A. Hertzog, R. Plougonven, and B. Legras, 2016: Lagrangian temperature and vertical velocity fluctuations due to gravity waves in the lower stratosphere. *Geophys. Res. Lett.*, **43**, 3543 - 3553, doi: 10.1002/2016GL068148.

Polvani L., *et al.*, 2019: Large impacts, past and future, of ozone-depleting substances on Brewer-Dobson circulation trends: A multi-model assessment. *J. Geophys. Res: Atmospheres*, **124**, 6669 - 6680 doi: 10.1029/2018JD029516.

Pommrich, R., *et al.*, 2014: Tropical troposphere to stratosphere transport of carbon monoxide and long-lived trace species in the Chemical Lagrangian Model of the Stratosphere (CLaMS). *Geosci. Model Dev.*, **7**, 2895 - 2916, doi: 10.5194/gmd-7-2895-2014.

Prather, M.J., 1986: Numerical advection by conservation of second-order moments. *J. Geophys. Res.,* **91**, 6671 - 6681, doi: 10.1029/JD091iD06p06671.

Ravishankara, A.R., S. Solomon, A.A. Turnipseed, and R.F. Warren, 1993: Atmospheric Lifetimes of Long-Lived Halogenated Species. *Science*, **259**, 194 - 199, doi: 10.1126/science.259.5092.194.

Ray, E.A., *et al.,* 1999: Transport into the Northern Hemisphere lower-most stratosphere revealed by in situ tracer measurements. *J. Geophys.Res.*, **104**, 26,565 - 26,580, doi: 10.1029/1999JD900323.

Ray, E.A., *et al.,* 2014: Improving stratospheric transport trend analysis based on SF_6 and CO_2 measurements. *J. Geophys. Res. Atmos.*, **119**, 14,110 - 14,128, doi: 10.1002/2014JD021802.

Reddmann, T., R. Ruhnke, and W. Kouker, 2001: Three-dimensional model simulations of SF6 with mesospheric chemistry. *J. Geophys. Res.*, **106**, 14525 - 14537, doi: 10.1029/2000JD900700.

Roscoe H.K., A.J. Charlton D.J. Fish, and J.G.T. Hill, 2001: Improvements to the accuracy of measurements of NO_2 by zenith-sky visible spectrometers II: errors in zero using a more complete chemical model. *J. Quant. Spectrosc. Ra.*, **68**, 337 - 349, doi: 10.1016/S0022-4073(00)00058-3.

Roscoe, H.K., 2004: A review of stratospheric H_2O and NO_2. *Adv. Space Res.*, **34**, 1747 - 1754, doi: 10.1016/j.asr.2003.01.025.

Rosenlof, K.H., 1995: Seasonal cycle of the residual mean meridional circulation in the stratosphere. *J. Geophys. Res.*, **100**, 5173 - 5191, doi: 10.1029/94JD03122.

Rotman, D.A., *et al.*, 2001: Global Modeling Initiative assessment model: Model description, integration, and testing of the transport shell. *J. Geophys. Res.*, **106**, 1669 - 1691, doi: 10.1029/2000JD900463.

Sato, K. and S. Hirano, 2019: The climatology of the Brewer–Dobson circulation and the contribution of gravity waves. *Atmos. Chem. Phys.*, **19**, 4517 - 4539, doi: 10.5194/acp-19-4517-2019.

Scheele, M.P., P.C. Siegmund, and P.F.J. van Velthoven, 2005: Stratospheric age of air computed with trajectories based on various 3D-Var and 4D-Var data sets. *Atmos. Chem. Phys.* **5**, 1-7, doi: 10.5194/acp-5-1-2005.

Schoeberl, M.R. and A.E. Dessler, 2011: Dehydration of the stratosphere. *Atmos. Chem. Phys.*, **11**, 8433 - 8446, doi: 10.5194/acp-11-8433-2011.

Schlutow, M., E. Becker, and H. Körnich, 2014: Positive definite and mass conserving tracer transport in spectral GCMs. *J. Geophys. Res. Atmos.*, **119**, 11,562 - 11,577, doi: 10.1002/2014JD021661.

Segers, A., van P. Velthoven, B. Bregman, and M. Krol, 2002: On the Computation of Mass Fluxes for Eulerian Transport Models from Spectral Meteorological Fields. In: Sloot P.M.A., Hoekstra A.G., Tan C.J.K., Dongarra J.J. (eds) Computational Science — ICCS 2002. ICCS 2002. *Lecture Notes in Computer Science*, vol **2330**. Springer, Berlin, Heidelberg. doi:10.1007/3-540-46080-2_81.

Skachko, S., *et al.*, 2014: Comparison of the ensemble Kalman filter and 4D-Var assimilation methods using a stratospheric tracer transport model. *Geosci. Model Dev.*, **7**, 1451 - 1465, doi: 10.5194/gmd-7-1451-2014.

SPARC, 2010: SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models. V. Eyring, T. Shepherd and D. Waugh (Eds.), *SPARC Report* **No. 5**, *WCRP-***30/2010**, *WMO/TD* – **No. 40**, available at **https://www.sparc-climate.org/publications/sparc-reports/**.

Stiller, G.P., *et al.*, 2012: Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period. *Atmos. Chem. Phys.*, **12**, 3311 - 3331, doi: 10.5194/acp-12-3311-2012.

Stiller, G.P., *et al.*, 2017: Shift of subtropical transport barriers explains observed hemispheric asymmetry of decadal trends of age of air. *Atmos. Chem. Phys.*, **17**, 11177 - 11192, doi: 10.5194/acp-17-11177-2017.

Tao, M., *et al.*, 2015: Impact of stratospheric major warmings and the quasi-biennial oscillation on the variability of stratospheric water vapor. *Geophys. Res. Lett.*, **42**, 4599 - 4607, doi: 10.1002/2015GL064443.

Tao, M., *et al.*, 2019: Multitimescale variations in modeled stratospheric water vapor derived from three modern reanalysis products. *Atmos. Chem. Phys.*, **19**, 6509 - 6534, doi: 10.5194/acp-19-6509-2019.

Tegtmeier *et al.*, 2020: Temperature and tropopause characteristics from reanalyses data in the tropical tropopause layer. *Atmos. Chem. Phys.*, **20**, 753 - 770, doi: 10.5194/acp-20-753-2020.

Uppala, S. M., *et al.*, 2005: The ERA-40 Re-analysis. *Q. J. R. Meteorol. Soc.*, **131**, 2961 - 3012, doi: 10.1256/qj.04.176.

van Noije, T.P.C., H.J. Eskes, M. van Weele, and P.F.J. van Velthoven, 2004: Implications of the enhanced Brewer-Dobson circulation in ECMWF reanalysis ERA-40 for the stratosphere-troposphere exchange of ozone in global chemistry transport models. *J. Geophys. Res.*, **109**, D19308, doi: 10.1029/2004JD004586.

Waters, J.W., *et al.*, 2006: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite. *IEEE T. Geosci. Remote*, **44**, 1075 - 1092, doi: 10.1109/TGRS.2006.873771.

Waugh, D.W. and T.M. Hall, 2002: Age of stratospheric air: Theory, observations, and models. *Rev. Geophys.*, **40**, 1 - 27, doi: 10.1029/2000RG000101.

Waugh, D.W., 2009: The age of stratospheric air. *Nat. Geosci.*, **2**, 14 - 16, doi: 10.1038/ngeo397.

Wolter, K., and M.S. Timlin, 1998: Measuring the strength of ENSO events: How does 1997/98 rank? *Weather*, **53**, 315 - 324, doi: 10.1002/j.1477-8696.1998.tb06408.x.

Wright, J. and S. Fueglistaler, 2013: Large differences in reanalyses of diabatic heating in the tropical upper troposphere and lower stratosphere. *Atmos. Chem. Phys.*, **13**, 9565 - 9576 doi: 10.5194/acp-13-9565-2013.

Major abbreviations and terms

AoA	Age-of-air
AOD	aerosol optical depth
AMSU	Advanced Microwave Sounding Unit
BAS	British Antarctic Survey
BDC	Brewer-Dobson Circulation
CCM	Chemistry Climate Model
CFSR	Climate Forecast System Reanalysis
CMAM	Canadian Middle Atmosphere Model
CCMI	Chemistry-Climate Model Initiative
CCMVal	Chemistry Climate Model Validation
CTM	Chemistry-transport model
DAS	Data assimilation system
DOE	Department of Energy
DJF	December-January-February
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño Southern Oscillation
EP-flux	Eliassen-Palm Flux
EPFD	Eliassen-Palm Flux Divergence
ERA-20C	ECMWF 20th century reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA-Interim	ECMWF interim reanalysis
ERA5	the fifth major global reanalysis produced by ECMWF

FFSL	Flux-Form Semi-Lagrangian
GEOSCCM	NASA Goddard Chemistry-Climate Model
GWD	Gravity Wave Drag
HALOE	Halogen Occultation Experiment
HATS	Halocarbons and other Atmospheric Trace Species
JJA	June-July-August
JRA-25	Japanese 25-year Reanalysis
JRA-55	Japanese 55-year Reanalysis
KASIMA	Karlsruhe Simulation of the Middle Atmosphere
LS	Lower stratosphere
MEI	Multivariate ENSO index
MERRA; MERRA-2	Modern Era Retrospective Analysis for Research and Applications / Version 2
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
MMC	Mean Meridional Circulation
MMM	Multi-Model Mean
MRM	Multi-Reanalysis Mean
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NCEP-DOE R2	Reanalysis 2 of the NCEP and DOE
NCEP-NCAR R1	Reanalysis 1 of the NCEP and NCAR
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
OMS	Observations of the Middle Stratosphere
POLARIS	Photochemistry of Ozone Loss in the Arctic Regions in Summer
QBO	Quasi-Biennial Oscillation
RCTT	Residual Circulation Transit Time
REM	Multi-Reanalysis Mean
SH	Southern Hemisphere
SST	Sea Surface Temperature
SSW	Sudden Stratospheric Warming
StratoClim	Stratospheric and upper tropospheric processes for better climate predictions
SWV	Stratospheric Water Vapour
TTL	Tropical Tropopause Layer
TOA	Top of Atmosphere
UKMO	United Kingdom Meteorological Office
UTLS	Upper troposphere and lower stratosphere
WMO	World Meteorological Organization

Chapter 6: Extratropical Stratosphere–Troposphere Coupling

Chapter lead authors

Edwin P. Gerber	Courant Institute of Mathematical Sciences, New York University	USA
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan

Co-authors

Blanca Ayarzagüena	(1)Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid (2)Instituto de Geociencias, CSIC-UCM	Spain
David Barriopedro	Instituto de Geociencias, CSIC-UCM	Spain
Thomas J. Bracegirdle	British Antarctic Survey	UK
Amy H. Butler	National Oceanic and Atmospheric Administration/Chemical Sciences Laboratory	USA
Natalia Calvo	Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid	Spain
Steven C. Hardiman	Met Office Hadley Centre	UK
Peter Hitchcock	Department of Earth and Atmospheric Sciences, Cornell University	USA
Maddalen Iza	Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid	Spain
Ulrike Langematz	Institut für Meteorologie, Freie Universität Berlin	Germany
Hua Lu	British Antarctic Survey	UK
Gareth Marshall	British Antarctic Survey	UK
Andrew Orr	British Antarctic Survey	UK
Froila M. Palmeiro	(1)Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid (2)Barcelona Supercomputing Center	Spain
Seok-Woo Son	School of Earth and Environmental Sciences, Seoul National University	South Korea
Masakazu Taguchi	Department of Earth Science, Aichi University of Education	Japan

Abstract. This chapter assesses the representation of the two-way coupling between the troposphere and the stratospheric polar vortices in the reanalysis products. This coupling is evaluated over a broad range of time scales, from sub-seasonal to decadal, with a particular emphasis on Sudden Stratospheric Warming (SSW) events, which are among the clearest manifestations of coupling between the tropospheric and stratospheric circulations. Coupled variability on synoptic to seasonal time scales is evaluated by comparing the timing, evolution, and dynamical consistency of SSW events and Final Warming events, and the representation of the Annular Mode indices. Variability on interannual time scales is evaluated by comparing the modulation of sub-seasonal stratosphere-troposphere coupling by El Niño-Southern Oscillation (ENSO) and the Quasi-Biennial Oscillation (QBO). Finally, variability on decadal time scales is evaluated by comparing atmospheric circulation trends driven by the depletion of stratospheric ozone over Antarctica.

As the large-scale circulation cannot easily be characterized from direct observations, this chapter has largely focused on the consistency between the reanalyses, asking the question: would the characterization of stratosphere-troposphere coupling provided by a given reanalysis differ from that provided by another? The internal self consistency of reanalyses has also been evaluated, allowing for more objective grading of the reanalyses. In the satellite era, there is generally good agreement among full-input reanalyses (which assimilate all available observations, including satellite measurements) on stratosphere-troposphere coupling on synoptic to interannual time scales. In addition, conventional-input reanalyses (which exclude satellite observations, and hence full-input reanalyses before the introduction of satellites) are fairly consistent as far back as 1958 in the Northern Hemisphere. There is, however, demonstrable evidence of improvement in the more recent reanalyses. While results in prior studies based on older reanalyses will generally not be significantly different from comparable results based on the modern reanalyses, due to large sampling uncertainty, we strongly recommend that users discontinue use of older reanalyses such as NCEP R1, NCEP R2 and ERA-40 since they provide limited data (*i.e.*, lower model top) and are biased with respect to modern products.

The dominance of sampling uncertainty implies that our assessment of stratosphere-troposphere coupling is limited by the length of the reanalysis records. Consequently, the availability of high quality pre-satellite era reanalysis in the Northern Hemisphere reduces our uncertainty in the tropospheric response to SSWs by approximately 20%.

Among the more modern reanalyses, a consistent trend in the coupled stratosphere-troposphere circulation is found, associated with ozone loss in the Southern Hemisphere. Caution should always be employed in the assessment of decadal variations and trends in stratosphere-troposphere coupling, however, due to changes in the observational network. It is also shown that uncertainties in older and conventional-input reanalyses increase with height, particularly above 10 hPa, and that satellite observations appear to be critical for an assessment of stratosphere-tropospheric coupling in the austral hemisphere. Finally, surface-input reanalyses have also been evaluated. While they should not be used in place of a full-input reanalysis, there is evidence that ERA-20C captures a substantial fraction of the variability between the troposphere and stratosphere, and so may be valuable for research into low frequency variations in stratospheric-troposphere coupling.

Contents

6.1 Introduction 223
6.2 Context and background 223
6.3 Reanalysis datasets 225
6.4 Sudden stratospheric warming events 225
6.4.1 Identifying SSW events 226
6.4.2 Characterizing SSW events 226
6.4.3 Sampling uncertainty vs. reanalysis uncertainty 231
6.4.4 Assessing the internal consistency of SSW events in reanalyses 233
6.5 Annular modes 235
6.5.1 Consistency of the annular mode index across reanalyses in the post and pre-satellite periods....235
6.5.2 Sampling uncertainty vs. reanalysis uncertainty 238
6.6 Stratospheric final warming events 240
6.7 Modulation of stratosphere-troposphere coupling by ENSO and QBO 240
6.7.1 Troposphere-stratosphere coupling through ENSO 242
6.7.2 Blocking patterns associated to SSWs and the modulation of ENSO 243
6.7.3 Nonlinear modulation of the extratropical stratosphere by ENSO and QBO 243
6.8 Stratosphere-troposphere coupling through the antarctic ozone hole 246
6.9 Outlook, key findings, and recommendations 249
References 255
Appendix A: Detection and classification of major SSW events 260
Major abbreviations and terms 264

6.1 Introduction

In this chapter, we assess the representation of coupling between the troposphere and stratosphere across all reanalyses, with a focus on interaction through the stratospheric polar vortices. While this coupling is primarily manifested on daily to seasonal time scales, low frequency modulation by other modes of internal variability (*e.g.*, the Quasi-Biennial Oscillation) and external forcings (*e.g.*, stratospheric ozone loss) require an analysis across a wide range of time scales. The global nature of these low frequency changes also requires consideration of links between variability in the tropics and extratropics.

Our focus on the influence of the stratosphere on tropospheric weather and variability presented two challenges to this chapter. First, this report has sought to evaluate reanalyses against direct measurements, ideally measurements that are not assimilated into the reanalyses themselves. The large-scale weather and variability of the troposphere, however, is not easily characterized or verified with single measurement records. We have attempted to compare with observation-constrained measures where available, but generally, this chapter evaluates the consistency of the reanalyses, or lack thereof, as opposed to verifying them against some objective standard.

A second challenge that we face in this chapter are limitations to our understanding imposed by the natural variability of the atmosphere. A common theme is the relative importance of sampling errors, associated with the finite length of the reanalysis records, compared to the differences between the reanalyses themselves. We term the latter a "reanalysis uncertainty", to differentiate it from the sampling uncertainty. While we find evidence for an improvement in more recent reanalysis products, overall we find that our characterization of stratosphere-troposphere coupling is dominated by sampling uncertainty. As such, the choice of one reanalysis over another would not affect the scientific conclusions of a particular study, with certain exceptions, *e.g.*, the use of restricted input reanalyses, as documented below.

Sampling uncertainty can appear in subtle ways. Stratosphere-troposphere coupling is often evaluated through the analysis of events that are identified by threshold criteria, *e.g.*, a Sudden Stratospheric Warming (SSW) is identified by a reversal of the winds at 10 hPa and 60°. As a result, subtle differences between reanalyses can lead to the identification and examination of different events. This effectively aliases sampling error into a comparison of reanalyses, giving a false impression of disagreement between different reanalysis products. To address this concern, we suggest the use of a uniform set of events when evaluating different reanalysis products.

After a brief introduction to stratosphere-troposphere coupling (*Section 6.2*), we describe the reanalysis datasets in *Section 6.3*. We then present our methodology for identifying, characterizing, and evaluating SSW events in *Section 6.4*. Stratosphere-troposphere coupling on daily to seasonal time scales is further evaluated in *Sections 6.5* and *6.6*, where we evaluate the representation of the annular modes and final warming events, respectively. *Section 6.7* then examines the modulation of stratosphere-troposphere coupling on interannual time scales by El Niño-Southern Oscillation (ENSO) and the Quasi-Biennial Oscillation (QBO). *Section 6.8* compares the representation of the vertical coupling forced by ozone depletion over interdecadal time scales. Finally, *Section 6.9*, provides a summary of our results and conclusions and a compact list of key findings and recommendations.

6.2 Context and background

The troposphere and the stratosphere, the two lowermost layers of the Earth's atmosphere, contain together about 99% of the atmospheric mass. The troposphere is the portion of the atmosphere in close contact with Earth's surface. It is the region where day-to-day weather systems evolve and impact human life; in this sense it could be viewed as a boundary layer, albeit one that occupies roughly 80 - 90% of the atmospheric mass. The stratosphere is found from about 10 - 16 km, depending on the latitude, to about 50 km above the surface (*Andrews et al.*, 1987). What sets these two layers apart is mainly the stability of the layers: whereas temperature decreases with height in the troposphere at a rate of about 7 K per kilometer - making it nearly neutral to moist convection - stratospheric temperatures increase with height owing to the absorption of ultraviolet radiation by ozone. This stratification gave the "sphere of layers" its name.

The stratosphere's large stability sets it dynamically apart from the troposphere as it prevents the penetration of atmospheric convection from the surface, and inhibits the propagation and growth of baroclinic disturbances that make up a great fraction of tropospheric weather. Yet, depending on the season, it can be a dynamically active region subject to large variability. Large equator-to-pole temperature gradients favor the formation of strong westerly vortices in the winter stratosphere (*Waugh et al.*, 2017). These strong westerlies act as a window for the propagation of tropospheric disturbances, allowing planetary-scale waves to go through while preventing the propagation of synoptic-scale systems (*Charney and Drazin*, 1961).

When planetary-scale waves propagate vertically from the troposphere to the stratosphere, they interact with the mean flow and sometimes break (*McIntyre and Palmer*, 1983, 1984) causing an irreversible mixing of potential vorticity leading to a long-lasting weakening of the westerly winds. One of the most extreme examples of stratospheric variability, Sudden Stratospheric Warming (SSW) events, which are characterized by an abrupt deceleration and reversal of the zonal-mean zonal wind, are the result of such interactions between planetary-scale waves and the stratospheric vortex (*Matsuno*, 1971; *Limpasuvan et al.*, 2004; *Polvani and Waugh*, 2004).

The vortices in the Northern Hemisphere and Southern Hemisphere are known to behave quite differently. While the Northern Hemisphere vortex is often disturbed by SSW events in December - January - February, the Southern Hemisphere vortex is more quiescent. These differences are attributable mainly to differences in topography and land-sea temperature contrasts which are known to generate stronger planetary-scale waves in the Northern Hemisphere (*Plumb*, 1989, 2010; *Randel*, 1988). Because of the comparatively weaker wave drag in the Southern Hemisphere, zonal winds are too strong to allow vertical propagation of waves which limit wave-mean flow interactions and the variability of the vortex (*Plumb*, 1989).

As mentioned earlier, a large fraction of stratospheric variability is the result of temporal fluctuations in planetary-scale wave propagation from the troposphere to the stratosphere. It is therefore of great importance to understand how these waves are amplified or reduced in the troposphere. *Garfinkel and Hartmann* (2010) has shown that the intensification of wavenumber-1 and wavenumber-2 waves in the Northern Hemisphere are important precursors of stratospheric polar vortex weakening. One specific tropospheric circulation pattern, atmospheric blocking, has garnered particular attention due to its ability to modulate planetary-scale wave fluxes. *Nishii et al.* (2011), for instance, have shown that there are preferred regions where upward-propagating wave packets from blocking events can interfere constructively with stationary waves to produce large bursts of upward-propagating wave activity, ultimately causing SSW events. The role of such interference in modulating stratospheric variability was also discussed in *Smith and Kushner* (2012). The coupling between the stratosphere and the troposphere is not limited to an upward coupling where the evolution of the stratosphere is influenced by upward-propagating waves. The coupling is actually two-way. Events of weak stratospheric vortex anomalies, such as SSW events, were shown to affect weather at the surface by, notably, favoring the negative phase of the related North Atlantic Oscillation (NAO) and Northern Annular Mode (NAM) patterns, and shifting the storm track southward (*Baldwin and Dunkerton*, 2001). This coupling, which is often attributed to balance arguments (*e.g.*, *Black*, 2002; *Haynes et al.*, 1991) and eddy feedback mechanisms (*e.g.*, *Kushner and Polvani*, 2004; *Song and Robinson*, 2004), has implications for the predictability of tropospheric weather. Mounting evidence suggests that the state of the stratosphere influences the skill of numerical weather forecasts (*e.g.*, *Domeisen et al.*, 2019b; *Tripathi et al.*, 2015; *Sigmond et al.*, 2013; *Baldwin et al.*, 2003).

Although the extratropical stratosphere itself has no interannual memory, essentially due to the opacity to wave propagation of the summertime easterly circulation which resets the state of the vortex every year, it does vary on interannual time scales because of dynamical linkages with other modes of atmospheric variability. A clear example of such influence is the connection between ENSO and the fluxes of planetary-scale waves from the troposphere to the stratosphere which can modulate the frequency of SSW events (*Domeisen et al.*, 2019a; *Weinberger et al.*, 2019; *Song and Son*, 2018; *Calvo et al.*, 2017; *Cagnazzo and Manzini*, 2009). In fact, the stratosphere can play a significant role in setting the extratropical response to ENSO events (*Polvani et al.*, 2017; *Butler et al.*, 2015; *Iza and Calvo*, 2015). Unlike the extratropics, the equatorial stratosphere does possess an intrinsic interannual memory which manifests itself as the Quasi-Biennial Oscillation (QBO). The QBO is characterized by an oscillation between westerly and easterly winds which occurs approximately every 28 months (*Baldwin et al.*, 2001). The QBO can induce interannual variability in the extratropical stratosphere through the modulation of upward fluxes of planetary-scale wave activity in the extratropics (*Holton and Tan*, 1980) and influence atmospheric circulation at the surface (*Gray et al.*, 2018).

On interdecadal to longer time scales, the stratospheric state is influenced by modes of sea surface temperature variability such as the Atlantic Multidecadal Oscillation (*Omrani et al.*, 2014) and the Pacific Decadal Oscillation (*Woo et al.*, 2015) and anthropogenic forcing. Perhaps the clearest example of human influence to date is the destruction of ozone which cools the polar stratosphere. This perturbation of the stratosphere has in turn affected the tropospheric circulation by inducing a poleward shift of the storm track and mid-latitude westerly jet through changes in wave forcing and wave mean-flow interactions (*Son et al.*, 2018; *Orr et al.*, 2012). Finally, increasing greenhouse gas concentrations continue to cool the stratosphere (*e.g.*, *Steiner et al.;* 2020, *Ramaswamy et al.*, 2001), and may ultimately have the largest impact as the ozone hole recovers over the next decades.

A substantial fraction of the progress made in understanding these features of the stratospheric circulation and its coupling to the troposphere is owed to the development of reanalysis data sets which have greatly facilitated the study of the dynamical phenomena that regulate the coupling. Reanalysis systems integrate both forecasts from numerical models and observations through data assimilation to produce a best guess of the true state of the atmosphere. However, as discussed in *Chapter 2* (see also *Fujiwara et al.*, 2017), reanalysis data sets differ by the models, observations and assimilation techniques they utilize. As such, they produce different versions of the thermodynamic and kinematic properties of the atmosphere.

As a notable example of the differences in the representation of the stratosphere among reanalyses, *Charlton and Polvani* (2007), and more recently *Butler et al.* (2017), have highlighted discrepancies in the onset dates of SSW events between NCEP R1 and ERA-40 data sets. However, subsequent studies have revealed that the depiction of the evolution of SSW events is fairly similar among data sets (*Martineau et al.*, 2018b; *Butler et al.*, 2015; *Palmeiro et al.*, 2015; *Martineau and Son*, 2010).

The biases among reanalyses are limited enough as to not significantly alter our understanding of the physical processes regulating the evolution of SSWs. More generally, vortex variability was also shown to be similar among reanalyses during both strong and weak stratospheric vortex states (*Martineau et al.*, 2016). On interannual time scales, *Mitchell et al.* (2015) recently compared reanalysis datasets and found a remarkable consistency between them in the context of the variability of the circulation associated with volcanic eruptions, ENSO, QBO and the solar cycle. Despite these recent findings, there is a growing need to better quantify and understand the differences in the representation of atmospheric processes among reanalyses as the number of available data sets grows with the development of more sophisticated reanalyses incorporating advanced modeling and assimilation components.

6.3 Reanalysis datasets

The reanalyses assessed in this chapter are listed in **Table 6.1**. The reader is referred to *Fujiwara et al.* (2017) and *Chapter 2* for an exhaustive description of reanalyses. Variables analyzed include geopotential height, temperature and three-dimensional wind components, all of which are analyzed on pressure levels, as well as the mean sea level pressure.

In order to facilitate the comparison of zonal-mean quantities, a standardized data set of zonal mean dynamical and thermodynamical variables, the S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels (*Martineau et al.*, 2018c; *Martineau*, 2017), was prepared for this chapter and made public at **http://dx.doi.org/10.5285/b241a7f536a244749662360bd7839312**. Details about the variables archived, the grids and numerical methods are provided in *Martineau et al.* (2018c). Analyses of the zonal mean circulation in this chapter made use of this data, with the exception of *Section 6.6. Sections 6.4.2, 6.7.1* and *6.7.2*, involved additional analysis of the full three-dimensional circulation.

Table 6.1: *List of reanalysis data sets compared.*

Name	type	Reference
ERA-40	full-input	*Uppala et al.* (2005)
ERA-Interim	full-input	*Dee et al.* (2011)
ERA-20C	surface-input	*Poli et al.* (2016)
JRA-25	full-input	*Onogi et al.* (2007)
JRA-55	full-input	*Kobayashi et al.* (2015)
JRA-55C	conventional-input	*Kobayashi et al.* (2014)
JRA-55AMIP	SSTs only	*Kobayashi et al.* (2014)
MERRA [b]	full-input	*Rienecker et al.* (2011)
MERRA-2 [b]	full-input	*Gelaro et al.* (2017)
NCEP-R1	full-input	*Kalnay et al.* (1996)
NCEP-R2	full-input	*Kanamitsu et al.* (2002)
CFSR	full-input	*Saha et al.* (2010)
CFSv2	full-input	*Saha et al.* (2014)
20CR v2	surface-input	*Compo et al.* (2011)
20CR v2c	surface-input	*Compo et al.* (2011)

[b] For MERRA and MERRA-2, only the assimilated state (ASM) products are used (see discussion in *Chapter 2* and *Fujiwara et al.*, 2017).

6.4 Sudden stratospheric warming events

A Sudden Stratospheric Warming is a dramatic breakdown of the climatological stratospheric polar vortex in the winter hemisphere, first observed in post-war Berlin by *Scherhag* (1952). The name itself encapsulates the essential features of these events. They are sudden, or, in the original language of Scherhag, explosive: the entire vortex breaks down in a few days, being associated with a remarkable warming of the winter pole, typically on the order of 10s of degrees Celsius at 10 hPa, sometimes exceeding 50 °C or 60 °C. They are primarily a Northern Hemisphere phenomenon, and only one major SSW (in 2002) has been observed in the Southern Hemisphere. We therefore focus exclusively on SSWs in the Northern Hemisphere.

SSWs tend to come in two flavors, splits and displacements. In the former, the climatological vortex splits into two vortices of similar size at the time of the warming, while in the latter, the vortex shifts off the pole. In both cases, the vortex(ices) are ultimately sheared apart, leading to an irreversible mixing of potential vorticity and the deceleration of the polar vortex. Equivalently, splits are associated with comparatively more wavenumber-2 activity, while displacements are primarily associated with wavenumber-1. Recent work has suggested that the type of warming may have significant implications to the mechanism of the warming and its impact on the surface (*Mitchell et al.*, 2013; *Esler and Matthewman*, 2011; *Matthewman and Esler*, 2011) although this sensitivity is not observed in all studies (*White et al.*, 2019; *Maycock and Hitchcock*, 2015). While many SSWs can be characterized unambiguously (*e.g.*, 22 February 1979 is a classic split), a nontrivial number (roughly 1/3, as we will see) are not so easy to classify. There are also substantial sampling uncertainty issues, particularly when assessing the influence of anthropogenic forcing (*e.g.*, *Maycock and Hitchcock*, 2015).

A key result of this section is shown in **Table 6.2** and **Figure 6.2**, where a standardized list of SSW event dates and classifications for the period 1957 to 2011 are provided. We refer the reader to the SSW compendium (*Butler et al.*, 2017) for an up-to-date list of SSW events (**https://www.esrl.noaa.gov/csd/groups/csd8/sswcompendium/majorevents.html**). Results of SSW classification performed independently for each reanalysis are listed in the *Appendix*, **Tables A6.4 - A6.7.**

6.4.1 Identifying SSW events

A number of definitions have been proposed to characterize SSWs in reanalyses and models, all ultimately establishing a key threshold to define the onset of an event. This threshold nature of SSWs makes them sensitive to subtle differences between the reanalyses (*Butler et al.*, 2015). For example, the most commonly used criteria, as adopted by the World Meteorological Organization (*WMO; McInturff*, 1978), requires that the zonal mean zonal wind reverses at 60° and 10 hPa. If the zonal mean winds drop just below zero in one reanalysis, but to only + 0.1 m s^{-1} in another, only one would count as an event (*Kim et al.*, 2017). Given the large variation between SSW events, this can alias sampling error into a comparison of events across reanalyses. In our hypothetical case above[1], a trivial difference in the reanalysis winds (0.1 m s^{-1} compared to a climatological variability on the order of 10 m s^{-1}) could mistakenly imply a large difference between two products that are actually very similar.

To account for this issue, we identify a standardized set of SSW dates for use across all reanalyses. This was obtained by first identifying events for each reanalysis individually, similarly to *Butler et al.* (2017), based on a reversal of the daily mean, zonal mean zonal wind at 60° N and 10 hPa from November to April, as listed in **Table A6.4**. The central date is defined by the day the daily mean wind first reverses, not necessarily the date on which the instantaneous zonal mean wind first reverses. Two criteria to ensure events are independent, and not the final reversal of the polar vortex to its summertime state, are also imposed. Following *Charlton and Polvani* (2007), the winds must return to a westerly direction for at least 20 consecutive days between independent events, and for at least 10 consecutive days prior to April 30.

The standard WMO definition also requires a reversal of the temperature gradient at 10 hPa. This gradient reversal is not well defined. Commonly it is interpreted that the zonal-mean temperature at the pole (here, 87.5° N is used to avoid the singular nature of the zonal mean at 90° N) must exceed the zonal-mean temperature at 60° N, but this puts a great deal of weight on the temperature near the pole. In practice, this criterion rarely matters; the stratosphere remains in geostrophic balance during an SSW, such that a reversal of the 10 hPa winds implies a reversal in the temperature gradients below 10 hPa, which are highly correlated with the 10 hPa temperatures. Only a few events would be excluded (two from NCEP R1, and just one from JRA-55, as delineated by the green boxes in **Table A6.4**). We therefore omit the temperature gradient criterion for classifying SSW events in this work.

To establish the standard set of dates listed in **Table 6.2**, events were defined when a majority of the reanalyses identify a SSW around the same time, *i.e.*, prior to 1979, 2 out of 3 reanalyses must detect the event and post 1979, at least 4 reanalyses must detect the event. The onset date was then set by taking the median across the dates given by each reanalysis. In recent decades, the dates rarely vary by more than a day or two across reanalyses, but there are a few events at the beginning of the reanalysis record, as in December 1965, where the spread was more than a week. In this case the date was set by the average of the two more modern reanalyses.

The frequency and seasonality of SSWs determined from each reanalysis separately were examined, as detailed in *Ayarzagüena et al.* (2019). In both periods, historical (1958 - 1978) and satellite (1979 - 2012), there is a good agreement in the mean frequency of SSWs between all reanalyses. This frequency is very similar in both eras, with 5.9 events per decade for the historical period and 6.5 events per decade for the satellite period.

In contrast, larger differences are found for the seasonality of SSWs. **Figure 6.1** shows the SSW decadal frequency distribution within ± 10-day periods. The historical period shows the largest spread. ERA-40 and JRA-55 display an increasing SSW occurrence from early winter that maximizes in January and decreases by late winter (**Figure 6.1a**). On the contrary, the intraseasonal distribution of SSWs for NCEP R1 shows three sharp maxima in early, mid and late winter, in agreement with the evolution of the standard deviation of the polar night jet (PNJ) for this reanalysis.

In the satellite period, the results are similar across reanalyses (**Figure 6.1b**). For this time period, the maximum occurrence is shifted to late winter in all datasets, unlike the distributions of ERA-40 and JRA-55 in the historical period. Similar differences in the intra-seasonal distribution of events were already documented by *Gómez-Escolar et al.* (2012) for the pre/post 1979 periods. The distribution of events in the two periods were compared with a two-sample Kolmogorov-Smirnov test: the null hypothesis that both samples came from the same probability distribution can be rejected. This may indicate low frequency variations in the seasonality of SSWs, although we have less confidence in the pre-satellite distribution given differences between reanalyses.

6.4.2 Characterizing SSW events

Recent work has suggested that there may be fundamental differences between the two types of sudden warmings. For instance, *Matthewman et al.* (2009) have shown that while split events typically have deep equivalent-barotropic structures, vortex displacement events have clear baroclinic structures. In addition, the impact of the event on the troposphere may differ between the two types of events (*e.g.*, *Mitchell et al.*, 2013).

[1] This case is actually not hypothetical; a similar situation, for example, occurred in February 2002, when MERRA missed an event detected by MERRA-2 by only 0.07 m s^{-1}.

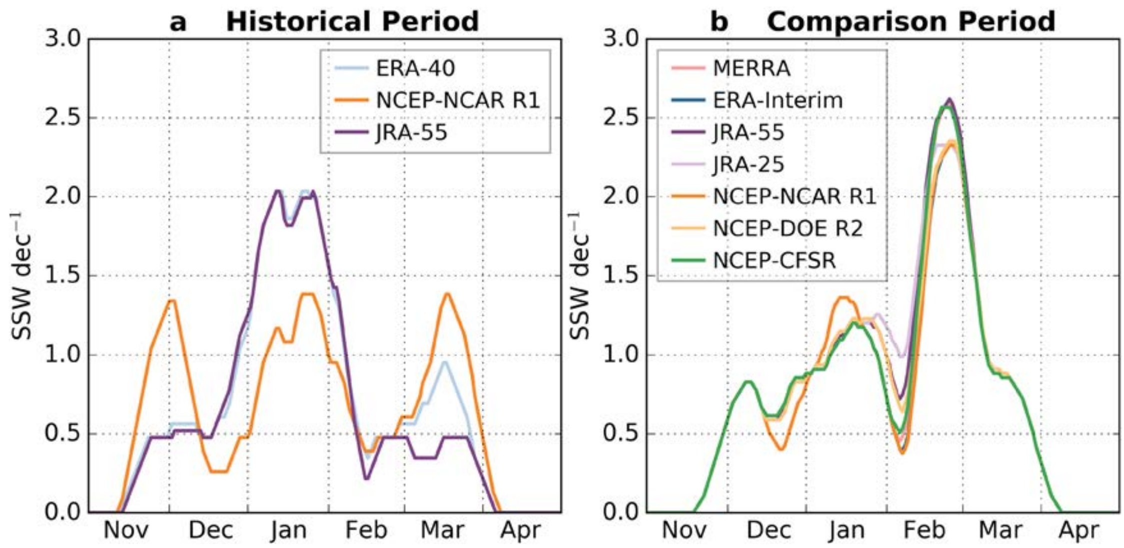


Figure 6.1: *Decadal frequency distribution of SSW events within ± 10 day-periods from the date displayed in the axis for: (a) the historical period (1958 - 1978) and (b) the satellite period (1979 - 2012). Data was smoothed with a 10-day running mean. Reproduced from Ayarzagüena et al. (2019).*

The limited sample size, however, leads to large uncertainty, such that alternative studies come to differing conclusions (*e.g.*, *Maycock and Hitchcock*, 2015). The topic is further muddied by the fact that different studies have utilized alternative definitions of SSWs, leading to a proverbial apples vs. oranges situation.

To provide greater clarity, while acknowledging that the topic is still an area of active research, we have taken the following approach. First, we consider only wind reversal events: classification schemes were applied to the 36 SSWs identified in the previous subsection. Second, we have applied four alternative classification schemes, described in more detail below, chosen to capture the range of ideas in the current literature. We provide a standardized classification of each event, listed in **Table 6.2**, based on the agreement of the classification applied separately to each reanalysis.

We compare three schemes designed to characterize whether the polar vortex is split (S) or displaced (D) during the warming event with another classification scheme that focuses primarily on the wave activity that precedes the vortex breakdown. The three schemes have been tuned to produce approximately the same rates of S and D events, all three reporting slightly more displacements than splits. The wave based diagnostic is different in that it focuses on the period leading up to the warming, as opposed to the evolution of the warming itself. It reflects the climatological dominance of wavenumber 1, classifying a clear majority of the events as wave 1-type. All the schemes are detailed below.

(1) The Seviour et al. (2013) classification scheme is based on geometric moment diagnostics of the geopotential height field at 10 hPa. The use of 10 hPa geopotential heights, which is output from all reanalyses, makes the scheme more practical than previous moment diagnostic techniques which rely on isentropic tracers, such as N_2O or potential vorticity (*Mitchell et al.*, 2011, 2013; *Waugh and Randel*, 1999; *Waugh*, 1997). The *Seviour et al.* (2013) approach was originally designed to characterize event dates as well; *e.g.*, a split event was triggered when the aspect ratio of the vortex remained higher than 2.4 for 7 days or more. However, only half of the major splits/displacements using this method are in common with those detected using the zonal-mean zonal wind reversal.

We therefore adapted the method to classify reversal events. We apply the same methodology as in *Seviour et al.* (2013), but only to days - 10 to + 10 surrounding the wind reversal. The diagnostic is based on both the aspect ratio of the vortex (the number of days the aspect ratio is above 2.4) and the displacement of the centroid (the number of days the centroid of the vortex stays below 66 degrees latitude). If the latter is greater than the former, then the event is classified as a displacement. Conversely, if the former is greater than the latter, the event is classified as a split. If the numbers are equal (or both are zero) we consider the event "unclassifiable". Note that if this adapted technique is applied to the events of *Seviour et al.* (2013), it yields identical classifications (W. Seviour, personal communication). **Table A6.5** shows results based on analysis of each individual reanalysis.

(2) The "Shibata" scheme was originally developed by Kiyotaka Shibata, and first described in *Ayarzagüena et al.* (2019). It focuses on non-zonal anomalies in the absolute vorticity at 10 hPa over a 16 day period starting 5 days before the central date of the SSW and ending 10 days later. Application of this scheme to each reanalysis is listed in **Table A6.6**.

The method is based on the algorithm suggested by *Charlton and Polvani* (2007), but with a few important modifications, as detailed in *Ayarzagüena et al.* (2019). Briefly, the algorithm identifies a local maximum in the vorticity.

Table 6.2: *Sudden Stratospheric Warming dates and classifications, according to the four schemes: D refers to a displacement, S to a split, and U, for an event that was unclassifiable, while W1 and W2 refer to events preceded by wave forcing at that number. The symbol * indicates that there was disagreement between the reanalyses; a † indicates that only a single reanalysis disagreed (after 1979 only).*

date	Seviour	Shibata	Lehtonen	Barriopedro and Calvo
30-Jan-58	D	S	S	W1
17-Jan-60	S*	D*	D*	W1
29-Jan-63	S	S	D*	W2
17-Dec-65	D	D*	D	W1
23-Feb-66	D*	D	S	W1
7-Jan-68	S	S	S	W2
28-Nov-68	D	D	D	W1
13-Mar-69	S*	D	D	W1
2-Jan-70	S	D	D*	W1
18-Jan-71	S	S	S	W2
20-Mar-71	D	D	D	W1
31-Jan-73	S	S	S	W1
9-Jan-77	S	D	S	W1
22-Feb-79	S	S	S	W2
29-Feb-80	D	D	D	W1
4-Mar-81	D	D	D	W1
4-Dec-81	U	D†	D	W1
24-Feb-84	D	D	D	W1
1-Jan-85	S	S*	S	W2
23-Jan-87	D	D	D	W1
8-Dec-87	S	S*	S	W1
14-Mar-88	S	S	S	W1
21-Feb-89	S†	S	S	W2
15-Dec-98	D*	S	D	W1
26-Feb-99	S	S	S	W1
20-Mar-00	U*	D†	D	W2
11-Feb-01	S	D†	S*	W1
31-Dec-01	S	D	D	W1
18-Jan-03	S	S†	S	W1
5-Jan-04	D	D	D	W1
21-Jan-06	D	D	D	W1
24-Feb-07	D	D	D	W1
22-Feb-08	D	D	D	W2
24-Jan-09	S	S	S	W2
9-Feb-10	U*	S	S	W1
24-Mar-10	D	D	D	W1

If two vorticity maxima are detected in diametrically opposing sectors, and the secondary maximum is at least half as strong as the first, the event is classified as a split. Otherwise it is a displacement. The main differences with the strategy of *Charlton and Polvani* (2007) consist in the definition of the sector around the strongest vorticity maximum, and the fact that the second sector must be located diametrically opposed to the first one.

(3) The Lehtonen and Karpechko (2016) classification, applied to all reanalyses in **Table A6.7**, shares features with both of the previous methods. It is based on a analysis of geopotential height at 10 hPa (as with the Seviour method), but with a goal similar to that of the Shibata approach: to separate cases where there are two independent vortices (as in a split event) from cases where there is essentially one vortex at any given time (as in a displacement).

The algorithm seeks out the two minima in the 10 hPa geopotential height, spaced apart by at least 1500 km in the horizontal and separated by a ridge of at least 375 m. If this condition is met on at least three consecutive days over the period from 5 days prior to the event onset to 10 days after, then the SSW is classified as a split. Otherwise, it is classified as a displacement. These parameters in the classification were selected to give the best agreement with the classification of major SSWs during 1958 - 2002 presented by *Charlton and Polvani* (2007).

Finally, the **(4) Barriopedro and Calvo (2014) method** classifies SSWs into wave 1 (W1) and wave 2 (W2) types by focusing on wave activity over just the period leading up to the SSW. The method is based on earlier work by *Bancalá et al.* (2012). It was applied to all reanalyses (with the exception of MERRA-2), and there was universal agreement on the classification of all 36 warmings across all the datasets. Briefly, this approach considers a Fourier decomposition of geopotential height anomalies at 50 hPa and 60 °N over an 11-day period; days -10 to 0 relative to onset. An SSW is classified as a W2 event if the amplitude over the 11-day period associated with wave 2 is equal or larger than that of wave 1, or if the wave 2 amplitude mean exceeds that of wave 1 by 200 m or more for at least one day of the period. Otherwise, the SSW is classified as a W1 event. In most cases, the former condition determines the type of SSW. The latter was included because the build up of W2 events is generally more abrupt than W1 events. The 50 hPa pressure level was chosen because wave 2 reaches its climatological maximum at this level.

As shown in **Figure 6.2**, 11 of the 36 SSWs observed between 1958 and 2011 are unanimously classified as splits by all three schemes, and 12 unanimously as displacements. The remaining 13 events differ depending on the classification scheme. These events, however, are more likely to be classified as a displacement: 8 events were displacements according to 2 of the 3 schemes, while only 5 were splits according to 2 of the 3 schemes.

We find that more than half of split events are preceded by enhanced wavenumber-2 activity (see the *Barriopedro and Calvo* (2014) method described above), as one might expect but the rest do not have prominent wavenumber-2 precursors. These may correspond to events that are preconditioned by wavenumber-1 forcing (*Bancalá et al.*, 2012; *Labitzke*, 1977) which reduces the necessity for large wavenumber-2 forcing prior to the onset in comparison to "pure" wavenumber-2 events. Perhaps more surprisingly, 2 displacement events (20 March 2000 and 22 February 2008) – one that was unambiguous across all classification schemes (the latter) – were also preceded by enhanced wavenumber-2 forcing.

Compared to the timing of event dates, there is more spread in the classification analysis between different reanalysis products. In a few instances, a tie had to be broken, in which case we gave greater weight to more modern reanalyses.

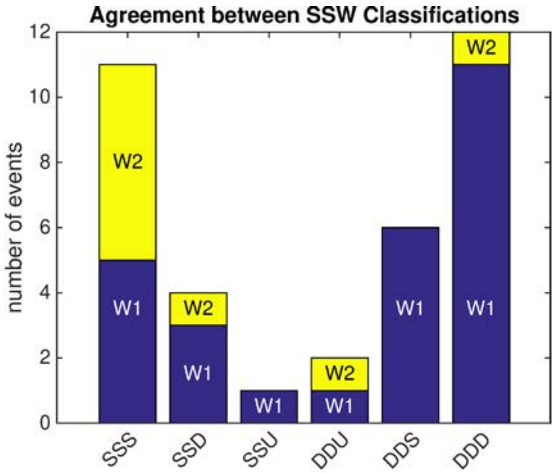


Figure 6.2: *Agreement between the SSW classification approaches. SSS and DDD refer to cases where all three schemes identified a split, displacement, respectively. SSD refers to cases where two schemes indicate the event is a split while one characterizes it as a displacement, SSU refers to a similar case, but where the third scheme was unable to classify the event, and so forth. Each bar is then divided into cases where the wave amplitude at 50 hPa over the 11 days preceding the event (see the Barriopedro and Calvo (2014) method for more detail) was primarily wave 1 (blue) or 2 (yellow).*

In addition, the Seviour scheme considers a few (3 of 36) events to be "unclassifiable", as they reflect too much of a mixture of properties of splits and displacements. In some cases, an event was unclassifiable for the individual reanalyses; in others, there was so much spread between products that we felt "unclassifiable" was the most reasonable designation. The classification schemes were applied to the wind reversal SSW events as above. There are a number of small differences in the dates and classifications based on individual reanalyses, as detailed in **Tables A6.4**, A**6.5**, A**6.6** and A**6.7**. Hence two studies based exclusively on two different reanalyses will not find the same SSW frequency, or produce the same composite fields. We find, however, that these differences are generally not significant if one accounts for sampling error. That is to say, the differences in the SSW frequency, or event composites, based on the two different reanalyses, would not be statistically significant.

As an example, consider a comparison of the dynamical evolution of W1 and W2 SSWs, classified with the *Barriopedro and Calvo* (2014) method, across different reanalyses. Key characteristics of SSWs, such as the warming of the lower and middle polar stratosphere, the deceleration of the polar vortex, and the injection of tropospheric wave activity, were compared across reanalyses by *Ayarzagüena et al.* (2019) based on the diagnostic benchmarks by *Charlton and Polvani* (2007).

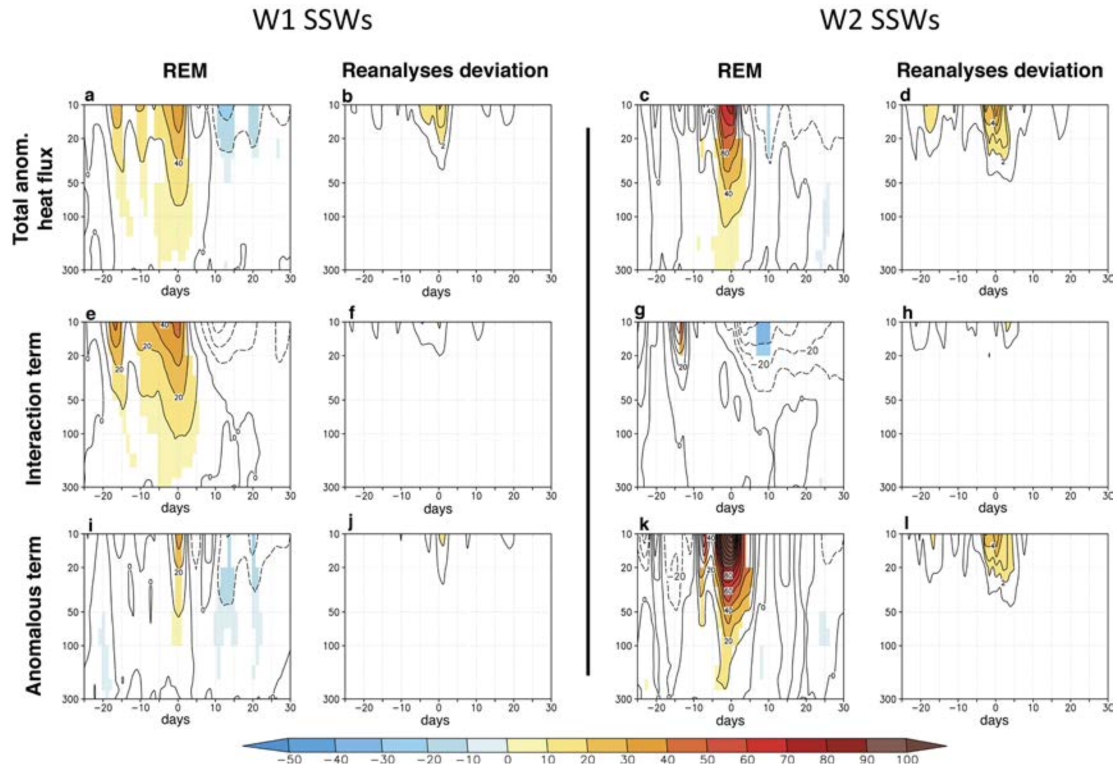


Figure 6.3: *(a) REM composited time evolution of the total anomalous eddy heat flux averaged over 45° N-75° N (K m s*$^{-1}$*) at different levels from 29 days before to 30 days after the occurrence of W1 SSWs in the comparison period. Contour interval: 20 K m s*$^{-1}$*. (b) Same as (a) but for the standard deviation of the reanalyses with respect to the REM. Contour interval: 2 K m s*$^{-1}$*. (e) and (f) Same as (a) and (b) but for the interaction between climatological and anomalous waves. Contour interval: 10 K m s*$^{-1}$*. (i) and (j) Same as (a) and (b) but for the contribution of the intrinsic wave activity associated with wave anomalies to the total anomalous heat flux. (c), (d), (g), (h), (k) and (l) Same as (a), (b), (e), (f), (i) and (j) respectively but for W2 SSWs. Shading in (a), (e), (i), (c), (g) and (k) denotes statistically significantly anomalies at a 95 % confidence level (Monte-Carlo test). Adapted from Ayarzagüena et al. (2019).*

Common events were considered to avoid possible discrepancies between reanalyses due to a different sampling. In both the pre- and post satellite periods of comparison, the agreement between datasets is very high. Only small discrepancies are found for the deceleration of the polar vortex at 10 hPa in the case of NCEP R1, particularly in the historical period. These discrepancies are probably related to the lowest model top and vertical resolution of the NCEP R1 model, since other SSW properties computed at lower levels do not present discrepancies between reanalyses.

As shown in **Figure 6.3**, anomalous meridional eddy heat flux (HF), averaged between 45 °N and 75 °N, and its different contributing terms (*Nishii et al.*, 2009) have been computed as a function of height about the onset date of SSWs. Since some previous studies have shown differences in mechanisms triggering different types of SSWs (*e.g.*, *Barriopedro and Calvo*, 2014; *Smith and Kushner*, 2012), the heat flux analysis is shown separately for W1 and W2 SSWs in the comparison period. The results of the Reanalysis Ensemble Mean (REM) resemble very much those by *Smith and Kushner* (2012) for D and S events, respectively, despite the lack of a one-to-one correspondence between W1 (W2) and D (S) SSWs.

W1 events are mainly triggered by the interaction between climatological and anomalous waves (**Figures 6.3a**, **e** and **i**) during persistent and moderately intense peaks of HF anomalies. Conversely, W2 events are related to intense but short pulses of HF arising from anomalous wave packets (**Figures 6.3c**, **g** and **k**). The comparison among reanalyses results reveals that all datasets can reproduce the different mechanisms involved in W2 and W1 SSWs. The spread is higher for W2 SSWs than for W1 SSWs particularly during the days immediately before the occurrence of SSWs (**Figures 6.3b**, **d**, **f**, **h**, **j**, and **l**). This is probably due to the smaller sample of W2 SSWs.

The tropospheric circulation associated with the occurrence of W1 and W2 SSWs in the satellite period has also been explored (**Figure 6.4**). The tropospheric patterns preceding the SSWs have been computed by analyzing the averaged geopotential height anomalies at 500 hPa in the [-10,0]-days prior to the central date of each type of SSW, while the surface signal after the occurrence of W1 and W2 SSWs has been analyzed by compositing the mean sea-level pressure (MSLP) anomalies in the [5,35]-days after these dates. The precursor signals for W1 SSWs and W2 SSWs show predominant W1-like and W2-like structures, respectively, that are similar to the precursors of the most intense events of stratospheric vortex deceleration (*Martineau and Son*, 2015).

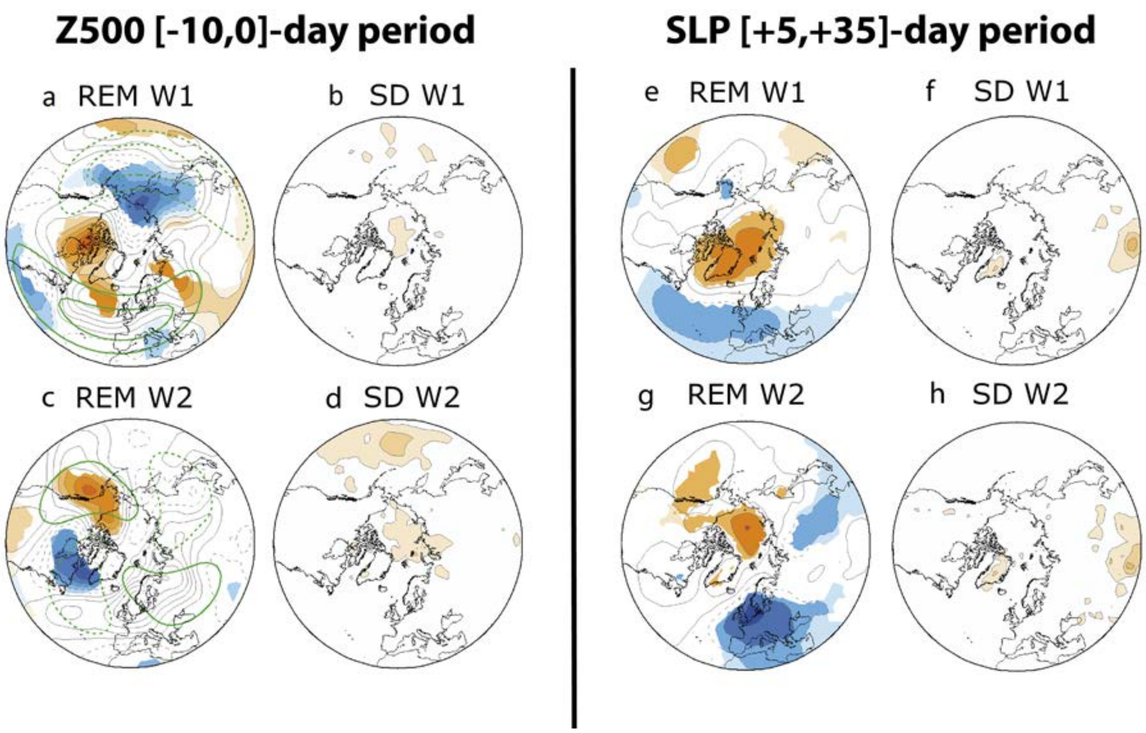


Figure 6.4: *(a) Reanalysis ensemble mean (REM) of W1 SSW-based composites of 500 hPa geopotential height anomalies (contour interval 20 m) over the [-10, 0]-day period before events for the comparison period (1979 - 2012). Only statistically significant anomalies at the 95 % confidence level of the same sign (Monte Carlo test) in at least 66.7 % of all reanalyses are shaded. (b) Standard deviation of the reanalyses with respect to the REM divided by the square root of the number of reanalyses for W1 SSWs (contour interval is 1 gpm). (c, d) Same as (a) and (b) but for WN2 SSWs, respectively. Green contours in (a) and (c) show the REM climatological W1 and W2 of 500 hPa geopotential height from November to March, respectively (contours: 40 and 80 gpm). To the right, the MSLP is composited over the [5, 35] day period after SSWs. The panels follow the same order as the Z500 precursors. Contour interval is 2 hPa for REM composites and differences and 0.1 hPa for the standard deviation of the reanalyses. Adapted from Ayarzagüena et al. (2019).*

We refer the reader to *Cohen and Jones* (2011) for earlier precursors. The SSW impact shows a negative Northern Annular Mode (NAM) pattern with positive MSLP anomalies over the polar cap in both cases, but some differences are found in lower latitudes of the Northeastern Pacific and Atlantic basins. The Pacific responses resemble the tropospheric precursor patterns therein, suggesting a possible remainder signal. In both cases (precursors and responses), the agreement among reanalyses is very good and almost no differences have been detected.

This analysis shows overall very good agreement among reanalyses in the representation of the main features, triggering mechanisms and surface fingerprint of SSWs. Despite this, some differences are found among reanalyses, particularly in the historical period and concerning the NCEP R1 reanalysis. Before 1979, SSWs in NCEP R1 show a lower mean frequency and a different seasonal distribution with respect to JRA-55 and ERA-40 (**Figure 6.1**). This disagreement also extends to climatological fields and their variability in upper levels. A plausible cause of this discrepancy is the strong artificial temperature trend affecting the early record of NCEP R1 (*Badin and Domeisen*, 2014). Arguably, the characteristics of the reanalysis models play an important role in this period, since the number of available data to be assimilated at upper levels is limited. Thus, we do not recommend the use of this reanalysis in the historical period for model evaluation initiatives.

6.4.3 Sampling uncertainty vs. reanalysis uncertainty

Studies of stratosphere-troposphere coupling are limited by the considerable dynamical variability present in both the stratosphere and the troposphere below. This variability introduces considerable sampling uncertainty into composite analyses, for example, and it is thus of interest to use all the data that is available. The amount of observational data increased considerably after 1979 when global satellite observations became broadly available. However, the basic theory underlying the occurrence of SSWs was formulated by *Matsuno* (1971), and several well-known reviews of the dynamics of these events were published well before a significant time series of satellite observations was available (*McIntyre*, 1982; *Labitzke*, 1977), indicating that the observational record largely based on radiosondes is of considerable value. This can be expected to be even more the case within the troposphere which is more easily observed with radiosondes.

Indeed, the uncertainty arising from dynamical variability that is intrinsic to the global circulation is far larger than the uncertainty arising from observational uncertainty and the process of assimilating this data into reanalysis products (*Hitchcock*, 2019). This is demonstrated in **Figure 6.5**. **Figure 6.5a** shows the time-series of zonal mean zonal wind at 10 hPa, 60 °N, around 36 major sudden stratospheric warmings from a single reanalysis, JRA-55. Events post 1979 are in solid lines, while those prior to 1979 are in dashed lines. The broad spread across events at all lags from the central date is evident, and the character of the variability in the two periods is not obviously different. This inter-event variability can be compared with the differences for individual events across reanalysis products.

Figure 6.5b shows the corresponding time series for one event (21 February 1989) during the post-1979 period, for each of the 12 reanalyses. With the exception of the two reanalyses that ingest only surface observations (ERA-20C, 20CR v2), the time-series are nearly indistinguishable relative to the inter-event variability highlighted in **Figure 6.5a**. This is even more the case if one omits NCEP-NCAR R1 and NCEP R2 whose forecast model top lies at 10 hPa. Although relatively few reanalyses extend prior to 1979 (and only one of the more modern products), this close agreement holds nearly as well for the pre-1979 period (**Figure 6.5c**).

Including the 21 years from 1958 to 1979, in addition to the 32 years from 1979 through 2010, can be expected to shrink confidence intervals by a factor of $\sqrt{32/53} = 0.78$; about a 20 % reduction. For instance, **Figure 6.6** shows the impact of including this period on the estimated frequency of SSWs. Here the *Lehtonen and Karpechko* (2016) classification method is used to define SSWs. Although not shown, similar results are found for the

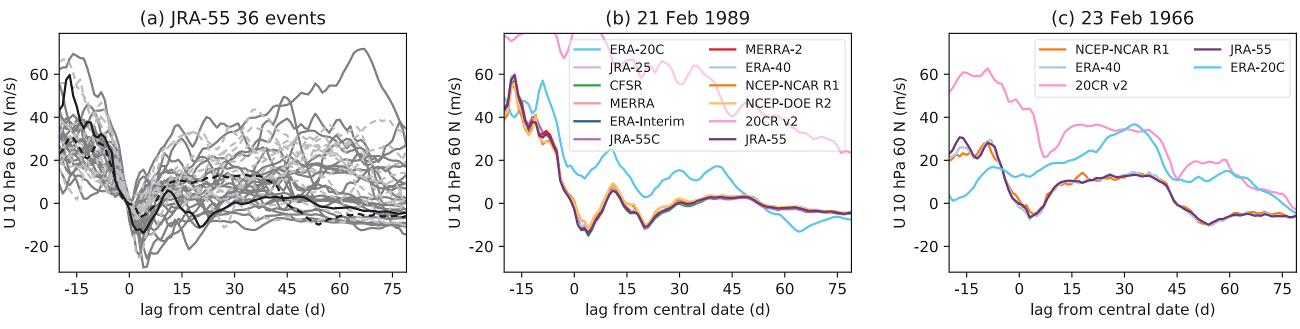


Figure 6.5: *(a) Zonal winds at 10 hPa and 60° N from JRA-55 for 36 sudden warmings. Events from the satellite period are in dark grey, those from the radiosonde period are in light grey and are dashed. (b) Winds for a single satellite-period event for all reanalyses; this event is shown by the black line in (a). (c) Zonal winds at 10 hPa and 60° N for a single radiosonde-period event for all reanalyses covering this period; this event is shown by the dashed black line in (a). Reproduced from Hitchcock (2019).*

other classification methods. The confidence intervals are generated by a bootstrapping procedure. For instance, for the post- 1979 period, sets of 32 years chosen at random (with replacement) from the period from 1958 through 2010; events that happen during these years are then used to generate an overall frequency. If a year is chosen multiple times, the events that occurred during these years are also included multiple times. This is carried out 10000 times; the 2.5th and 97.5th percentiles then define the confidence interval. A similar procedure is used for the confidence intervals on the whole 1958 - 2010 period but using sets of 53 years.

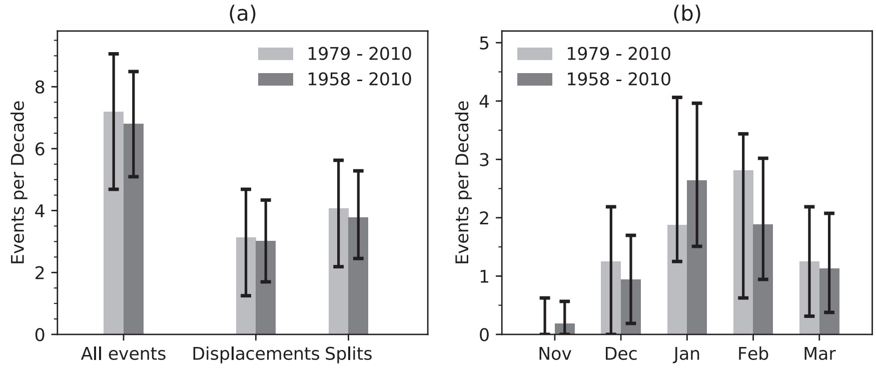


Figure 6.6: *(a) Frequency of all SSW events, and of events classified as splits or displacements for the satellite period versus the entire period where quality reanalyses are available. (b) Same as (a) but for each month of extended winter. Error bars indicate 95% confidence intervals, see text for details. Reproduced from Hitchcock (2019).*

The resulting confidence intervals are indeed reduced by a factor close to the 20 % estimate given above. The overall event frequency and the frequency of splits and displacements are somewhat reduced. The seasonal distribution of events is more substantially affected; within the broader record more events occur in January that in any other month; the period from 1979 - 2010 had relatively few January events and relatively many February events resulting in a rather different seasonal distribution as shown in **Figure 6.1** (though one well within sampling uncertainty).

Similar reductions in confidence intervals can be found for more dynamical quantities. **Figure 6.7** shows three such examples. **Figure 6.7a** shows the anomalous zonal wind, integrated from 1000 hPa to 100 hPa, from days 5 to 60 following the central date. **Figure 6.7b** shows the anomalous meridional momentum flux, also integrated from 1000 hPa to 100 hPa and averaged from days 5 to 60 following the central date. Finally, **Figure 6.7c** shows the meridional heat flux at 100 hPa, averaged from days -15 to 0, prior to the central date (in red), and averaged from days 5 to 60 following the central date (in blue). In all cases confidence intervals are generated by a similar bootstrapping procedure; however in this case the events themselves are sub-selected, rather than the years.

Again, in all cases, including the whole period results in a slightly different meridional structure. The low-latitude easterly response is somewhat weaker in **Figure 6.7a**, the momentum flux response is somewhat more positive at all latitudes (**Figure 6.7b**), and the heat fluxes in the recovery period are somewhat more reduced (**Figure 6.7c**). More importantly, the reduction in confidence intervals provides a stronger constraint for dynamical understanding and for model evaluation.

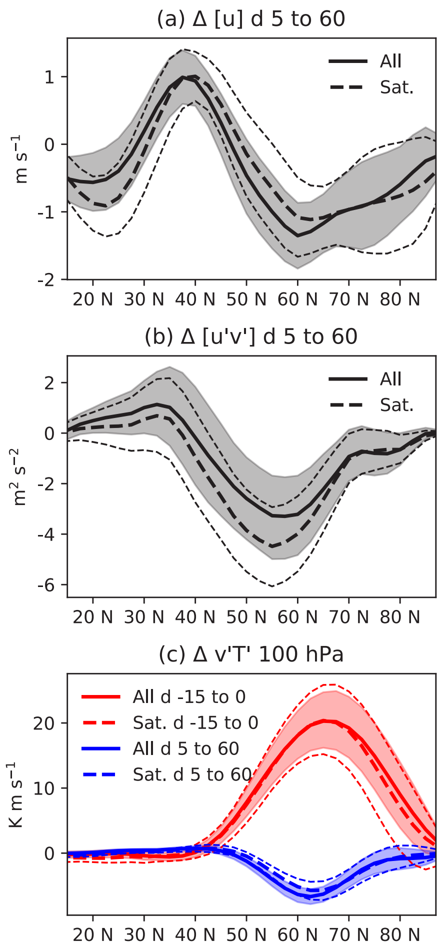


Figure 6.7: *(a) Composite mean of vertically averaged (100 to 1000 hPa) zonal wind anomalies, averaged over lags 5 to 60 days following major warmings. The solid line shows the composite for all events while the dashed line shows the composite for the satellite era alone. Confidence intervals for the whole period are shaded while those for the satellite era are indicated by thin dashed lines. (b) Similar but for vertically integrated momentum fluxes. (c) Similar but for meridional heat fluxes at 100 hPa, averaged over lags -15 to 0 (in red), and over lags 5 to 60 (in blue). Reproduced from Hitchcock (2019).*

6.4.4 Assessing the internal consistency of SSW events in Reanalyses

Given that the sampling error tends to overwhelm differences in the representation of SSWs in different reanalysis products, we consider an alternative approach to evaluating their fitness: an assessment of their internal consistency. Many studies have investigated the evolution of zonal mean zonal wind using zonal-mean momentum budgets applied to reanalysis data (*e.g.*, *Martineau and Son*, 2015; *Limpasuvan et al.*, 2004). Reanalysis data sets, however, are known to present biases with respect to observations and with respect to each other. For instance, recent studies by *Lu et al.* (2015) and *Martineau et al.* (2016) have highlighted discrepancies among data sets concerning the momentum budget. Here we summarize and show key figures from the analysis of *Martineau et al.* (2018b), which quantified uncertainties in the zonal momentum budget among the reanalysis data sets.

The comparison is performed among all conventional reanalysis data sets except for ERA-40 whose deficiencies are well documented in the literature (*e.g.*, *Martineau et al.*, 2016) and which terminates in 2002, limiting the sample of SSW events. The common dates identified in **Table 6.2**, beginning with the 29-Feb-80 event and ending with the 24-Mar-10 event, are used to perform composites of the momentum budget for SSW events. The zonal-mean momentum budget can be written as follows:

$$\frac{\partial \bar{u}}{\partial t} = \underbrace{f\bar{v}}_{fv} \underbrace{-\frac{1}{a\cos^2\phi}\frac{\partial(\cos^2\phi\overline{u'v'})}{\partial\phi}}_{du'v'/dy} \underbrace{-\bar{v}\frac{1}{a\cos\phi}\frac{\partial(\bar{u}\cos\phi)}{\partial\phi}}_{Adv_\phi} \underbrace{-\bar{\omega}\frac{\partial\bar{u}}{\partial p}}_{Adv_p} \underbrace{-\frac{\partial(\overline{u'\omega'})}{\partial p}}_{du'\omega'/dp} + R \quad (6.1),$$

where f is the Coriolis parameter, u, v, ω are the zonal, meridional, and vertical components of wind, ϕ is the latitude, p is the pressure, and a is the mean radius of the Earth (6371 km). Overbars and primes denote zonal mean and anomalies with respect to the zonal mean, respectively. While the left-hand side term expresses the zonal-mean zonal wind tendency, terms of the right-hand side represent forcing terms. They are, in order, the acceleration due to the Coriolis torque, the meridional convergence of momentum fluxes, the advection of zonal momentum by the meridional wind, the vertical advection of zonal momentum by the vertical wind, and the vertical convergence of vertical momentum fluxes. The last term, R, is referred to as the residual and represents sub-grid scale processes such as gravity wave drag and numerical diffusion. It also includes imbalances in the momentum equation introduced by the data assimilation process (analysis increment), errors due to the interpolation from model levels to pressure levels, and errors related to the numerical methods employed to evaluate each term of the equation. All calculations are based on the zonal-mean data set of global atmospheric reanalyses on pressure levels (*Martineau et al.*, 2018c; *Martineau*, 2017) which provides dynamical variables on a common 2.5° by 2.5° latitude-longitude grid for all reanalysis datasets at six-hour intervals. The diagnostics presented here are markedly more sensitive to the choice of data set than horizontal resolution (*Martineau et al.*, 2018c).

Figure 6.8 shows the composite evolution of all terms of the zonal-mean momentum equation during SSW events. In addition to the terms evaluated and shown for each individual data set, the standard deviation among an ensemble of the latest reanalysis data (CFSR, ERA-Interim, JRA-55 and MERRA-2) is displayed. SSW events are characterized by an intense deceleration (up to $-7\,\mathrm{m\,s^{-1}\,day^{-1}}$ at 3 hPa) of the zonal-mean zonal wind in the mid-stratosphere. Uncertainties in the zonal wind tendency are typically small in comparison to other terms of the momentum equation and are largest several days before the onset date (day 0). The dominant forcing terms are those that are typically included in the quasigeostrophic version of the momentum equation – *i.e.*, the acceleration due to the Coriolis torque and the convergence of meridional fluxes of momentum. These two forcings are strongly opposed, but not completely. Their sum results in a net deceleration before the onset of SSW events. Uncertainties in these forcing terms due to inter-reanalysis discrepancies typically peak several days before the onset of SSW events. Other forcing terms that are left out of the QG approximation have smaller magnitudes and show better agreement among the reanalyses. Finally, the residual is typically negative before the onset of SSW events, in part due to the exclusion of gravity wave drag from our analysis (*Martineau et al.*, 2016). It becomes more neutral after the onset, suggesting a more dynamically quiet period.

It is worth noting that preceding lag 0, JRA-25 shows a markedly larger residual in comparison to other reanalyses both in the mid and upper stratosphere. This large negative residual may be attributed to an underestimation of deceleration by the Coriolis torque in the mid stratosphere and an overly strong momentum flux convergence in the upper stratosphere in comparison to other reanalyses (not shown, see *Martineau et al.* (2016) for more details). Note that NCEP R1 and NCEP R2 are also clear outliers for these two forcings in the mid-stratosphere. Their residual is however not shown here since vertical motion is not provided in the stratosphere.

The vertical profiles of the forcing terms and their uncertainties are shown in **Figure 6.9**. Here, the inter-reanalysis standard deviation is shown separately for the ensemble of latest reanalyses and an ensemble of all reanalyses (listed in legend). Overall, all forcing terms display an exponential increase of uncertainties with height in the stratosphere. Again, the Coriolis torque and the convergence of meridional momentum fluxes dominate in terms of uncertainty. It is also noteworthy that uncertainties of the latest reanalysis ensemble are always smaller than the all reanalysis ensemble in the stratosphere which suggests an enhanced consistency in the representation of the atmospheric circulation in the modern reanalysis products.

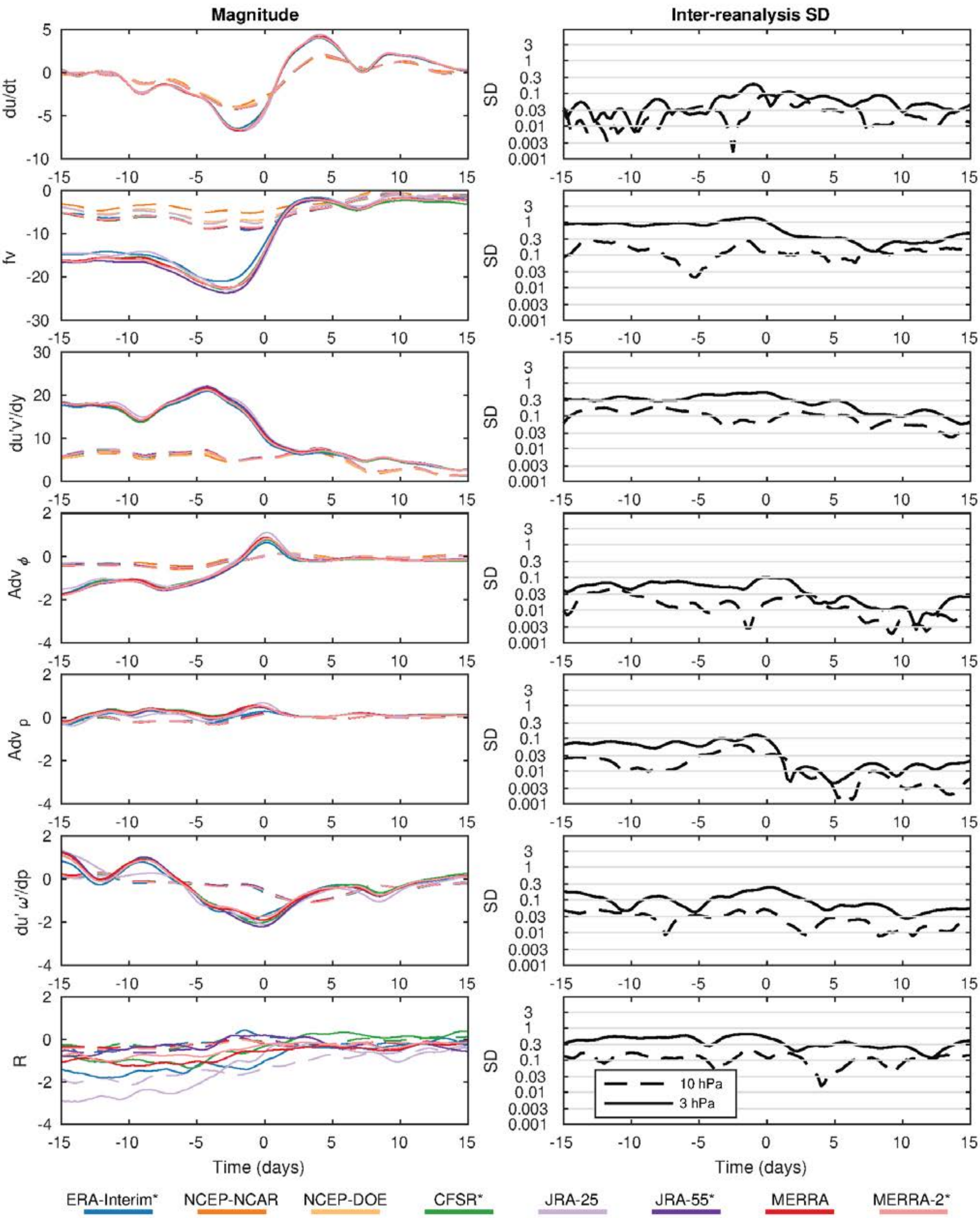


Figure 6.8: *Evolution of forcing terms of the zonal-mean momentum equation at 10 hPa (dashed lines) and 3 hPa (solid lines) in the course of SSW events. All variables are averaged from 45°N to 85°N. Note that the range of the y axis in each panel is different. (b) The inter-reanalysis spread (standard deviation) of the corresponding terms are shown for the latest reanalysis ensemble members (indicated with a * in the legend). The standard deviation is shown on a logarithmic scale: the spacing between tick marks represents a decrease or increase of the standard deviation by a factor of about 3. All quantities are expressed in m s^{-1}day^{-1}. Reproduced from Martineau et al. (2018b).*

Martineau et al. (2018a) have noted that not only the mean forcings between the Coriolis torque and momentum flux convergence are strongly opposed, but also the inter-reanalysis discrepancies in the Coriolis torque are often compensated for by inter-reanalysis discrepancies in the momentum fluxes. This results in a seemingly better self-consistency of the momentum equation (small residual) although the disagreement between data sets about the dominant momentum forcing terms can be large. This compensation could be the result of an induced meridional overturning circulation in response to biases in wave drag from planetary waves or gravity waves among the data sets. The meridional overturning circulation is an ageostrophic circulation and is thus not constrained by the thermal structure of the atmosphere like the zonal mean zonal winds which largely obey geostrophic and hydrostatic balance in the extratropics.

The aforementioned results characterized uncertainties of the momentum budget in reanalysis data sets by considering all SSW events but the study of *Martineau et al.* (2018a) provides a more thorough analysis by investigating differences between SSW events characterized by a split or displacement of the stratospheric polar vortex. The classification is done by both using vortex moment diagnostics (see *Section 6.4.2*) and by identifying the dominant fluxes of wave activity from the troposphere to the stratosphere (whether dominated by wavenumber 1 or 2) prior to the events. Overall, there is no striking difference in the uncertainties of the momentum budget between these different types of events. It is rather found that the intensity of the event, evaluated by the magnitude of the deceleration of zonal-mean zonal wind prior to the reversal, is more relevant for the agreement between reanalysis data sets. As is somewhat intuitive, the events that showed the strongest deceleration and largest forcing terms were shown to suffer from larger inter-reanalysis uncertainties.

In summary, there is generally a good agreement between the various terms of the zonal-mean momentum budget among reanalysis data sets. The discrepancies are small enough as to not introduce important uncertainties in our understanding of the dynamical evolution of SSW events. Inter-reanalysis uncertainty typically increases exponentially with height as the forcing terms also grow in magnitude. The dominant forcing terms, *i.e.*, momentum flux convergence and the Coriolis force, dominate the budget and have the largest uncertainties. The residual also increases with height, indicative of the greater role played by gravity waves in the momentum budget in the mid- to-upper stratosphere. Differences in the contribution of gravity waves to the momentum budget among reanalyses are hard to evaluate since gravity wave drag is not commonly provided for the reanalysis data sets; we therefore recommend that future data sets provide daily parameterized gravity wave drag on the standard pressure levels.

6.5 Annular modes

The annular modes have been used to quantify the coupling between the stratosphere and troposphere, particularly that associated with SSW events (*e.g.*, *Kushner*, 2010; *Baldwin and Dunkerton*, 2001; *Thompson and Wallace*, 2000). In the troposphere, the annular modes characterize meridional shifts in the extratropical jet streams; a positive index indicates the jet is located poleward of its climatological position. The jet streams are associated with the extratropical storm tracks, so that the annular modes are linked with shifts in storm activity, particularly in Northern Europe and eastern North America (*e.g.*, *Thompson and Wallace*, 1998). In the stratosphere, the annular modes chiefly characterize variations in the strength of the polar vortex. A positive index indicates a stronger than average vortex, so that the breakdown of the vortex in an SSW is associated with an abrupt shift to a very negative annular mode index in the stratosphere.

The negative shift in the stratospheric annular mode index associated with an SSW typically precedes a similar (albeit weaker) shift towards a negative annular mode index in the troposphere by a few days (*Baldwin and Dunkerton*, 2001; *Karpechko et al.*, 2017). The equatorward shift in the tropospheric jet stream persists on the order of 30 to 60 days, associated with the slow recovery time scale of the lower stratospheric vortex (*e.g.*, *Gerber et al.*, 2010) and potential feedback with baroclinic eddies in the troposphere (*e.g.*, *Song and Robinson*, 2004). SSWs are therefore important for seasonal to subseasonal forecasts (*e.g.*, *Butler et al.*, 2019; *Domeisen et al.*, 2019b; *Sigmond et al.*, 2013).

In addition, the annular modes have been used to investigate cases where the polar vortex is stronger than average (*Baldwin and Dunkerton*, 2001; *McDaniel and Black*, 2005). These "Polar Vortex Intensification" events (hereafter strong vortex events) are somewhat of an opposite analogue to a SSW, but lack a clear, abrupt onset. A stronger than average polar vortex (*i.e.*, a positive annular mode state in the stratosphere) is typically associated with a poleward shift in the tropospheric jet (*i.e.*, a positive annular mode in the troposphere).

6.5.1 Consistency of the annular mode index across reanalyses in the post and pre-satellite periods

As detailed by *Gerber and Martineau* (2018), we use a simplified procedure to compute the daily annular mode indices from the reanalyses. As proposed by *Baldwin and Thompson* (2009), the annular mode index is defined by the polar cap averaged geopotential height (all latitudes poleward of 65°), normalized to have zero mean and unit variance. To ensure that the annular mode indices characterize meridional shifts in geopotential height at all levels, the global mean geopotential height on each pressure level is first removed at each time step before computing the polar cap averages (*Gerber et al.*, 2010).

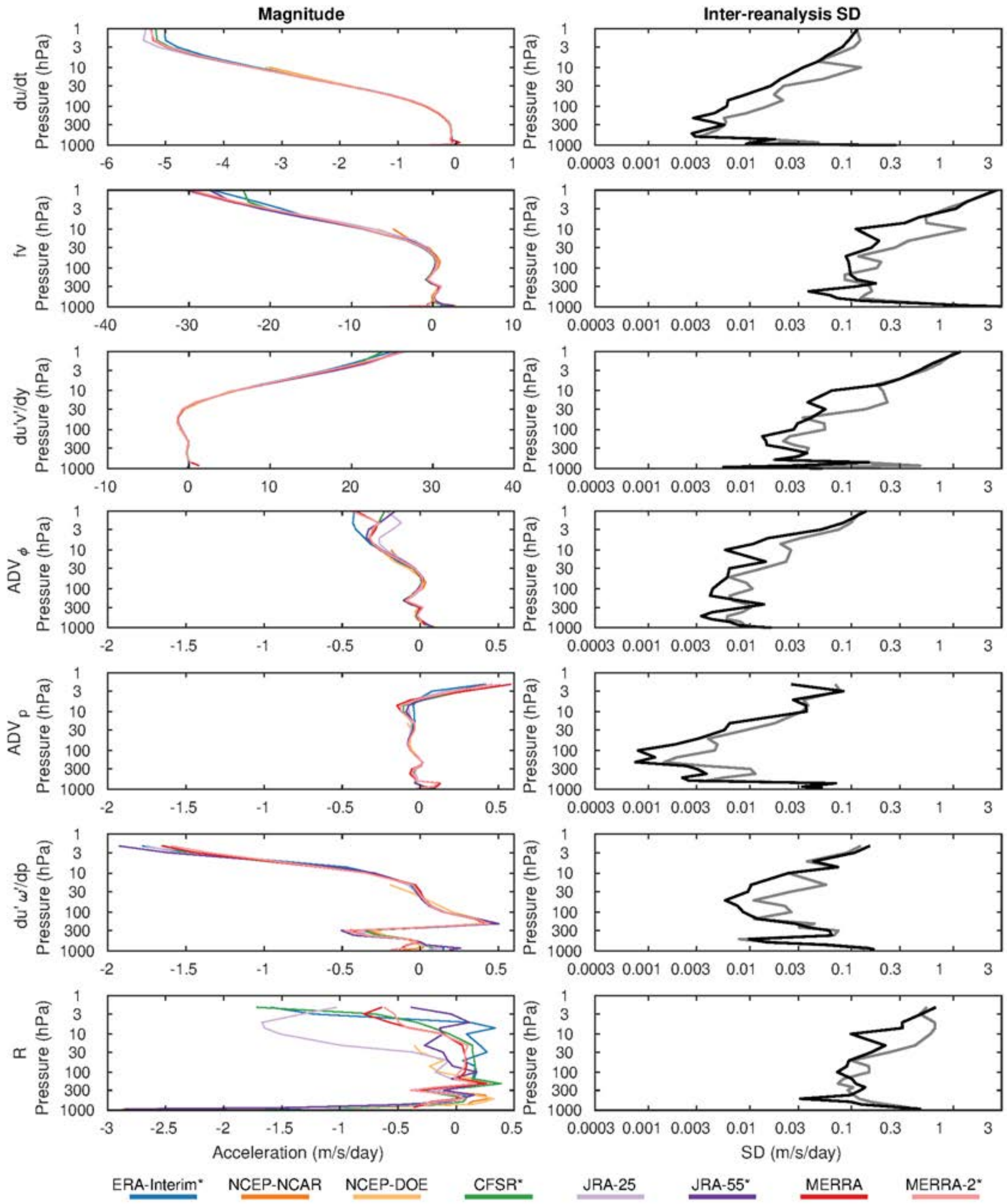


Figure 6.9: *Vertical profiles of each term in the momentum equation averaged from lags 5 to 0 days before SSW events. All variables are averaged between 45°N and 85°N. Individual reanalyses are shown to the left and the inter-reanalysis standard deviation is shown to the right on a logarithmic scale. The latter is shown for all reanalyses (grey) and for just the modern reanalyses (black; indicated with a * in the legend). All quantities are expressed in units of m s^{-1} day^{-1}. Reproduced from Martineau et al. (2018b).*

In keeping with the sign convention of *Thompson and Wallace* (2000), we also reverse the sign, so that a high index state is associated with a lower than average polar cap geopotential height.

This definition of the annular mode requires extrapolation of data to pressure levels below the surface in regions of high topography, which was done by the reanalysis centers with the exception of the MERRA products. To avoid introducing extrapolation errors, we omit MERRA and MERRA-2 from comparisons below 700 hPa. We focus on a subset of the pressure levels between 1000 hPa and 1 hPa that were shared by all reanalyses. Levels above 10 hPa, however, are unavailable for NCEP R1/NCEP R2 and 20CR v2/v2c reanalyses.

For the satellite era, 1979 onward, *Gerber and Martineau* (2018) found that a reanalysis ensemble mean (REM) constructed from the most recent reanalyses (ERA-Interim, JRA-55, and CFSR) provided a reliable benchmark for comparison. MERRA-2 was not included in the REM due to missing data below 700 hPa, but the results are nearly identical if it is included. The annular mode indices in the modern reanalyses are correlated $R^2 > 0.96$ with each other at all levels in the Northern Hemisphere and up to 3 hPa in the Southern Hemisphere; CFSR's correlation with the others drops to $R^2 = 0.9$ at 1 hPa in the Southern Hemisphere. For the pre- satellite period, it was unclear if a REM was meaningful, particularly in the austral hemisphere. In the analysis shown in **Figure 6.10**, JRA-55 is chosen among modern full-input reanalyses as an arbitrary point of comparison.

Figure 6.10 contrasts consistency between the reanalyses in the post- and pre-satellite periods. To assess performance during the satellite era, **Figures 6.10a** and **c** correlate the annular modes computed from each individual reanalysis with the REM index over the standard WMO climatological period, 1981 - 2010. Essentially the same results would be found for any period after 1979, with some evidence of greater agreement in the last decades at upper levels (not shown). In the Northern Hemisphere, the annular mode indices computed from all of the full-input reanalyses are almost indistinguishable (the squared correlations are near one). In the Southern Hemisphere, there is reasonable agreement between all the full-input reanalyses ($R^2 > 0.95$ up to 10 hPa), but with evidence of tighter agreement amongst the more recent reanalyses ($R^2 \approx 0.99$ up to 3 hPa). While not shown here, an early output of ERA5 (2008 - 2016) was compared with the other modern reanalyses by *Gerber and Martineau* (2018) and shown to be as good as the other modern reanalyses.

In the Northern Hemisphere, the conventional-input JRA-55C reanalysis provides a very good estimate of the state of the annular mode up to 10 hPa. JRA-55C's annular mode index, however, is noticeably less correlated with the REM in the Southern Hemisphere, suggesting the satellite observations are critical for quantifying the large-scale circulation of the austral hemisphere. At the surface, and throughout most of the troposphere, the surface-input reanalyses 20CRv2/v2c and ERA-20C are also well correlated with the REM. The annular mode indices in the 20CR reanalyses, however, quickly decorrelate with the REM above the tropopause, suggesting that these reanalyses cannot effectively capture stratospheric variability. ERA-20C also loses skill in the stratosphere, but much more slowly, particularly in the Northern Hemisphere. The R^2 of approximately 0.6 at 10 hPa indicates that ERA-20C captures 60 % of the variance in the annular mode at this height in the stratosphere. As discussed in greater detail by *Gerber and Martineau* (2018), ERA-20C appears to capture approximately half of the observed SSWs, and simulates the same frequency of events overall.

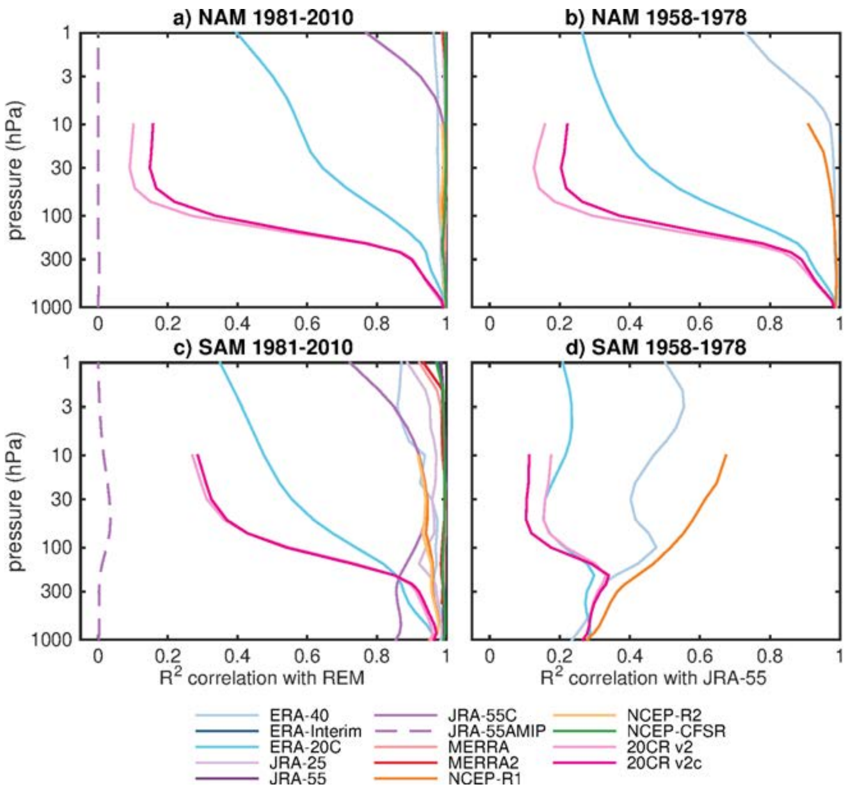


Figure 6.10: *The squared correlation between the (a, b) Northern and (c, d) Southern Annular Mode indices computed from each individual reanalysis with (a, c) a Reanalysis Ensemble Mean (REM) for the period 1981 - 2010, and (b, d) with the Annular Mode index of JRA-55 for the pre-satellite period, 1958 - 1978. As detailed in the text, the REM for the more recent period is constructed from three of the most recent reanalyses (ERA-Interim, JRA-55, and CFSR). In the pre-satellite period, a REM proved less meaningful. Comparable plots are obtained if NCEP R1 or ERA-40 are used instead of JRA-55. Adapted from Gerber and Martineau (2018).*

We note that the JRA-55AMIP integration does not meaningfully capture any of the annular mode variability. This was an expected result; this integration is not a reanalysis, but rather the JRA-55 model forced with observed SSTs, as in a standard Atmospheric Model Intercomparison Project (AMIP) simulation. Knowledge of the sea surface temperature is not sufficient to constrain the large-scale circulation of either hemisphere.

Only six reanalyses provide coverage in the pre-satellite era. Here we restrict ourselves to the period 1958 - 1979, as only NCEP R1 and the surface-input reanalyses extend further back in time, but *Gerber and Martineau* (2018) consider earlier periods. We have arbitrarily chosen JRA-55 as the reference time series among the modern full-input reanalyses for **Figures 6.10b** and **d**, but a qualitatively similar structure is found if ERA-40 or NCEP R1 is used instead. In the Northern Hemisphere, we find that the annular mode is consistently represented in the full-input reanalyses, with growing uncertainty above 10 hPa (where NCEP R1 is not available). This result is consistent with the ability of the conventional input reanalysis

JRA-55C to capture Northern Annular Mode variability in the satellite period.

While ERA-20C still captures more of the variability in the stratosphere in comparison to the 20CR reanalyses, the R^2 correlation is weaker in the pre-satellite period. At 10 hPa, ERA-20C captures only 40 % of the variability in the full-input reanalysis JRA-55 (or equivalently, ERA-40 and NCEP R1), compared to 60 % in in the satellite era. This could be due to fewer surface observations during this earlier period.

In the Southern Hemisphere, the situation is different. There is little agreement between JRA- 55 and the other reanalyses. Similarly poor agreement is found if NCEP R1 or ERA-40 is chosen as the reference time series (not shown), though we do find the NCEP R1 is somewhat better correlated with the surface-input reanalyses in the troposphere than either JRA-55 or ERA-40. The poor consistency between the reanalyses in the pre-satellite period was somewhat expected, given the inability of JRA-55C to capture the Southern Annular Mode in recent years. But the fact that JRA-55C still captures 85 % or more of the variance in the REM at nearly all levels suggests that a scarcity of conventional observations before 1979 is a larger part of the problem.

As discussed in *Gerber and Martineau* (2018), it is difficult to assess the synoptic variability of the Southern Annular Mode from direct measurements. On monthly time scales, *Marshall* (2003) has constructed a station based index that is correlated at approximately $R = 0.85$ with the 850 hPa Southern Annular Mode index in all reanalyses over the period 1979 - 2001. (This period was chosen to allow comparison with ERA-40.) For JRA-55 and ERA-40, this correlation drops markedly (to approximately 0.5) in the pre-satellite period 1958 - 1978. NCEP R1's correlation also weakens, but only drops to approximately 0.7. In contrast, the surface based reanalyses ERA-20C and 20CR maintain their correlation with the *Marshall* (2003) index.

The 20CR products, however, have been shown to miss most of stratospheric variability in earlier periods. Thus, for probing the large-scale circulation of the stratosphere-troposphere in the pre-satellite Southern Hemisphere atmosphere, ERA-20C might actually provide a more reliable estimate, even though NCEP R1, ERA-40, and JRA-55 assimilate radiosonde data and other free atmosphere observations.

6.5.2 Sampling uncertainty vs. reanalysis uncertainty

As found with the evolution of the stratosphere during an SSW event in *Section 6.4.3*, our ability to quantify the large-scale tropospheric response to SSWs and strong vortex events is primarily limited by the finite length of the reanalysis records, not differences between the reanalyses. **Figure 6.11** compares the sampling uncertainty in the "dripping paint" plots of *Baldwin and Dunkerton* (2001) to uncertainty associated with differences in the reanalyses. Panels (a) and (b) provide an update on the evolution of the annular mode index about weak and strong vortex events, now based on almost 6 decades of JRA-55 reanalysis. Following *Baldwin and Dunkerton* (2001), composites are centered about the date the 10 hPa index drops below -3 (rises above 1.5) standard deviations. The asymmetry in event criteria was based in part on the fact that the annular mode index at this level is skewed negative on account of SSWs, but 1.5 standard deviations is a much weaker threshold, such that more strong events are identified.

Gerber and Martineau (2018) show that using a consistent set of event dates is important for this comparison. The threshold nature of the event detection implies that very small differences between reanalyses can lead to the detection of different events (or more frequently, a shift in the timing of a given event). This effectively aliases sampling uncertainty in a comparison of reanalyses: the key is that the annular mode indices vary very little between reanalyses (differences are on the order of 1 %), but the inter-event variance is of order unity.

The weak vortex composite (**Figure 6.11a**) shows a rapid breakdown of the stratospheric polar vortex in the week preceding an event, evident first at upper levels, but become nearly synchronous in height by the time of onset. The stratospheric vortex then slowly recovers, from top to bottom, taking nearly three months in the lower stratosphere. During this long period of recovery, the tropospheric annular mode tends to be weakly negative, indicating an equatorward shift in the jet stream.

The strong vortex events (**Figure 6.11b**) exhibit a similar structure, but shifted earlier in time relative to date of event onset. The stratospheric vortex exhibits a positive annular mode (*i.e.*, is stronger than average) for over a month in advance, associated with a positive tropospheric annular mode (poleward shift in the tropospheric jet) that is already fully developed by the onset. This shift is partly due to the fact that strong vortex events tend to build slowly, on the time scale of radiative forcing, and so are harder to align in time. With respect to the amplitude of these events, pay close attention to the color scale. A weak vortex event is associated with a 3 standard deviation drop in the annular mode index in approximately 1 week, corresponding to a 1.4 km rise in the 10 hPa surface height at the pole. In contrast, the strong vortex event is associated with a more gradual 0.7 km drop in the 10 hPa surface over a month.

Figures 6.11c and **d** show the 1 standard deviation error bound on the weak and strong vortex composites, respectively. As shown by *Gerber and Martineau* (2018), inter-event variance of the annular mode indices is on the order of unity at all times except in the stratosphere at event onset (which occurs by construction: the 10 hPa index annular mode is always approximately -3 or 1.5 at lag 0). The sampling uncertainty of the composite is thus approximately 1 over the square root of the number of events. For weak vortex cases, where we have only 32 events, this is approximately 0.2, of the same order as the signal at any

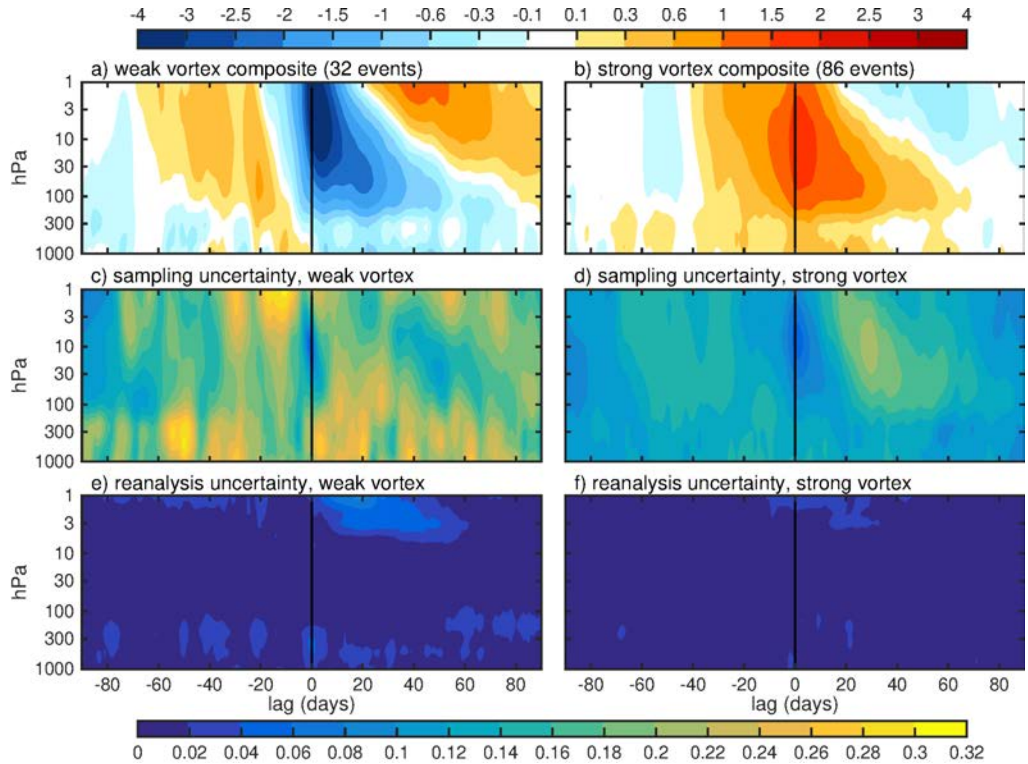


Figure 6.11: *Composites of the Northern Annular Mode indices as a function of lag and pressure for (a) weak and (b) strong vortex events, based on JRA-55 reanalyses over the period 1958-2016. Following Baldwin and Dunkerton (2001), weak (strong) events are identified when the NAM index at 10 hPa drops below -3 (rises above 1.5), and must be separated by a minimum of 30 days. The remaining panels quantify the uncertainty in the NAM index evolution as a function of lag and pressure. (c) and (d) show the sampling uncertainty in the mean weak/strong composites shown in **Figs. 6.11 (a) and (b)**, expressed as a one standard deviation error bound. Panels (e) and (f) show the reanalysis uncertainty: the standard deviation between composites of weak/strong vortex events based on the 4 most recent reanalysis products (ERA-Interim, JRA-55, MERRA-2, and CFSR/CFSv2) separately, for the period 1980 - 2016. As discussed in the text, a standardized set of event dates are used to prevent the aliasing of sampling error. Adapted from Gerber and Martineau (2018).*

given time! As argued by *Baldwin and Dunkerton* (2001), the tropospheric response is only significant if one averages over an extended period. This takes advantage of the fact that the tropospheric annular mode tends to exhibit memory on the order of 10 days (*e.g.*, *Gerber et al.*, 2010). If we ask for a 95 % confidence interval at any given time, we need the signal to be about equal to two standard deviations, requiring on the order of 100 events, a point we just approach in the case for strong vortex events.

Differences between the reanalyses are an order of magnitude smaller than the sampling error, as shown in **Figures 6.11e** and **f**. This measure of the "reanalysis uncertainty" was constructed by comparing weak and strong vortex composites based on the most recent reanalyses (ERA- Interim, JRA-55, MERRA-2, and CFSR/CFSv2) separately. We find that composites based on one reanalysis versus another are almost indistinguishable, provided one uses a standardized set of event dates. As the uncertainty is more than 10 times smaller than the sampling uncertainty, we'd need a record 100 times as long (*i.e.*, 6000 years!) for the choice of reanalysis to become as important as sampling uncertainty.

A similar conclusion applies to other measures of the coupling between the stratosphere and troposphere through the polar vortex, such as the variance and persistence of the annular mode indices as a function of season explored by *Baldwin et al.* (2003) and *Gerber et al.* (2010): results based on one reanalysis are not significantly different from those based on another with respect to the sampling uncertainty. This suggests that lengthening the reanalysis record has a substantial effect on our ability to quantify the coupling between the troposphere and stratosphere.

The sampling uncertainty shown in **Figures 6.11c** and **d** was based on JRA-55, which provides two additional decades (30 % more years) than the other most recent reanalyses which are restricted to the satellite era. As the sampling error decays with the square root of the number of events, these error bounds are 20 % smaller than could be obtained from the other modern reanalyses. This reduction depends on the assumption that JRA-55's reanalysis from 1958 - 1978 is of sufficiently high quality, supported by our comparison of the pre-satellite era reanalyses in **Figure 6.10**, and the fact that JRA-55C does a good job of capturing annular variability since 1979 without the aid of satellite observations. We look forward to assessing the ERA5 reanalysis, which is planned to extend back to 1950.

6.6 Stratospheric final warming events

The extratropical stratosphere exhibits a pronounced seasonal cycle with westerly winds in the winter hemisphere (with the exception of SSW events) and easterly winds in the summer hemisphere. The final transition from the westerlies to the easterlies, which occurs every year, is referred to as a Stratospheric Final Warming (SFW) event. Similar to SSW events, SFW events show a signature of zonal-mean zonal wind deceleration in the troposphere, indicative of a downward coupling, and a signature of enhanced upward Eliassen-Palm (EP) flux propagation to the stratosphere prior to the events (*Sun and Robinson*, 2009; *Black and McDaniel*, 2007). As such, they allow us to evaluate the representation of stratosphere-troposphere dynamical coupling in both hemispheres.

There is greater variability of final warmings in the Northern Hemisphere compared to the Southern Hemisphere, but stratospheric ozone loss has influenced their statistics in the Southern Hemisphere. Given their influence on the troposphere, the timing of the final warming has implications for seasonal forecasting (*e.g.*, *Butler et al.*, 2019; *Byrne and Shepherd*, 2018; *Lim et al.*, 2018; *Hardiman et al.*, 2011; *Ayarzagüena and Serrano*, 2009).

The final warming of the polar vortex is of key importance in chemistry-climate models. Once the polar vortex has broken down, ozone rich air can be transported to polar latitudes again. In the Southern Hemisphere, a late final warming in models will mean that the simulated Antarctic ozone hole persists longer through the year than is observed. A bias in the final warming time is also an indication of polar temperature biases, which will adversely affect the modelling of heterogeneous ozone destruction there (*Eyring et al.*, 2006). Adequate representation of the timing of the final warming in reanalysis data sets therefore has important implications for the evaluation of chemistry-climate models.

The final warming date is defined here as the day on which the zonal mean zonal wind at 60° becomes easterly for the final time during winter/spring. This can be sufficiently diagnosed using monthly mean data (calculating the day of the final warming using linear interpolation and assuming the monthly mean value represents the value on day 15 of the month) and occurs first in the mesosphere in the Southern Hemisphere (**Fig. 6.12**; shown only up to 1 hPa) but first in the mid-stratosphere in the Northern Hemisphere (Fig 6.12). With the exception of 20CR, all reanalysis products agree on the mean final warming date to within 6 days.

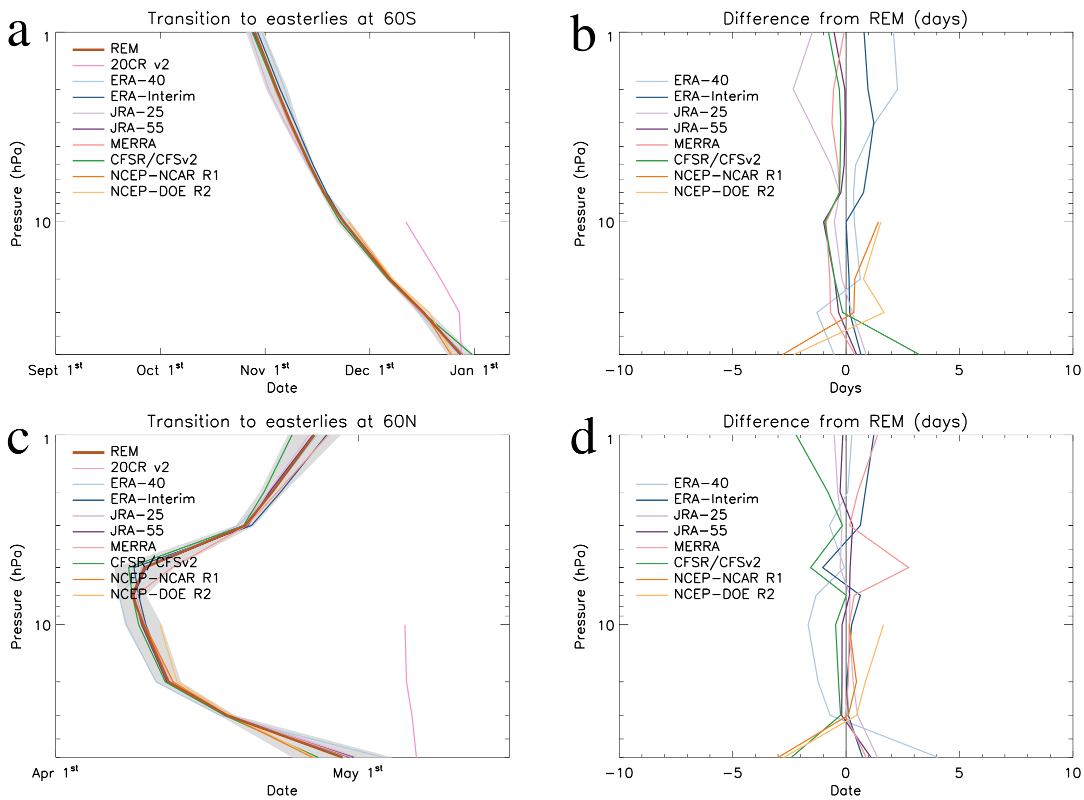


Figure 6.12: *The final transition of zonal mean zonal wind from westerly to easterly at (a) 60° S and (c) 60° N is shown for the period 1979-2010 for all reanalysis data products except 20CR v2 (which uses 1979-2009) and ERA-40 (which uses 1979-2002). The reanalysis ensemble mean (REM) is shown as a thick brown line, and uses data from all products except 20CR v2. The dark gray shading indicates the inter-reanalysis standard error (again excluding 20CR v2), scaled to represent a 95% confidence interval. The difference, in the final warming times shown in panels a and c, of each reanalysis from the multi-reanalysis mean is shown in (b,d). 20CR v2 is excluded from the REM since final warming times, especially in the Northern Hemisphere, are significantly later in this reanalysis and, given the remarkable agreement in final warming times across all other reanalysis datasets, the final warmings in 20CR v2 are very likely to be biased late.*

A closer study of the final warming in the Northern Hemisphere reveals that in some years the final warming occurs first in the mid-stratosphere ("10hPa-first years"), but in some years occurs first in the mesosphere ("1hPa-first years") (**Figure 6.13**). In 27 of the 32 years used, the reanalysis products all agree on the final warming type. Although there is generally a good agreement among full-input reanalyses, ERA-40 shows larger discrepancies in the mid- to lower- stratosphere transition date with respect to other data sets.

Correctly simulating the proportion of 10 hPa-first years and 1 hPa-first years is an area in which climate models do not currently perform well. In the reanalyses 68 - 79 % of years are 10 hPa-first years, whereas only 36 % of all modeled years, using the chemistry-climate models participating in phase 2 of the Chemistry-Climate model Validation activity (CCMVal-2) are 10 hPa-first years (*Hardiman et al.*, 2011). *Thiéblemont et al.* (2019) note a similar underestimation of 10 hPa-first years in the CESM and EMAC climate models.

6.7 Modulation of stratosphere-troposphere coupling by ENSO and QBO

The Northern-Hemisphere winter stratospheric polar vortex varies in strength from year to year with several external factors (*Yoden et al.*, 2002). One prominent source for this interannual variability is ENSO, the main mode of interannual variability in the tropical troposphere. During its warm phase (El Niño), Rossby wave trains propagate towards mid-latitudes in the Northern Hemisphere (NH) in boreal winter, strengthening the Aleutian low (*e.g.*, *Horel and Wallace*, 1981). As a consequence, upward propagation of planetary waves into the stratosphere is enhanced, which results in a weaker and a warmer polar stratosphere (*e.g.*, *Cagnazzo and Manzini*, 2009; *Brönnimann*, 2007). Although its teleconnectivity to the stratosphere is weaker than El Niño, the cold ENSO phase, La Niña, weakens the Aleutian low leading to reduced upward-propagating wave activity into the stratosphere and a strengthening of the polar vortex (*Iza et al.*, 2016; *Butler and Polvani*, 2011). For a comprehensive review of ENSO-stratosphere teleconnections, see *Domeisen et al.* (2019a).

ENSO's influence on the extratropical circulation is not limited to the time-mean flow. *Barriopedro and Calvo* (2014) found an ENSO modulation of the blocking precursors of SSWs, leading to distinctive wave signatures of SSWs during opposite ENSO phases: during El Niño, SSWs are predominantly associated with wavenumber-1 amplification in the lower stratosphere, whereas La Niña SSWs tend to occur after wavenumber-2 amplification (see also *Song and Son* (2018)). The way blocking events interfere with stationary waves and either amplify or damp the total injection of wave activity into the stratosphere depends critically on their location (*e.g.*, *Nishii et al.*, 2011; *Castanheira and Barriopedro*, 2010; *Woollings et al.*, 2010; *Martius et al.*, 2009).

Another source of interannual variability of the strength of the stratospheric polar vortex is the QBO which can modulate the nature and propagation of extratropical planetary-scale waves (*Garfinkel et al.*, 2012b; *Holton and Tan*, 1980). Several studies (*Taguchi*, 2015; *Richter et al.*, 2011; *Calvo et al.*, 2009; *Garfinkel and Hartmann*, 2007) further suggested some nonlinear influence of QBO and ENSO onto the stratospheric polar vortex such that when the QBO is in a westerly phase in the lower stratosphere, the polar night jet weakens and SSW probability increases for the warm ENSO phase (El Niño), whereas the changes are opposite for the QBO easterly winters.

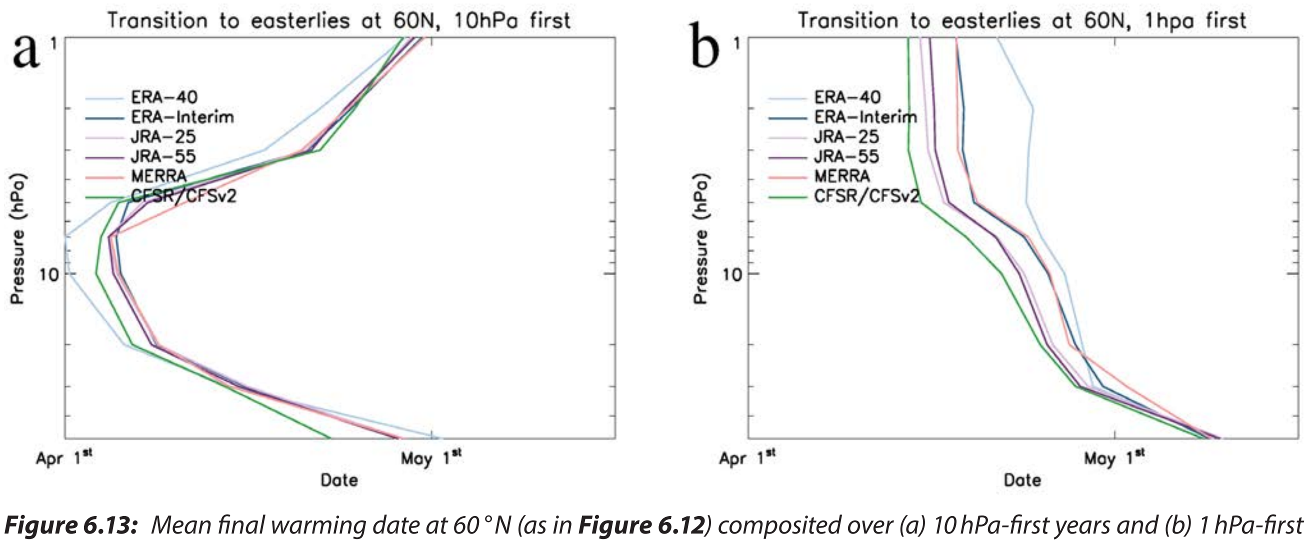


Figure 6.13: *Mean final warming date at 60 °N (as in **Figure 6.12**) composited over (a) 10 hPa-first years and (b) 1 hPa-first years (defined in text). The percentage of 10 hPa-first years is: 73.9 in ERA-40, 78.1 in ERA-Interim, 75.0 in JRA-25, 75.0 in JRA-55, 78.1 in MERRA, and 68.8 in CFSR. Data from the reanalyses 20CR, NCEP R1, and NCEP R2 does not extend above 10 hPa, so these products cannot be used for this diagnostic.*

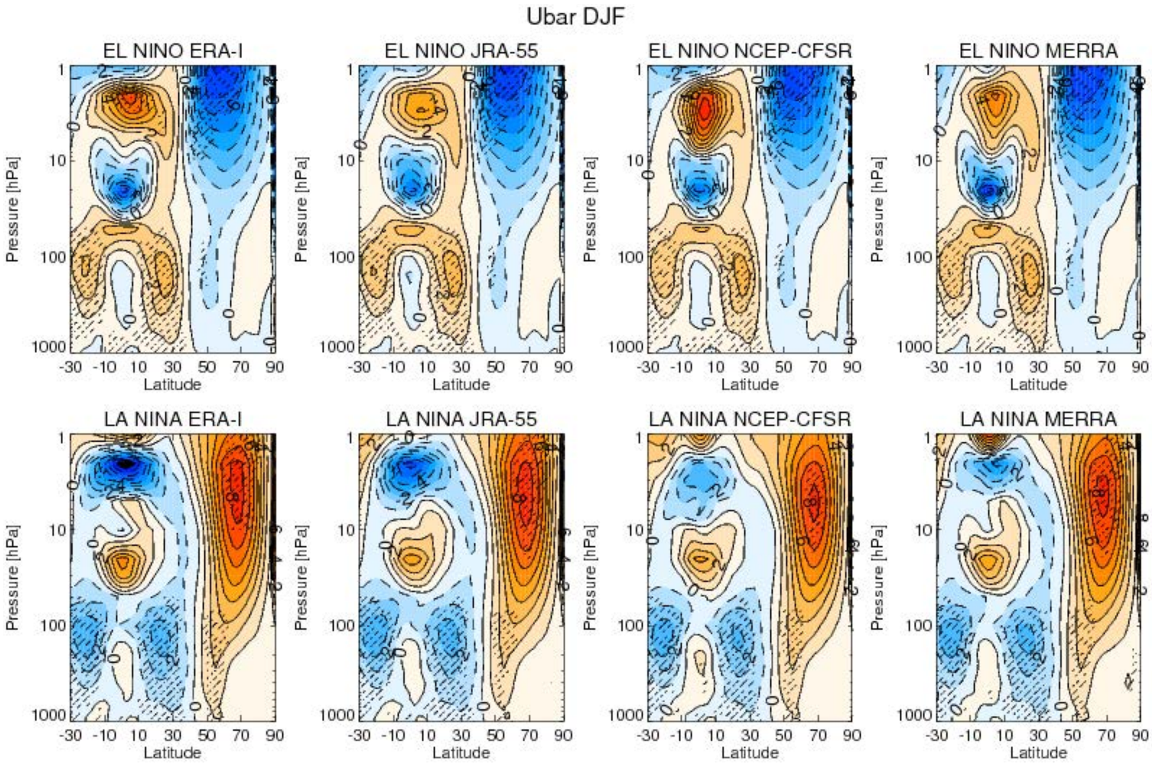


Figure 6.14: *Latitude-pressure cross sections of the composited DJF average of monthly zonal mean zonal wind anomalies for (top) El Niño and (bottom) La Niña events, from left to right for ERA-Interim, JRA-55, CFSR, and MERRA reanalyses. Contour intervals are ± 1 m s⁻¹. Solid (dashed) contours denote positive (negative) anomalies. Stippling indicates significance at the 95 % level.*

Alternative criteria have been used in the literature to define cold and warm ENSO phases. In the following analyses, we have focused on the most commonly used Niño 3.4 index based on monthly mean SST anomalies in the region from 5°S-5°N and 170°E-20°W with reference to 1981-2010 climatology. Standard El Niño and La Niña phases are defined by plus or minus 0.5K anomalies in this region, as done in *Section 6.7.2*. In *Sections 6.7.1* and *6.7.3*, more restrictive criteria (1 standard deviation anomalies) were applied to focus on more extreme events. The period of averaging, DJF vs. a more extended winter season, was also varied depending on the scientific focus. The criteria for selecting warm and cold phases and the resulting years are therefore listed in each section; winters are identified by the year in January, *e.g.*, 1983 refers to the 1982-1983 winter.

6.7.1 Troposphere-stratosphere coupling through ENSO

The wintertime-mean stratospheric response to El Niño and La Niña conditions is first compared among reanalysis datasets. Monthly mean data from ERA-Interim, JRA-55, CFSR, and MERRA reanalyses are used. First, for each field, time series from 1979 to 2013 are detrended and anomalies are computed with respect to the 1981-2010 climatology. El Niño and La Niña events are defined using the standardized NDJF sea surface temperature anomaly of the Niño 3.4 index from the NCEP-CPC. El Niño (La Niña) winters are selected above (below) 1 SD (-1 SD). The composites include 7 El Niño winters (1983, 1987, 1992, 1995, 1998, 2003, 2010) and 5 La Niña winters (1989, 1999, 2000, 2008, 2011). The statistical significance of the composites is assessed with a Monte Carlo test at the 95 % confidence level.

Figure 6.14 shows the latitude-pressure cross-section of December-January-February (DJF) average of the zonal mean zonal wind anomalies composited for El Niño (up) and La Niña (bottom) events. In the polar stratosphere the El Niño (La Niña) signal is characterized by a robust weakening (strengthening) of the zonal mean zonal wind in all reanalyses. All reanalyses agree on the significant area and the sign of the anomalies, with the largest polar stratospheric signal peaking at -7 m s⁻¹ for El Niño and 8 m s⁻¹ for La Niña. Therefore, a good agreement across reanalyses is found for El Niño and La Niña polar stratospheric responses.

To quantify the relationship between the strength of the Aleutian low, modulated by ENSO, and the response of the stratospheric polar vortex, **Figure 6.15** shows the scatter plot of the Z index at 500 hPa (average of geopotential height anomalies between 40°N-60°N and 180°E-210°E) versus the U index (zonal mean zonal wind averaged at 60°N between 10 hPa and 30 hPa), similar to *Cagnazzo et al.* (2009). It is important to note that these Z and U index values for each event are very similar among reanalyses. El Niño winters (squares) are associated with negative values of the Z and U indices. This corroborates that the deepened Aleutian low related

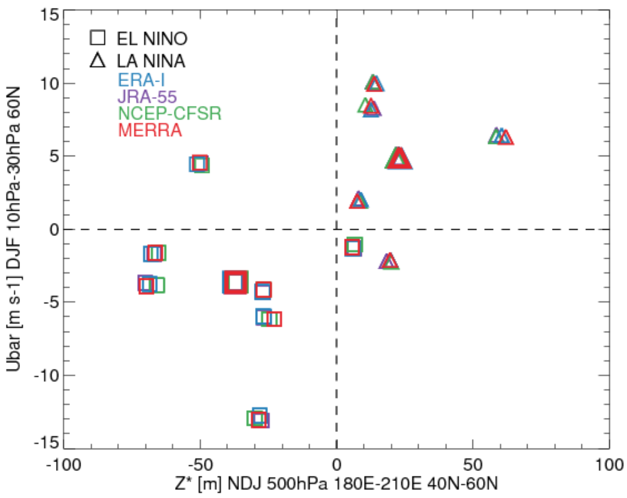


Figure 6.15: *Scatter plot of the NDJ mean Z index, versus DJF U index. Squares (triangles) represent each El Niño (La Niña) event and the corresponding larger symbols represent El Niño and La Niña events composite.*

to the negative Z index increases the upward wave propagation into the stratosphere leading to a weaker polar vortex. In contrast, La Niña winters (triangles) are mainly related to positive Z and U indices, due to an anomalously weak Aleutian low and in agreement with the observed positive wind anomalies respectively, since a weakened Aleutian low inhibits the upward wave propagation leading to a stronger polar vortex. Results show an excellent agreement among reanalyses. Therefore, we conclude that for the purpose of studying the coupling between the stratosphere and the troposphere during El Niño and La Niña events, any of the compared reanalyses is equally suitable, as *Iza et al.* (2016) noted for La Niña events.

6.7.2 Blocking patterns associated to SSWs and the modulation of ENSO

The intercomparison of ENSO's influence on the stratosphere among reanalyses is then extended to ENSO's influence on SSW events and their blocking precursors. The analysis contrasts inter- dataset uncertainties with the uncertainties associated with the definition of blocking events by using three different blocking definitions.

Daily mean geopotential height at 500 hPa (Z500) and 100 hPa (Z100) is used for this analysis which is performed for the full 1958 - 2012 period and the 1979 - 2012 satellite period. For the latter period, the REM is computed from the CFSR, ERA-Interim, JRA-55 and MERRA reanalyses. The REM of the full period is based on the NCEP R1, ERA-40 (completed with ERA-Interim from 2002 to 2012 since the two agree well over their overlapping period from 1979 to 2002) and JRA-55 reanalyses. Fields were interpolated (if required) to the same common 2.5° × 2.5° grid before any further analysis is carried out. Anomalies are defined with respect to the daily climatology of 1981 - 2010.

SSW central dates are chosen from the common dates identified in **Table 6.2**. ENSO winters were characterized by the NDJFM average of the monthly Niño 3.4 index (**http://www.cpc.ncep.noaa.gov/**). EN and LN winters were identified when Niño 3.4 ≥ 0.50 °C and Niño 3.4 ≤ 0.50 °C, respectively. The resulting warm phase years are 1958, 1966, 1969, 1973, 1983, 1987, 1988, 1992, 1995, 1998, 2003, 2005, 2007, and 2010. Cold phases were identified in 1962, 1963, 1965, 1967, 1968, 1971, 1972, 1974, 1975, 1976, 1984, 1985, 1989, 1996, 1999, 2000, 2001, 2006, 2008, 2009, 2011, and 2012.

We employed three blocking detection methods, which cover most approaches to blocking definition: 1) the absolute method (ABS), based on the detection of reversals in the meridional Z500 gradient; 2) the anomaly method (ANO), using Z500 anomalies above a given threshold; 3) the mixed method (MIX), a hybrid definition of the two previous approaches. These definitions are described in more details in *Woollings et al.* (2018). All methods give two preferred regions for blocking occurrence: one over the Atlantic and one over the Pacific basins, with maximum blocking frequencies of about 15 % of days in NDJFM. However, there are substantial differences among definitions in the blocking location within each basin as well as in the relative frequencies of Atlantic vs Pacific blocking (*Woollings et al.*, 2018).

Blocking precursors of SSWs were identified for each reanalysis by performing 2-D composites of blocking frequency for the [-10,0]-day period before the central dates of SSWs. This was carried out separately for SSWs occurring during El Niño and La Niña winters. The REM for the full period is shown in **Figure 6.16**. There is a spatial preference for different blocking precursors of SSWs depending on the ENSO phase, with enhanced (reduced) blocking frequencies over eastern North America and the North Alantic (eastern Pacific) during El Niño, and nearly opposite patterns for La Niña winters. Thus, SSWs are often preceded by North Atlantic sector blocking during El Niño, while eastern Pacific blocks are the preferred precursors of SSWs in La Niña winters. The comparison across reanalyses reveals a good agreement, with differences that are much smaller than among blocking definitions (everywhere except the blue dots in **Figure 6.16**). The intensity, significance and spatial extension of the signal weaken for the satellite period (1979 - 2012, not shown).

The composites of blocking precursors of SSWs for El Niño and La Niña winters are similar to those obtained for W1 and W2 SSWs, respectively (*Ayarzagüena et al.*, 2019; *Song and Son*, 2018), which hints at a modulation of the characteristics of SSW events. To further illustrate the association between ENSO and the dominant wave signatures of SSWs, the temporal evolution of Z100 wavenumber components are evaluated for the [-30, 30]-day period surrounding the central date of SSWs (**Figure 6.17**). The results confirm that SSWs are significantly preceded by wavenumber-1 amplification during El Niño, whereas SSWs preferably occur after wavenumber-2 amplification in La Niña winters (*Taguchi and Hartmann*, 2006).

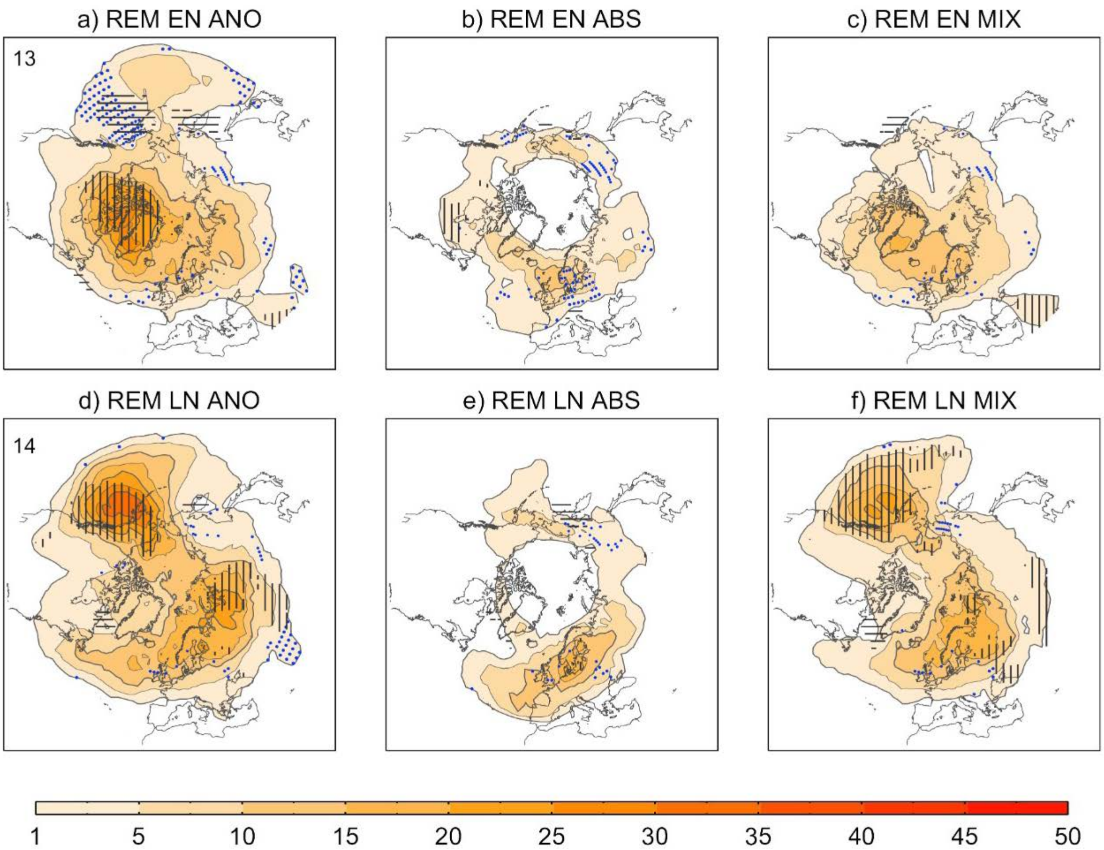


Figure 6.16: *Reanalysis ensemble mean composites of blocking frequency for the [-10,0]-day period before the central dates of SSWs occurring during El Niño (top) and La Niña (bottom) winters of the 1958-2012 period for three different blocking definitions (columns). The blocking frequency is expressed as the percentage of time (over the 11-day period) during which a blocking was detected at each grid point. Vertical (horizontal) black lines indicate regions with blocking activity significantly higher (lower) than the climatology at the 90% confidence level in at least 66% of the reanalyses. The significance is derived from a bootstrap of 1000 members, each one containing the same number of cases and dates as the SSWs of each composite but with random years of occurrence. Blue dots highlight grid points where the inter-reanalysis spread for a given blocking definition is larger than the spread across the reanalysis ensemble mean of blocking definitions. The numbers in the upper left corner of panels a), d) indicate the sample size of SSWs during El Niño and La Niña winters, respectively. Adapted from Ayarzagüena et al. (2019).*

During La Niña, the wavenumber-2 signal is accompanied by significant anomalies in wavenumber-1, albeit they are smaller and/or shorter-lasting. This difference in wave driving does not, however, necessarily affect the ratio of vortex splits to displacement (*Garfinkel et al.*, 2012a). This modulation of SSW characteristics by ENSO is achieved through a change in the preferred blocking location, which injects different scales of wave activity into the stratosphere, and thus forces different types of SSWs (*e.g., Barriopedro and Calvo*, 2014). We note that the sensitivity of these results to the choice of reanalysis is very weak.

The modulation of SSW properties by ENSO is robustly observed across reanalyses when the 1958-2012 period is used, but less evident in the 1979-2012 period. This suggests decadal variability in the ENSO-blocking-SSW relationship (*e.g., Rao et al.*, 2019), biases in the pre-satellite period or sampling issues affecting the shorter satellite period. The differences among blocking definitions are much larger than differences among reanalyses, likely contributing to the discrepancies in the blocking-SSW relationship reported in the literature.

6.7.3 Nonlinear modulation of the extratropical stratosphere by ENSO and QBO

Finally, we evaluate the representation of nonlinearities in the modulation of DJF-mean polar vortex strength and SSW occurrence with ENSO and QBO among reanalyses. SSW onset dates are defined by the common dates established in *Section 6.4.1* and the DJF zonal mean zonal wind at 60°N and 10hPa is used as a proxy for strength of the polar night jet. The analysis period ranges from 1979 to 2011, except for 20CR v2 (1979-2010). The DJF climatology for each reanalysis is based on the 1981-2010 period.

In order to define ENSO phases, the monthly Niño 3.4 index (provided by NOAA/CPC) is averaged over DJF. The DJF mean of the zonal mean zonal wind at the equator and 50hPa in the respective reanalyses is used to define QBO phases. All DJF seasons are classified into six groups defined by three ENSO and two QBO conditions. Two of the three ENSO conditions are El Niño and La Niña,

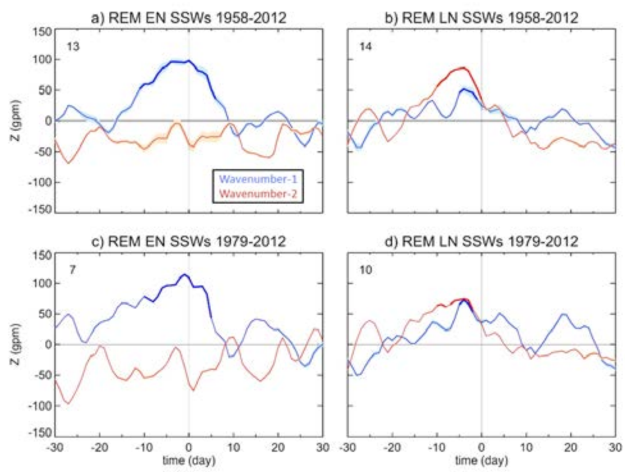


*Figure 6.17: REM composites of the temporal evolution of 100 hPa geopotential height wavenumber-1 (blue) and wavenumber-2 (red) amplitude anomalies at 60° N (gpm) for the [-30,30]-day period around the central dates of SSWs occurring during El Niño (left) and La Niña (right) winters of the full (1958-2012, top) and satellite (1978-2012, bottom) period. Shading denotes the ±2 sigma level across reanalyses. The time intervals highlighted with thick lines indicate significant differences with respect to climatology at the 95% confidence level in at least 66% of the reanalyses. The significance is assessed with a bootstrap test of 1000 samples with the same number of cases and calendar days as the SSWs of each composite but with random years of occurrence. The numbers in the upper left corner of each panel indicate the sample size of the composite. This figure differs from **Fig. 6** of Barriopedro and Calvo (2014) due to the addition of an extra SSW and a different ENSO classification with updated ENSO indices.*

when the DJF mean Niño 3.4 index exceeds ±1 standard deviation (both inclusive); warm phase years (EN) were identified in 1983, 1987, 1992, 1998, and 2010, and cold phases (LN) in 1989, 1999, 2000, 2008, and 2011. The third ENSO condition is neutral (NT) for remaining years. The mean and standard deviation of the ENSO index are calculated for the 1981-2010 period.

The two QBO conditions are easterly and westerly, when the DJF mean zonal wind in the equatorial lower stratosphere is negative or positive (the latter includes zero). The resultant grouping is slightly different among the reanalyses, as the equatorial zonal wind is different. Alternatively, we also use 50 hPa equatorial zonal wind data archived at Free University Berlin to standardize the QBO classification for all reanalyses, which yields similar results (not shown). It is noted that the classification of about 30 years into the six groups implies that in some cases the sample size is small and therefore it is difficult to obtain statistically significant results. Before focusing on changes with ENSO and QBO, it will be useful to mention that except for 20CR v2, the interannual variability of the DJF mean vortex strength is highly correlated between the reanalyses, with correlation coefficients over +0.99.

Figure 6.18 shows heat maps based on composite zonal wind anomalies of each reanalysis for the six groups. As expected from the high correlations of the interannual variability of the zonal wind, the plots show that the changes in the vortex strength with ENSO and QBO are more or less similar among the reanalyses, except for 20CR v2 whose climatological vortex strength is notably more than twice that of other reanalyses. Specifically, the zonal wind anomalies tend to be slightly negative for easterly QBO winters regardless of the ENSO conditions. For westerly QBO conditions, the zonal wind anomalies exhibit a clear decreasing tendency with the ENSO SST conditions, from La Niña (positive wind anomalies) through neutral to El Niño (negative anomalies).

It is noted that the sample sizes of the six groups are not the same among all reanalyses, implying that the equatorial zonal wind and hence QBO classification are different in some cases. This may matter when one extracts changes with QBO conditions that are defined using an equatorial zonal wind index in the respective reanalyses, but this effect seems limited here since most reanalyses show similar results.

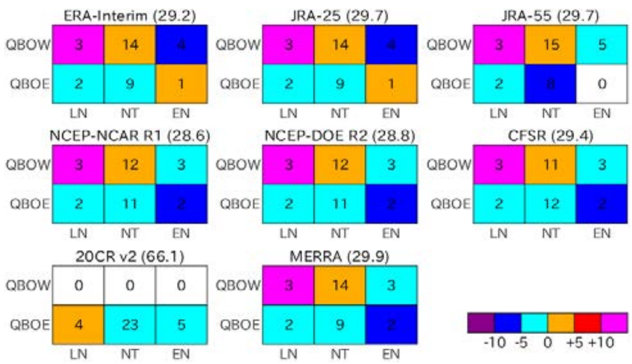


Figure 6.19: Heat maps showing composite anomalies of the DJF-mean zonal wind at 60° N, 10 hPa for the six groups in the eight reanalyses as indicated above each panel. The number indicated in each cell denotes the sample size. Each panel also includes the DJF climatological wind value in ms^{-1} in parentheses.

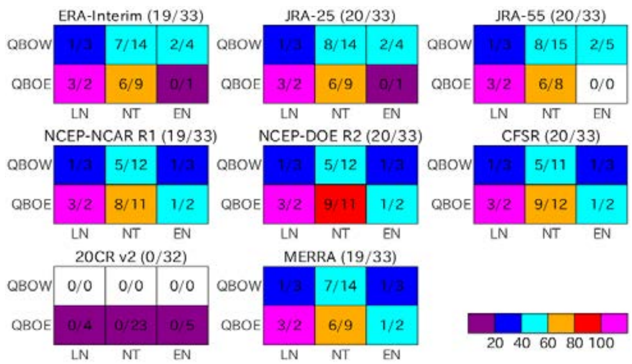


*Figure 6.18: Same as **Figure 6.18**, but for SSW probabilities (in %) computed as the ratio of the number of SSWs to the number of years for each group (indicated in each cell). The numbers in the title indicate total number of SSW / total number of years.*

Figure 6.19 similarly shows heat maps of SSW probabilities for the six groups. Here, for each group, the SSW probability is a ratio (times 100) of the number of SSWs to the number of years. The charts show that although the classification of QBO years is slightly different among the reanalyses, the changes in the SSW probabilities are similar among the data sets, except for 20CR v2 which has no SSWs. For easterly QBO, the probabilities tend to decrease from La Niña, through neutral, to El Niño. A characteristic feature is the highest SSW probability for La Niña and easterly QBO (group 1). On the other hand, for westerly QBO, the probabilities slightly increase in the opposite way, from La Niña, through neutral, to El Niño, consistent with the changes in the zonal wind anomalies (**Figure 6.18**). These changes in the SSW probabilities do not necessarily match changes in DJF zonal wind anomalies (**Figure 6.18**), since the occurrence or absence of a SSW during each winter depends not only on the DJF mean vortex strength but also on its variance.

6.8 Stratosphere-troposphere coupling through the antarctic hzone hole

In recent decades, severe stratospheric ozone depletion has led to the Antarctic ozone hole in austral spring (*Thompson and Solomon*, 2002). This has resulted in substantial cooling in the lower stratosphere, leading to an increase in the latitudinal temperature gradient and a consequent strengthening of the stratospheric polar vortex. Through mid-to-late spring, this mid-to-high latitude circulation anomaly descends from the lower stratosphere to reach the troposphere during austral summer (*e.g.*, *Son et al.*, 2018). The anomalous tropospheric circulation is associated with a noticeable increase in zonal mean sea level pressure difference between the mid and high latitudes, commonly referred to as an increase in the positive phase of the Southern Annular Mode (SAM). The positive SAM is generally marked by a poleward displacement and intensification of the tropospheric mid-latitude jet.

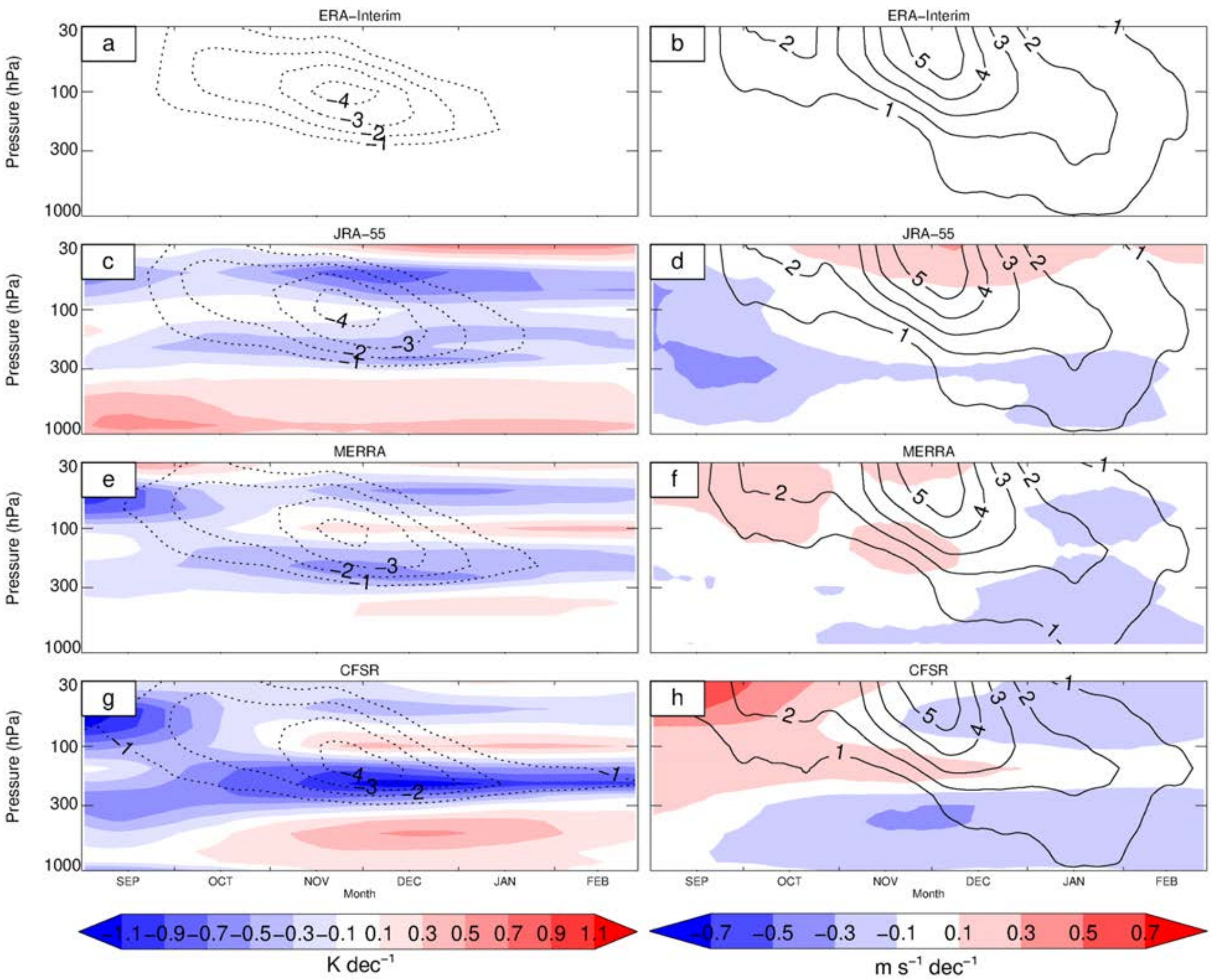


***Figure 6.20:** (left) The dashed contours show time-height cross sections of zonal-mean temperature trend (with contour intervals of 1 K dec⁻¹) averaged over latitudes 60-90° S during 1979-2001 for a) ERA-Interim, c) JRA-55, e) MERRA, and g) CFSR. The shadings show the differences between the various reanalyses and ERA-Interim at intervals of ±0.1, ±0.3, ±0.5, ±0.7, ±0.9 and ±1.1 K dec⁻¹. (right) The contours show time-height cross sections of zonal-mean zonal wind trend (with contour intervals of 1 m s⁻¹ dec⁻¹) averaged over latitudes 50-70° S during 1979 - 2001 for b) ERA-Interim, d) JRA-55, f) MERRA, and h) CFSR. The shadings show the differences between the various reanalyses and ERA-Interim at intervals of ±0.1, ±0.3, ±0.5, and ±0.7 m s⁻¹ dec⁻¹.*

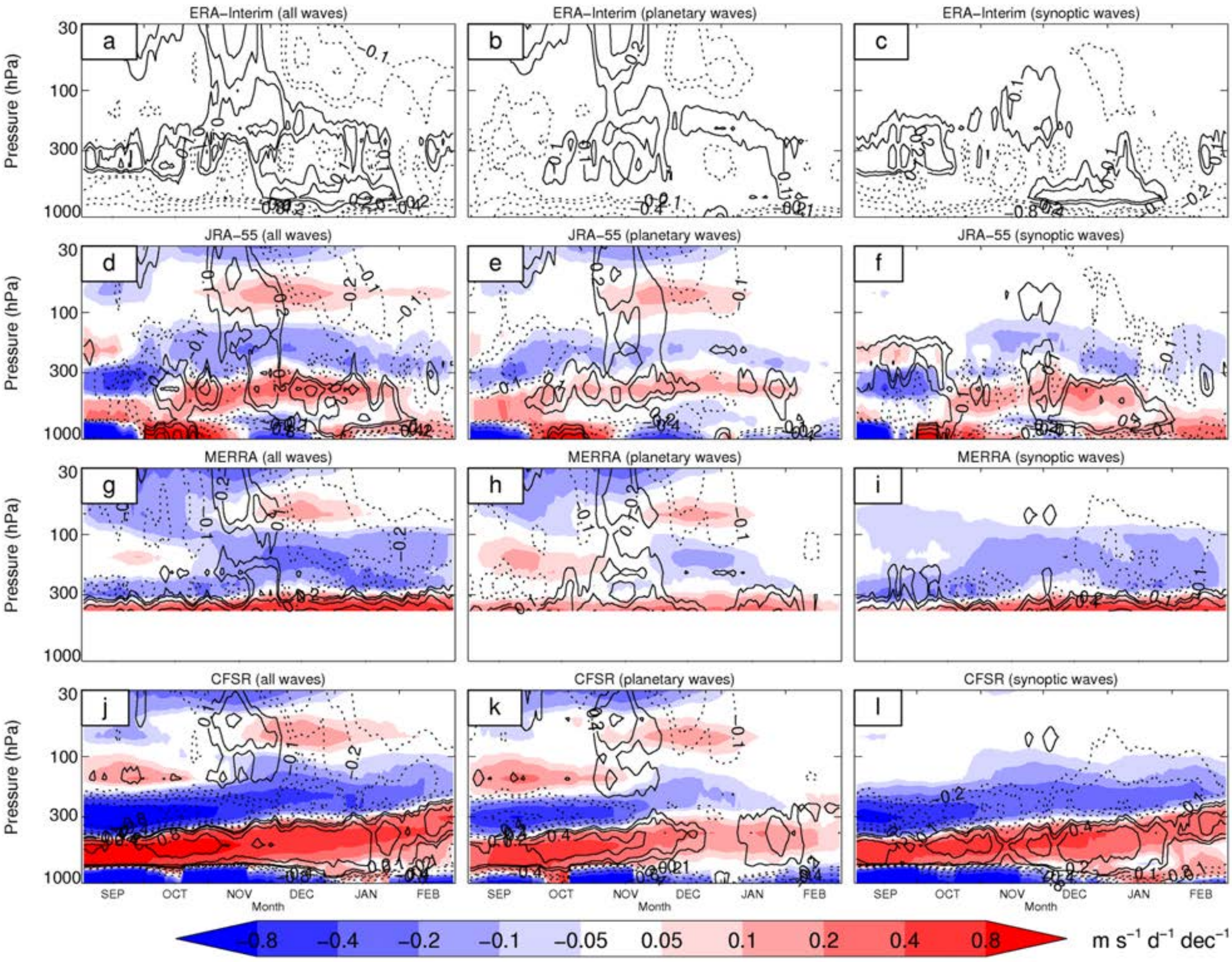


Figure 6.21: *The contours show time-height cross sections of the trend in the EP flux divergence due to all waves (left), planetary-scale waves (middle) and synoptic-scale waves (right) averaged over the latitude band of 40-80° S during 1979 - 2001 for (a-c) ERA-Interim, (d-f) JRA-55, (g-i) MERRA, and (j-l) CFSR at intervals of ±0.1, ±0.2, ±0.4 and ±0.8 m s^{-1} d^{-1} dec^{-1}. Solid and dashed contours indicate positive and negative values, respectively. The shadings show the differences between the various reanalyses and ERA-Interim at intervals of ±0.05, ±0.1, ±0.2, ±0.4 and ±0.8 m s^{-1} d^{-1} dec^{-1}. Note that MERRA data below 400 hPa is excluded.*

Although ozone loss has a direct impact on stratospheric temperatures by reducing the absorption of incoming solar radiation, a number of studies show that the anomalous circulation is strongly influenced by changes to wave forcing and wave mean-flow interaction (*Orr et al.*, 2012). Here, the impacts of the ozone hole on the dynamical coupling between the stratosphere and the troposphere in the spring and summer Southern Hemisphere are examined in the ERA-Interim, JRA-55, MERRA, and CFSR reanalyses datasets. A more detailed analysis is provided in Orr et al. (2021).

Figure 6.20 shows the trends in zonal mean temperature over the SH polar region between 1979 and 2001 for the four datasets. This period is chosen for two reasons. First, the size of the ozone hole increased steadily during this period (*Huck et al.*, 2007). Second, the trends in the four reanalyses were largest for this period (not shown), which allows to identify important differences between the datasets. In ERA-Interim, the cooling starts at 30 hPa in October and peaks at around 100 hPa between mid-November and early December (with trends reaching -4 K per decade), which is in good agreement with radiosonde data from Antarctica (*Thompson and Solomon*, 2002). The other three reanalyses all show broadly similar results with downward descent pattern from 30 hPa to 300 hPa. However, compared to ERA-Interim, CFSR shows considerably stronger and longer-lasting cooling (by up to -1 K dec^{-1}) between 100 hPa and 300 hPa, and enhanced warming below 300 hPa (by around 0.5 K dec^{-1}). This would lead to a comparative weakening of the atmospheric stability near the tropopause. In both CFSR and MERRA, the cooling also starts noticeably earlier than ERA-Interim.

Figure 6.20 also shows the corresponding trends in zonal wind over the SH polar regions, with all four reanalyses showing the expected strengthening of the SH circumpolar winds from the lower stratosphere down to the surface. In ERA-Interim, the strengthening starts in mid-to-late September at 30 hPa, peaks at around 5 m s^{-1} dec^{-1} between late November and early December, and reaches the lower troposphere in January. The results from ERA-Interim, JRA-55 and MERRA are in relatively good agreement, with differences not exceeding ± 0.3 m s^{-1} dec^{-1}.

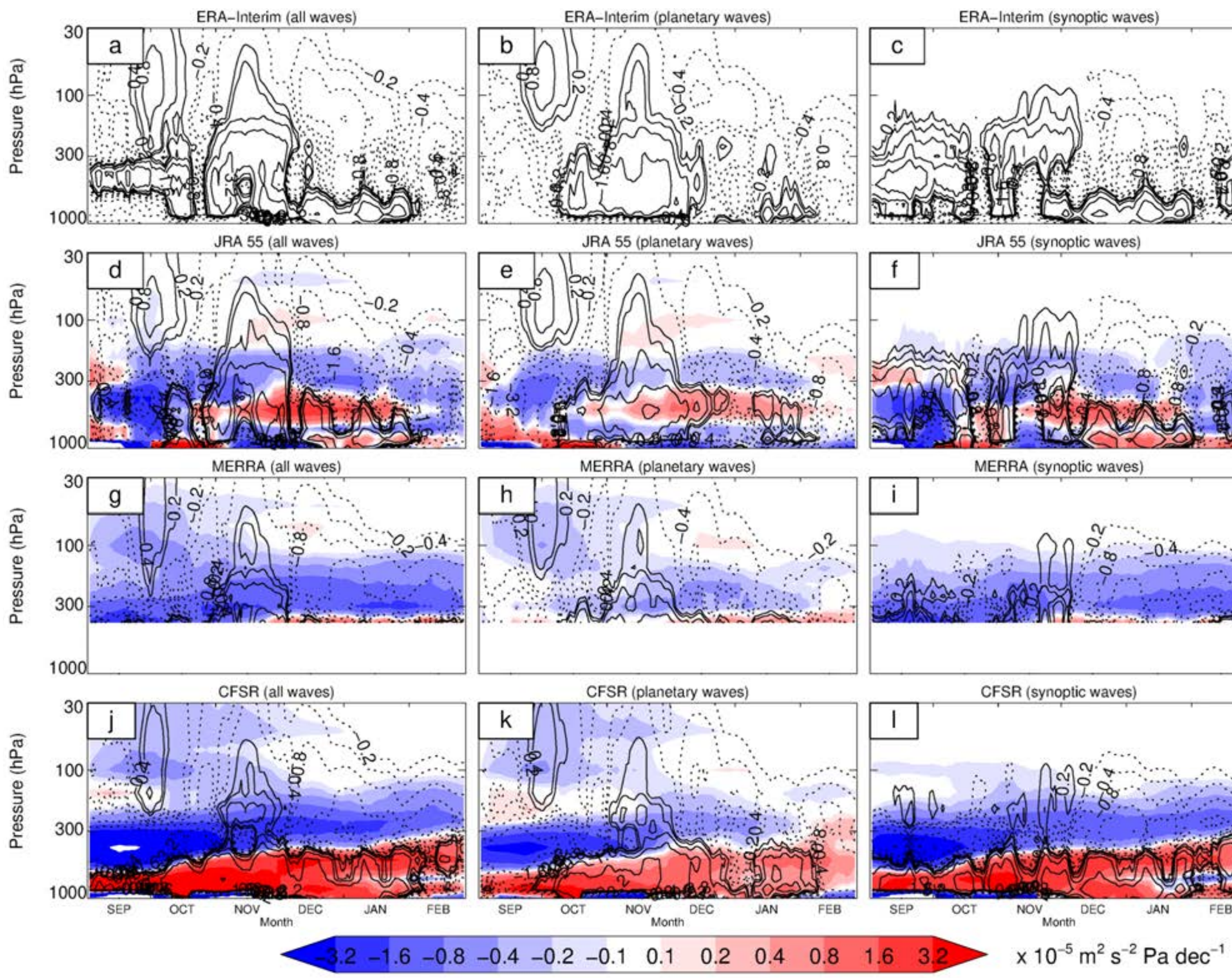


Figure 6.22: *The contours show time-height cross sections of the trend in the vertical component of EP flux due to all waves (left), planetary-scale waves (middle) and synoptic-scale waves (right) averaged over the latitude band of 40-80°S during 1979-2001 for (a-c) ERA-Interim, (d-f) JRA-55, (g-i) MERRA, and (j-l) CFSR at intervals of 0.2, 0.4, 0.8, 1.6, 3.2 and 10.0 x $10^{-5} m^2 s^{-2}$ Pa dec^{-1}. Solid and dashed contours indicate positive and negative values, respectively. The shadings show the differences between the various reanalyses and ERA-Interim at intervals of ±0.1, 0.2, 0.4, 0.8, 1.6 and 3.2 x $10^{-5} m^2 s^{-2}$ Pa dec^{-1}. Note that MERRA data below 400 hPa is excluded.*

However, in CFSR the initial strengthening of the winds in the lower stratosphere occurs earlier than in ERA-Interim, while in the lower troposphere they are delayed, indicating a comparatively slower downward descent rate in CFSR.

Figure 6.21 shows the trends of the total Eliassen-Palm (EP) flux divergence from 40-80°S derived from the four datasets. We note that MERRA data is excluded from this analysis below 400 hPa because, unlike other reanalyses, data is not extrapolated below the surface and thus zonal-mean diagnostics are not comparable. The EP flux is assessed from a common grid for all data sets (*Martineau et al.*, 2018c; *Martineau*, 2017). In ERA-Interim, there are positive EP flux divergence anomalies from September to November and negative EP flux divergence anomalies from December to February in the lower stratosphere, which imply a strengthening of the polar vortex in spring followed by a delayed breakup of the vortex in summer. This is consistent with the circulation changes shown in **Figure 6.20**. In the stratosphere, the anomalies of EP flux divergence are dominated by planetary waves. In the troposphere, both planetary and synoptic waves are affected. In late austral spring, a region of positive EP flux descends from the upper troposphere down to the surface, which is dominated by planetary waves in the upper troposphere and synoptic waves in the lower troposphere. These wave forcing anomalies are consistent with the downward descent of strengthened circumpolar winds, shown in **Figure 6.20**. The other three reanalyses show a broadly similar pattern in the stratosphere, particularly JRA-55, although the negative EP flux divergence trend in summer is typically strongest in ERA-Interim (by around -0.2 $m s^{-1} d^{-1} dec^{-1}$). Considerable differences are detected when compared to MERRA and particularly CFSR, which take the form of alternating positive and negative horizontally-orientated bands in total (planetary and synoptic) wave contributions. The disagreement is most profound in the troposphere, with differences reaching ± 0.8 $m s^{-1} d^{-1} dec^{-1}$. In all four reanalyses the region of negative EP flux divergence descends into the upper troposphere during summer, but is less pronounced in ERA-Interim largely due to differences in the synoptic wave component.

We also found that the corresponding trend in the vertical component of EP flux (**Fig. 6.22**) is characterized by reduced planetary wave propagation from the troposphere into the stratosphere in austral spring and enhanced planetary wave propagation in austral summer. All four reanalyses demonstrate similar broad features. Nevertheless, it is apparent that ERA-Interim and JRA-55 show stronger and longer lasting upward wave propagation in austral spring compared to MERRA and CFSR. In the troposphere, the intensification of winds during summer is associated with anomalies of both vertical and horizontal (not shown) synoptic EP flux divergence anomalies. The results for CFSR, in particular, show considerable differences when compared with the other three reanalyses. The disagreement again takes the form of alternating positive and negative horizontally-orientated bands.

These banded features most likely originate from the stability parameter in the vertical component of the EP flux, which is affected by the banded structure of zonally-averaged temperature trend anomalies (**Fig. 6.20**). This may be due in part to model drift induced by radiative heating imbalance during data assimilation, rather than observational errors (*e.g.*, *Lu et al.*, 2015). Similar banded structures are observed in temperature anomalies (**Fig. 1** of *Long et al.*, 2017, see also *Chapter 3*) and may result from discontinuities in the assimilation of temperatures retrieved from satellite sensors, which are known to show vertical oscillations when compared among sensors.

The four modern reanalyses support the notion that ozone depletion leads to a strengthening of the stratospheric polar vortex and consequent downward movement of zonal mean anomalies. They broadly agree on characterising the dynamical evolution of circulation anomalies and associated wave forcing in high southern latitudes during the period of formation of the ozone hole (*Thompson et al.*, 2011; *Son et al.*, 2010, 2018). The wave driving characteristics associated with the circulation changes are in general agreement with the hypothesis examined by *Orr et al.* (2012). Noticeably large differences in EP fluxes and divergence are found in CFSR compared to the other three reanalyses datasets, which appear to be related to the aforementioned vertically alternating positive and negative anomalies in temperature.

6.9 Outlook, key findings, and recommendations

We have assessed the reanalyses' representation of large-scale coupling between the troposphere and the stratospheric polar vortices, which are present during the extended winter season (or, polar night) of each hemisphere. This coupling is chiefly effected through major Sudden Stratospheric Warming (SSW) events, which are found almost exclusively in the Northern Hemisphere. Much of our focus has thus been on the boreal extratropical atmosphere on synoptic to intraseasonal time scales (*Section 6.4*). The influence of the tropics on the Northern Hemisphere polar vortex, however, is felt through modulation of SSWs by the tropical ocean (ENSO) and stratosphere (QBO) on lower frequencies (*Section 6.7*). Large-scale coupling on synoptic to seasonal timescales in both hemispheres was assessed by comparing the annular mode indices and final warming events in *Sections 6.5* and *6.6*, respectively. Finally, anthropogenic induced ozone loss caused significant trends in the polar vortex over Antarctica, as assessed in *Section 6.8*. After summarizing the results of this chapter in this section, we conclude with a list of key findings and recommendations.

Our assessment has largely focused on the self-consistency of a given reanalysis, and the consistency between the different reanalyses, as opposed to a direct validation against measurements. The large-scale circulation cannot be easily assessed from measurements directly. Surface based observations (*e.g.*, radiosondes) generally provide a very localized (point) measurement, while satellite irradiance measurements provide indirect information about composition and temperature [2]. These measurements can of course be directly linked to the large-scale circulation, but the best way of doing so is through a reanalysis, which allows one to interpolate between localized measurements and incorporate retrieval information to infer temperature, and hence the balanced circulation.

Figures 6.23 and **6.24** provide an overview of reanalyses performance for the satellite (1979-) and pre-satellite (1958 - 78) periods respectively, based on metrics discussed in *Sections 6.4* to *6.8*. We have used the 4 point scale used by all chapters in this report. In some cases, we struggled to find entirely objective measures to provide these scores, and therefore urge the reader to consult the relevant sections of the report for a more careful analysis. Demonstrated suitable indicates that a reanalysis provides a self-consistent representation of the large-scale circulation that is very similar to other reanalyses at the same level. For the very large-scale structures (*e.g.*, planetary wave structure preceding an SSW), nearly all full-input reanalyses provide a comparable representation. As detailed in previous sections, on finer scales, and particularly at higher elevations, the more recent reanalyses become more clearly superior.

[2] The new European Space Agency Aeolus mission, launched in 2018 is an exception, designed to provide direct wind measurements.

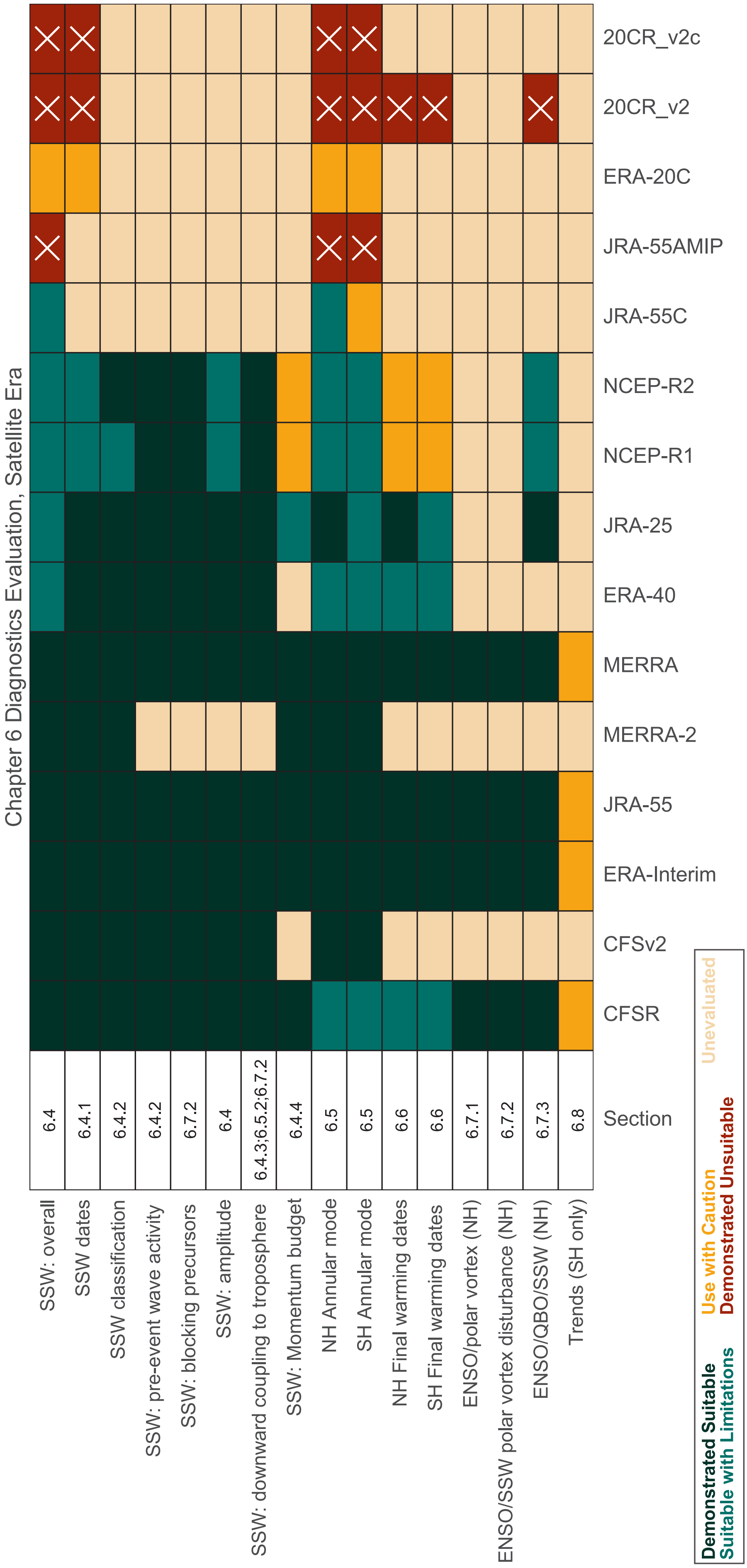


Figure 6.23: *Metric based evaluation of the reanalyses during the satellite era, 1979 onward. Please see text for further details.*

Suitable with Limitations indicates that a reanalysis provides a fairly consistent representation of circulation; conclusions from previous studies that used these reanalyses would not differ significantly if redone with a Demonstrated Suitable reanalysis. However, absent a compelling reason, these reanalyses should not be used for further research. Use with Caution has generally been applied only to the surface-input reanalyses, and the older NCEP products, which exhibit clear inconsistencies, particularly near their upper boundary at 10 hPa. Surface-input reanalyses are severely handicapped when it comes to the representation of the stratosphere, but in some cases could be used to explore variability on longer time scales. We also generally recommend to use caution when evaluating trends since reanalysis data is affected by artificial jumps caused by discontinuities in assimilated observations (*Long et al.*, 2017; *Lu et al.*, 2015; see also *Chapter 3*). Finally, as the name would imply, Demonstrated Unsuitable indicates the presence of clear problems in a reanalysis product. In particular, all of the full-input reanalyses show clear sign of divergence from basic measurements in the Southern Hemisphere before 1979. This is not to say, however, that there is no useful information in them. We also found significant biases in the mean state and variability of the polar vortex in the 20CR surface-input reanalysis, such that we do not recommend it for the purpose of investigating stratosphere-troposphere coupling.

We find that nearly all measures of large-scale coupling between the extratropical stratosphere and the troposphere are dominated by sampling uncertainty, as opposed to uncertainty in the reanalyses. As a result, conclusions based on any full (or conventional-input) reanalysis during the satellite era are generally valid. To put this more precisely, differences between the reanalyses are always smaller than the sampling uncertainty. One would not obtain results that are significantly different if you picked one reanalysis over another. The dominance of sampling uncertainty implies that our characterization of stratosphere-troposphere coupling is limited by the length of record; in a sense, we have a “small data” problem.

In the Northern Hemisphere, there is evidence that conventional observations are sufficient to constrain reanalyses from at least 1958 onward, as indicated in **Figure 6.24**. Given the dominance of sampling uncertainty, the longer record available in the boreal hemisphere is important. An additional two decades of high-quality reanalysis, as provided by JRA-55, reduces uncertainty in stratosphere-coupling processes by about 20 %. This reduction in uncertainty dwarfs the differences between the modern reanalysis over the satellite period, and makes a case for using JRA-55. We are excited that ERA5 will provide a reanalysis of the atmosphere from 1950, and it is a high priority for future work to more fully assess and compare this reanalysis.

The dominance of sampling uncertainty has implications for event based diagnostics, notably SSWs. Results based on different reanalyses may appear to diverge from one another more substantively if one does not compare the same events, i.e., use the same dates. This divergence, however, is really sampling uncertainty, aliasing into the signal.

All this said, we find that the modern reanalyses, ERA-Interim, JRA-55, MERRA 1 and 2, and to a slightly lesser extent, CFSR/CFSv2, are demonstrably superior to earlier reanalyses, providing a more dynamically consistent representation of the circulation. Over the limited period for which it is available, ERA5 also appears to be equally high quality as well. As a matter of best practice, we would urge all users to avoid earlier reanalyses unless there is specific need for them. As a practical note, modern reanalyses are available at reduced resolution. Based largely on anecdotal evidence, this appears to be a common reason why NCEP R1 is still used widely: it ’s volume of data is smaller, and thus simply easier to download. Reduced resolution is appropriate for many analyses of the large-scale circulation, but it is recommended to use a modern reanalysis with a reduced resolution instead of NCEP R1. An exception is when real-time data availability is required but we note that by mid 2020, ERA5 will be provided five days behind real time.

The surface-input reanalyses are generally inferior in their representation of stratospheric variability, but may still provide research value. We do not find evidence that NOAA-20CR reanalyses accurately capture stratospheric variability; they are therefore not recommended for use. There is evidence that ERA-20C has accurate climatological variability in the stratosphere, and substantial skill in recent decades of capturing the actual variability. It is not recommended for use if restricted to periods where other reanalyses are available, but could be valuable for analysis of stratosphere- troposphere coupling on longer time scales. It should, however, be viewed as a mixture of a high quality free running model and a reanalysis, as stratospheric variability is only partially constrained by observations.

To conclude, we provide an overall, albeit more subjective, assessment of the reanalyses in **Table 6.3**. Full-input reanalyses, which make use of all available observations at a given time, have been marked recommended, consistent, or inconsistent. Recommended does not necessarily mean error-free, but indicates a self-consistent representation of the coupled variability, and consistency with other recommended reanalyses and observational constraints where available. We have marked other reanalyses consistent when differences between them and the recommended reanalyses are small relative to sampling uncertainty. Hence published results based on these reanalyses would not be significantly different if they were redone with a recommended reanalysis. A mark of inconsistent indicates that the reanalysis differs substantially with respect to other reanalysis data sets and/or available observational constraints. While “inconsistent” is meant to convey a

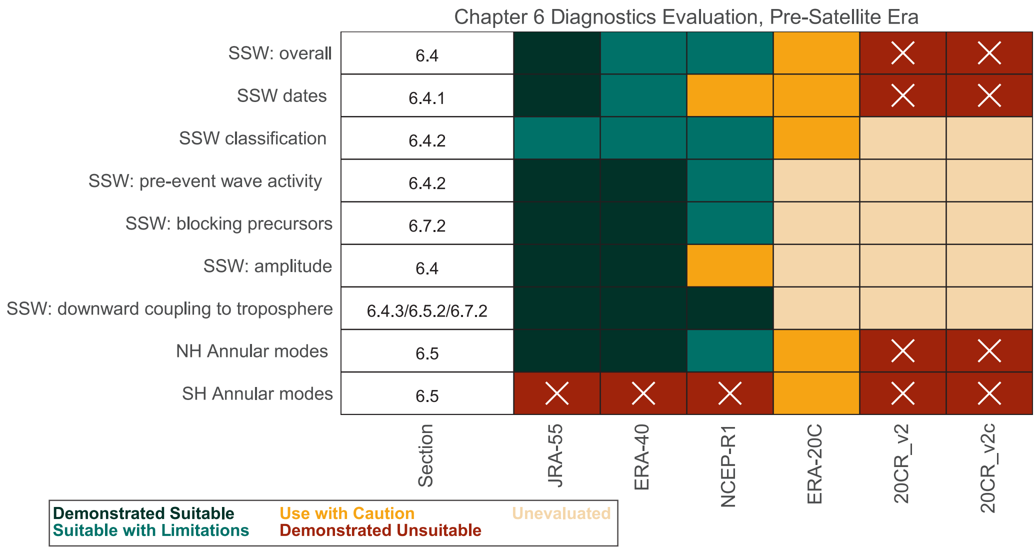


Figure 6.24: *Metric based evaluation of the reanalyses during the pre-satellite era from 1958-1978.*

clear warning, it does not imply that there is no useful information in these reanalysis products.

Given the dominance of sampling uncertainty, we may be able to glean additional confidence in stratosphere-troposphere coupling by careful use of earlier records and limited input reanalyses (*Hitchcock*, 2019). Use w/ caution has been applied to alternative reanalyses (JRA-55C and ERA- 20C), the latter of which can be used to explore variability on longer time scales. 20CR may be suitable for analysis of the troposphere, but exhibits clear biases in the variability of the stratosphere. ERA-20C, while clearly not as accurate as modern, full-input reanalysis, does appear capable of capturing information

Table 6.3: *Recommendations on the use of atmospheric reanalyses to evaluate the large-scale coupling between the stratospheric polar vortex and the tropospheric circulation on synoptic to interannual time scales. This endorsement does not include the analysis of trends, where greater caution must be employed, as discussed in Section 6.8.*

Name	post-satellite era, 1979 - present		pe-satellite era, 1958 - 979	
	NH	SH	NH	SH
ERA-40	consistent	consistent	consistent *	inconsistent
ERA-Interim†	recommended	recommended	n.a.	n.a.
ERA-20C	use w/ caution	use w/ caution	use w/ caution	use w/ caution
JRA-25	consistent	consistent	n.a.	n.a.
JRA-55	recommended	recommended	recommended *	inconsistent
JRA-55C	consistent *	use w/ caution	n.a.	n.a.
JRA-55AMIP	inconsistent	inconsistent	inconsistent	inconsistent
MERRA	consistent	consistent	n.a.	n.a.
MERRA-2	recommended	recommended	n.a.	n.a.
NCEP-R1	consistent *	consistent *	consistent *	inconsistent
NCEP-R2	consistent *	consistent *	n.a.	n.a.
CFSR	recommended	recommended	n.a.	n.a.
CFSv2	recommended	recommended	n.a.	n.a.
20CR v2	inconsistent	inconsistent	inconsistent	inconsistent
20CR v2c	inconsistent	inconsistent	inconsistent	inconsistent

* There are few conventional observations above 10 hPa, and caution must be employed above this level (or the reanalysis itself does not extend past 10 hPa).

† ERA-Interim is being supplanted by the ERA5 reanalysis. Tentative analysis suggests thatERA5 is as good as ERA-Interim, if not better, but we do not have sufficient evidence to make a full recommendation. It will be particularly important to evaluate its performance in the Northern Hemisphere during the pre-satellite era.

about the variability of the stratosphere given only surface data. This feat alone establishes the remarkably tight coupling between the troposphere and stratosphere in our atmosphere.

Key findings

- In the satellite era (1979 - onward), the representation of large-scale stratosphere-troposphere circulation is very consistent across all full-input (including satellite observations) reanalyses. On synoptic scales, the more recent reanalyses (ERA-Interim, JRA-55, MERRA and MERRA 2, and to a slightly lesser extent, CFSR/CFSv2) become more clearly superior.

- Our ability to assess and understand stratosphere-troposphere coupling is primarily limited by sampling uncertainty, that is, by the comparatively large natural variability of the circulation relative to the length of the satellite record. As an example, various efforts have sought to characterize the break-down of the polar vortex during a Sudden Stratospheric Warmings (SSW) as a split or displacement event. Methodological differences among the classifications proposed in the literature, however, result in a partial agreement (for two-thirds of SSW events). In contrast, applying the same definition to different reanalyses yields nearly identical results.

- Although measures of stratosphere-troposphere coupling determined from earlier reanalyses are generally not statistically distinct from results obtained with a more recent reanalysis, the more recent products show demonstrable improvement, particularly with respect to internal consistency (*e.g.*, the momentum budget) and at higher levels (10 hPa and above).

- Reanalysis datasets broadly agree on temperture, wind, and wave forcing trends in the austral polar vortex related to ozone depletion from 1979 to 2001. In contrast, there are no discernible trends in Northern Hemisphere polar vortex variability over the same period.

- Pre-satellite era reanalyses (1958 - 1978) appear to be of good quality in the Northern Hemisphere, and therefore can be used to reduce sampling uncertainty in measures of stratosphere-troposphere coupling by approximately 20 %. We emphasize that this represents a more significant reduction in uncertainty than achieved by shifting from an earlier generation reanalysis to a more recent reanalysis.

- Pre-satellite era reanalyses of the Southern Hemisphere are generally of poor quality, and can only be used to reduce sampling uncertainty with great caution.

- A conventional-input (excluding satellite observations) reanalysis of the Northern Hemisphere (JRA-55C) matches full-input reanalyses well up to 10 hPa, supporting the validity of pre-satellite reanalysis products in this hemisphere. JRA-55C's representation of the Southern Hemisphere is not as accurate, suggesting that satellite measurements are more critical in this hemisphere due to the reduced density of conventional observations.

- Surface-input reanalyses have also been evaluated. ERA-20C captures not only the correct statistical climatology of the Northern Hemisphere stratospheric polar vortex, but also much of its actual variability (correctly representing the timing of about half of observed SSWs). This suggests it may be suitable for exploring low-frequency variability of the stratosphere-troposphere coupled system. The representation of the stratospheric vortex in NOAA 20CR v2/v2c, however, is demonstrably poor.

Recommendations

- We recommend the use of more recent reanalysis products. As a matter of best practice, we urge all users to avoid the use of earlier reanalyses unless the project requires the use of an older product, and special care is taken to justify that the older product is otherwise consistent with more recent reanalyses. In particular, we note for users that modern reanalyses can be obtained, in addition to their native high-resolution grids, at a coarser resolution that is comparable to that of earlier reanalyses and thus more manageable in size, but which still captures the best representation of the large-scale circulation.

- The consistency of trends associated with the Antarctic ozone hole (for the period 1979 forward) suggest that reanalyses may be reliably capturing the influence of stratospheric ozone loss. One must exercise great caution in the interpretation of trends in the reanalyses, however, as they can be spuriously caused by changes in the observations assimilated over time, an issue that could systematically affect all products. Additional support from direct observations and/or understanding of the mechanism(s) help build confidence in trends found in the reanalyses.

- When an extended record is needed to reduce sampling uncertainty, we recommend the use of pre-satellite era reanalyses (1958 - 1978) in the Northern Hemisphere, but caution against their use in the Southern Hemisphere.

- Due to significant biases in the mean state and variability of the polar vortex in the 20CR surface-input reanalysis, we do not recommend it for the purpose of investigating stratosphere-troposphere coupling.

- ERA-20C may be suitable, with caution, for exploring the low-frequency variability of the stratosphere-troposphere coupled system.

- As our ability to quantify the large-scale coupling between the stratosphere and troposphere is primarily limited by sampling uncertainty, we recommend that future reanalysis products extend their analysis prior to the satellite era.

Code availability

Code can be provided by the authors upon request.

Data availability

The S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels (Martineau et al., 2018c; Martineau, 2017) is publicly available. More refined data can be provided by the authors upon request.

Acknowledgements

We thank Ilari Lehtonen for providing the classification of the split/displacement events shown in **Table A6.7**. EPG acknowledges support from the US NSF through grants AGS-1546585 and AGS-1852727. PM was a Japan Society for the Promotion of Science (JSPS) international research fellow. BA was funded by “Ayudas para la contratación de personal postdoctoral de formación en docencia e investigación en los departamentos de la UCM”. BA, DB, NC, FMP and MI were supported by the European Project 603557-STRATOCLIM under program FP7-ENV.2013.6.1-2 and the Spanish Government through the PALEOSTRAT (CGL2015-69699-R) and JEDiS (RTI2018- 096402-B-I00) projects. The work of SCH was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. The work by MT was supported by JSPS KAKENHI Grant Number JP24224011.

References

Andrews, D.G., J.R. Holton, and C.B. Leovy, 1987: Middle Atmosphere Dynamics. *Academic Press*, San Diego, xi + 489 pp., doi: 10.1002/qj.49711548612.

Ayarzagüena, B. and E. Serrano, 2009: Monthly Characterization of the Tropospheric Circulation over the Euro-Atlantic Area in Relation with the Timing of Stratospheric Final Warmings. *J. Climate*, **22**, 6313 - 6324, doi: 10.1175/2009JCLI2913.1.

Ayarzagüena, B., *et al.*, 2019: On the representation of major stratospheric warmings in reanalyses. *Atmos. Chem. Phys.*, **19**, 9469 - 9484, doi: 0.5194/acp-19-9469-2019.

Badin, G. and D.I. Domeisen, 2014: A Search for chaotic behavior in Northern Hemisphere stratospheric variability. *J. Atmos. Sci.*, **71**, 1494 - 1507, doi: 10.1175/JAS-D-13-0225.1.

Baldwin, M.P. and T.J. Dunkerton, 2001: Stratospheric Harbingers of Anomalous Weather Regimes. *Science*, **294**, 581 - 584, 2001, doi: 10.1126/science.1063315.

Baldwin, M.P. and D.W.J. Thompson, 2009: A critical comparison of stratosphere-troposphere coupling indices. *Quart. J. Roy. Meteor. Soc.*, **135**, 1661 - 1672, doi 10.1002/qj.479.

Baldwin, M.P., *et al.*, 2001: The Quasi-Biennial Oscillation. *Rev. Geophys*., **39**, 179 - 229, doi: 10.1029/1999RG000073.

Baldwin, M.P., *et al.*, 2003: Stratospheric Memory and Skill of Extended-Range Weather Forecasts. *Science*, **301**, 636 - 640, doi 10.1126/science.1087143.

Bancalá, S., K. Krüger, and M. Giorgetta, 2012: The preconditioning of major sudden stratospheric warmings. *J. Geophys. Res. Atm.*, **117**, D04101, doi: 10.1029/2011JD016769.

Barriopedro, D. and N. Calvo, 2014: On the Relationship between ENSO, Stratospheric Sudden Warmings, and Blocking. *J. Climate*, **27**, 4704 - 4720, doi: 10.1175/JCLI-D-13-00770.1.

Black, R.X., 2002: Stratopsheric Forcing of Surface Climate in the Arctic Oscillation. *J. Clim.*, **15**, 268 - 277, doi: 10.1175/1520-0442(2002)015<0268:SFOSCI>2.0.CO;2.

Black, R.X. and B.A. McDaniel, 2007: The Dynamics of Northern Hemisphere Stratospheric Final Warming Events. *J. Atmos. Sci.*, **64**, 2932 - 2946, doi: 10.1175/JAS3981.1.

Brönnimann, S., 2007: Impact of El Niño-Southern Oscillation on European climate. *Rev. Geophys.*, **45**, RG3003, doi: 10.1029/2006RG000199.

Butler, A.H. and L.M. Polvani, 2011: El Niño, La Niña, and stratospheric sudden warmings: A reevaluation in light of the observational record. *Geophys. Res. Lett.*, **38**, 1 - 5, doi: 10.1029/2011GL048084.

Butler, A.H., *et al.*, 2015: Defining Sudden Stratospheric Warmings. *Bull. Am. Meteorol. Soc.*, **96**, 1913 - 1928, doi: 10.1175/BAMS- D-13-00173.1.

Butler, A.H., J.P. Sjoberg, D.J. Seidel, and K.H. Rosenlof, 2017: A sudden stratospheric warming compendium. *Earth Syst. Sci. Data*, **9**, 63 - 76, doi: 10.5194/essd-9-63-2017.

Butler, A.H., *et al.*, 2019: Predictability of Northern Hemisphere Final Stratospheric Warmings and Their Surface Impacts. *Geophys. Res. Lett.*, **46**, 10 578 - 10 588, doi: 10.1029/2019GL083346.

Byrne, N.J. and T.G. Shepherd, 2018: Seasonal persistence of circulation anomalies in the Southern Hemisphere stratosphere, and its implications for the troposphere. *J. Clim.*, 3467 - 3483, doi: 10.1175/JCLI-D-17-0557.1.

Cagnazzo, C. and E. Manzini, 2009: Impact of the stratosphere on the winter tropospheric teleconnections between ENSO and the North Atlantic and European Region. *J. Clim.*, **22**, 1223 - 1238, doi: 10.1175/2008JCLI2549.1.

Cagnazzo, C., *et al.*, 2009: Northern winter stratospheric temperature and ozone responses to ENSO inferred from an ensemble of Chemistry Climate Models. *Atmos. Chem. Phys.*, **9**, 8935 - 8948, doi: 10.5194/acp-9-8935-2009.

Calvo, N., M.A. Giorgetta, R. Garcia-Herrera, and E. Manzini, 2009: Nonlinearity of the combined warm ENSO and QBO effects on the Northern Hemisphere polar vortex in MAECHAM5 simulations. *J. Geophys. Res.*, **114**, D13109, doi: 10.1029/2008JD011445.

Calvo, N., *et al.*, 2017: Northern Hemisphere Stratospheric Pathway of different El Niño flavors in stratosphere-resolving CMIP5 models. *J. Clim.*, **30**, 4351 - 4371, doi: 10.1175/JCLI-D-16-0132.1.

Castanheira, J.M. and D. Barriopedro, 2010: Dynamical connection between tropospheric blockings and stratospheric polar vortex. *Geophys. Res. Lett.*, **37**, 1 - 5, doi: 10.1029/2010GL043819.

Charlton, A.J. and L.M. Polvani, 2007: A New Look at Stratospheric Sudden Warmings. Part I: Climatology and Modeling Benchmarks. *J. Climate*, **20**, 449 - 469, doi: 10.1175/JCLI3996.1.

Charney, J.G. and P.G. Drazin, 1961: Propagation of Planetary-Scale Disturbances from the Lower into the Upper Atmosphere. *J. Geophys. Res.*, **66**, 83 - 109, doi: 10.1029/JZ066i001p00083.

Cohen, J. and J. Jones, 2011: Tropospheric Precursors and Stratospheric Warmings. *J. Clim.*, **24**, 6562 - 6572, doi: 10.1175/2011JCLI4160.1.

Compo, G.P., *et al.*, 2011: The Twentieth Century Reanalysis Project. *Q. J. R. Meteorol. Soc.*, **137**, 1 - 28, doi: 10.1002/qj.776.

Dee, D.P., *et al.*, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, **137**, 553 - 597, doi: 10.1002/qj.828.

Domeisen, D.I., C.I. Garfinkel, and A.H. Butler, 2019a: The Teleconnection of El Niño Southern Oscillation to the Stratosphere. *Rev. Geophys.*, **57**, 5 - 47, doi: 10.1029/2018RG000596.

Domeisen, D.I.V., *et al.*, 2019b: The Role of the Stratosphere in Subseasonal to Seasonal Prediction: 2. Predictability Arising From Stratosphere-Troposphere Coupling. *J. Geophys. Res. Atmos.*, **125**, doi: 10.1029/2019JD030923.

Esler, J.G. and N.J. Matthewman, 2011: Stratospheric Sudden Warmings as Self-Tuning Resonances. Part II: Vortex Displacement Events. *J. Atmos. Sci.*, **68**, 2505 - 2523, doi: 10.1175/JAS-D-11- 08.1.

Eyring, V., *et al.*, 2006: Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *J. Geophys. Res.*, *111*, D22 308, doi: 10.1029/2006JD007327.

Fujiwara, M., *et al.*, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp- 17-1417-2017.

Garfinkel, C.I. and D.L. Hartmann, 2007: Effects of the El Niño–Southern Oscillation and the Quasi- Biennial Oscillation on polar temperatures in the stratosphere. *J. Geophys. Res.*, **112**, D19112, doi: 10.1029/2007JD008481.

Garfinkel, C.I. and D.L. Hartmann, 2010: Tropospheric Precursors of Anomalous Northern Hemisphere Stratospheric Polar Vortices. *J. Clim.*, **23**, 3282 - 3299, doi: 10.1175/2010JCLI3010.1.

Garfinkel, C.I., *et al.*, 2012a: Why might stratospheric sudden warmings occur with similar frequency in El Niño and la Niña winters? *J. Geophys. Res. Atmos.*, **117**, 1 - 11, doi: 10.1029/2012JD017777.

Garfinkel, C.I., T.A. Shaw, D.L. Hartmann, and D.W. Waugh, 2012b: Does the Holton-Tan Mechanism Explain How the Quasi-Biennial Oscillation Modulates the Arctic Polar Vortex? *J. Atmos. Sci.*, **69**, 1713 - 1733, doi: 10.1175/JAS-D-11-0209.1.

Gelaro, R., *et al.*, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.*, **30**, 5419 - 5454, doi: 10.1175/JCLI-D-16-0758.1.

Gerber, E.P. and P. Martineau, 2018: Quantifying the variability of the annular modes: Reanalysis uncertainty vs. sampling uncertainty. *Atmos. Chem. Phys.*, **18**, 17 099 - 17 117, doi: 10.5194/acp-18-17099-2018.

Gerber, E.P., *et al.*, 2010: Stratosphere-Troposphere Coupling and Annular Mode Variability in Chemistry-Climate Models. *J. Geophys. Res.*, **115**, D00M06, doi: 10.1029/2009JD013770.

Gómez-Escolar, M., S. Fueglistaler, N. Calvo, and D. Barriopedro, 2012: Changes in polar stratospheric temperature climatology in relation to stratospheric sudden warming occurrence. *Geophys. Res. Lett.*, **39**, L22802, doi: 10.1029/2012GL053632, 2012.

Gray, L.J., *et al.*, 2018: Surface impacts of the Quasi Biennial Oscillation. *Atmos. Chem. Phys.*, **18**, 8227 - 8247, doi: 10.5194/acp-18- 8227-2018.

Hardiman, S.C., *et al.*, 2011: Improved predictability of the troposphere using stratospheric final warmings. J. *Geophys. Res. Atmos.*, **116**, 1 - 11, doi: 10.1029/2011JD015914.

Haynes, P.H., *et al.*, 1991: On the "Downward Control" of Extratropical Diabatic Circulations by Eddy-Induced Mean Zonal Forces. *J. Atmos. Sci.*, **48**, 651 - 678, doi: 10.1175/1520-0469(1991)048<0651:OTCOED>2.0.CO;2.

Hitchcock, P., 2019: On the value of reanalyses prior to 1979 for dynamical studies of stratosphere–troposphere coupling. *Atmos. Chem. Phys.*, **19**, 2749 - 2764, doi: 10.5194/acp-19-2749-2019.

Holton, J.R. and H.-C. Tan, 1980: The Influence of the Equatorial Quasi-Biennial Oscillation on the Global Circulation at 50 mb. *J. Atmos. Sci.*, **37**, 2200 - 2208, doi: 10.1175/1520-0469(1980)037<2200:TIOTEQ>2.0.CO;2.

Horel, J.D. and J.M. Wallace, 1981: Planetary-Scale Atmospheric Phenomena Associated with the Southern Oscillation, *Mon. Wea. Rev.*, **109**, 813 - 829, doi: 10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2.

Huck, P.E., *et al.*, 2007: An improved measure of ozone depletion in the Antarctic stratosphere. *J. Geophys. Res.*, **112**, D11104, doi: 10.1029/2006JD007860.

Iza, M. and N. Calvo, 2015: Role of Stratospheric Sudden Warmings on the response to Central Pacific El Niño. *Geophys. Res. Lett.*, **42**, 2482 - 2489, doi: 10.1002/2014GL062935.

Iza, M., N. Calvo, and E. Manzini, 2016: The Stratospheric Pathway of La Niña. *J. Clim.*, **29**, 8899 - 8914, doi: 10.1175/JCLI-D-16-0230.1.

Kalnay, E., *et al.*, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.*, **77**, 437 - 472, doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kanamitsu, M., *et al.*, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Am. Meteorol. Soc.*, **83**, 1631 - 1644, doi: 10.1175/BAMS-83-11-1631.

Karpechko, A.Y., P. Hitchcock, D.H. Peters, and A. Schneidereit, 2017: Predictability of Downward Propagation of Major Sudden Stratospheric Warmings. *Q. J. R. Meteorol. Soc.*, **60**, doi: 10.1002/QJ.3017.

Kim, J., S.-W. Son, E.P. Gerber, and H.-S. Park, 2017: Defining Sudden Stratospheric Warming in Climate Models: Accounting for Biases in Model Climatologies. *J. Clim.*, **30**, 5529 - 5546, doi: 10.1175/JCLI-D-16-0465.1.

Kobayashi, C., *et al.*, 2014: Preliminary Results of the JRA-55C, an Atmospheric Reanalysis Assimilating Conventional Observations Only. *SOLA*, **10**, 78 - 82, doi: 10.2151/sola.2014-016.

Kobayashi, S., *et al.*, 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *J. Meteorol. Soc. Japan. Ser. II*, **93**, 5 - 48, doi: 10.2151/jmsj.2015-001, 2015.

Kushner, P.J., 2010: Annular modes of the troposphere and stratosphere. *Geophys. Monogr. Ser.*, **190**, 59 - 91, doi: 10.1029/2009GM000924.

Kushner, P.J. and L.M. Polvani, 2004: Stratosphere-Troposphere Coupling in a Relatively Simple AGCM: The Role of Eddies. *J. Clim.*, **17**, 629 - 639, doi: 10.1175/1520-0442(2004)017<0629:SCIARS>2.0.CO;2.

Labitzke, K., 1977: Interannual Variability of the Winter Stratosphere in the Northern Hemisphere. *Mon. Weather Rev.*, **105**, 762–770, doi: 10.1175/1520-0493(1977)105<0762:IVOTWS>2.0.CO;2.

Lehtonen, I. and A.Y. Karpechko, 2016: Observed and modeled tropospheric cold anomalies associated with sudden stratospheric warmings. *J. Geophys. Res. Atmos.*, **121**, 1591 - 1610, doi: 10.1002/2015JD023860.

Lim, E., H.H. Hendon, and D.W.J. Thompson, 2018: Seasonal Evolution of Stratosphere-Troposphere Coupling in the Southern Hemisphere and Implications for the Predictability of Surface Climate. *J. Geophys. Res. Atmos.*, **123**, 12,002 - 12,016, doi: 10.1029/2018JD029321.

Limpasuvan, V., D.W.J. Thompson, and D.L. Hartmann, 2004: The Life Cycle of the Northern Hemisphere Sudden Stratospheric Warmings. *J. Clim.*, **17**, 2584 - 2596, doi: 10.1175/1520-0442(2004)017<2584:TLCOTN>2.0.CO;2.

Long, C.S., *et al.*, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14593 - 14629, doi: 10.5194/acp-17-14593-2017.

Lu, H., T.J. Bracegirdle, T. Phillips, and J. Turner, 2015: A Comparative Study of Wave Forcing Derived from the ERA-40 and ERA-Interim Reanalysis Datasets. *J. Clim.*, **28**, 2291 - 2311.

Marshall, G.J., 2003: Trends in the Southern Annular Mode from Observations and Reanalyses. *J. Climate*, **16**, 4134 - 4143, doi: 10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2.

Martineau, P., 2017: S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels. *Centre for Environmental Data Analysis*, doi: 10.5285/b241a7f536a244749662360bd7839312.

Martineau, P. and S.-W. Son, 2010: Quality of reanalysis data during stratospheric vortex weakening and intensification events. *Geophys. Res. Lett.*, **37**, doi: 10.1029/2010GL045237.

Martineau, P. and S.-W. Son, 2015: Onset of Circulation Anomalies during Stratospheric Vortex Weakening Events: The Role of Planetary-Scale Waves. *J. Clim.*, **28**, 7347 - 7370, doi: 10.1175/JCLI- D-14-00478.1.

Martineau, P., S.-W. Son, and M. Taguchi, 2016: Dynamical Consistency of Reanalysis Datasets in the Extratropical Stratosphere. *J. Clim.*, **29**, 3057 - 3074, doi: 10.1175/JCLI-D-15-0469.1.

Martineau, P., G. Chen, S.-W. Son, and J. Kim, 2018a: Lower-Stratospheric Control of the Frequency of Sudden Stratospheric Warming Events. *J. Geophys. Res. Atmos.*, **123**, 3051 - 3070, doi: 10.1002/2017JD027648.

Martineau, P., S.-W. Son, M. Taguchi, and A.H. Butler, 2018b: A comparison of the momentum budget in reanalysis datasets during sudden stratospheric warming events. *Atmos. Chem. Phys.*, **18**, 7169 - 7187, doi: 10.5194/acp-18-7169-2018.

Martineau, P., J.S. Wright, N. Zhu, and M. Fujiwara, 2018c: Zonal-mean data set of global atmospheric reanalyses on pressure levels. *Earth Syst. Sci. Data*, **10**, 1925 - 1941, doi: 10.5194/essd-10-1925- 2018.

Martius, O., L.M.Polvani, and H.C. Davies, 2009: Blocking precursors to stratospheric sudden warming events. *Geophys. Res. Lett.*, **36**, doi: 10.1029/2009GL038776.

Matsuno, T., 1971: A Dynamical Model of the Stratospheric Sudden Warming. *J. Atmos. Sci.*, **28**, 1479 - 1491, doi: 10.1175/1520-0469(1971)028<1479:ADMOTS>2.0.CO;2.

Matthewman, N.J. and J.G. Esler, 2011: Stratospheric Sudden Warmings as Self-Tuning Resonances. Part I: Vortex Splitting Events. *J. Atmos. Sci.*, **68**, 2481 - 2504, doi: 10.1175/JAS-D-11-07.1.

Matthewman, N.J., J.G. Esler, A.J. Charlton-Perez, and L.M. Polvani, 2009: A New Look at Stratospheric Sudden Warmings. Part III: Polar Vortex Evolution and Vertical Structure. *J. Clim.*, **22**, 1566 - 1585, doi: 10.1175/2008JCLI2365.1, 2009.

Maycock, A.C. and P. Hitchcock, 2015: Do split and displacement sudden stratospheric warmings have different annular mode signatures? *Geophys. Res. Lett.*, **42**, 10,943 - 10,951, doi: 10.1002/2015GL066754.

McDaniel, B.A. and R.X. Black, 2005: Intraseasonal Dynamical Evolution of the Northern Annular Mode, *J. Clim.*, **18**, 3820 - 3839, doi: 10.1175/JCLI3467.1.

McInturff, R.M., 1978: Stratospheric warmings: Synoptic, dynamic and general-circulation aspects. *NASA Reference Publ.* **NASA-RP-1017**, p. 174, available at **https://ntrs.nasa.gov/archive/nasa/casi.nt rs.nasa.gov/19780010687.pdf**.

McIntyre, E., 1982: How well do we Understand the Dynamics of Stratospheric Warmings? *J. Meteorol. Soc. Japan*, **60**, 37–65, doi: 10.2151/jmsj1965.60.1_37.

McIntyre, M.E. and T.N. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature*, **305**, 593 - 600, doi: 10.1038/305593a0, 1983.

McIntyre, M. and T. Palmer, 1984: The 'surf zone' in the stratosphere. *J. Atmos. Terr. Phys.*, **46**, 825 - 849, doi: 10.1016/0021-9169(84)90063-1.

Mitchell, D.M., A.J. Charlton-Perez, and L.J. Gray, 2011: Characterizing the variability and extremes of the stratospheric polar vortices using 2D moment analysis. *J. Atmos. Sci.*, **68**, 1194 - 1213, 2011, doi: 10.1175/2010JAS3555.1.

Mitchell, D.M., *et al.*, 2013: The Influence of Stratospheric Vortex Displacements and Splits on Surface Climate. *J. Climate*, **26**, 2668 - 2682, doi: 10.1175/JCLI-D-12-00030.1.

Mitchell, D.M., *et al.*, 2015: Signatures of naturally induced variability in the atmosphere using multiple reanalysis datasets. *Q. J. R. Meteorol. Soc.*, **141**, 2011 - 2031, doi: 10.1002/qj.2492.

Nishii, K., H. Nakamura, and T. Miyasaka, 2009: Modulations in the planetary wave field induced by upward-propagating Rossby wave packets prior to stratospheric sudden warming events: A case-study. *Quart. J. Roy. Meteor. Soc.*, **135**, 39 - 52, doi: 10.1002/qj.359.

Nishii, K., H. Nakamura, and Y.J. Orsolini, 2011: Geographical Dependence Observed in Blocking High Influence on the Stratospheric Variability through Enhancement and Suppression of Upward Planetary-Wave Propagation. *J. Clim.*, **24**, 6408 - 6423, doi: 10.1175/JCLI-D-10-05021.1.

Omrani, N.E., N.S. Keenlyside, J. Bader, and E. Manzini, 2014: Stratosphere key for wintertime atmospheric response to warm Atlantic decadal conditions. *Clim. Dyn.*, **42**, 649 - 663, doi: 10.1007/s00382-013-1860-3.

Onogi, K., *et al.*, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369 - 432, doi: 10.2151/jmsj.85.369.

Orr, A., *et al.*, 2012: Possible Dynamical Mechanisms for Southern Hemisphere Climate Change due to the Ozone Hole. *J. Atmos. Sci.*, **69**, 2917 - 2932, doi: 10.1175/JAS-D-11-0210.1.

Orr, A., *et al.*, 2021: Is our dynamical understanding of the circulation changes associated with the Antarctic ozone hole sensitive to the choice of reanalysis dataset? *Atmos. Chem. Phys.*, **21**, 7451 - 7472, doi: 10.5194/acp-21-7451-2021.

Palmeiro, F.M., D. Barriopedro, D., R. García-Herrera, and N. Calvo, 2015: Comparing Sudden Stratospheric Warming Definitions in Reanalysis Data. *J. Clim.*, **28**, 6823 - 6840, doi: 10.1175/JCLI- D-15-0004.1.

Plumb, R.A., 1989: On the seasonal cycle of stratospheric planetary waves. *Pure Appl. Geophys.* **130**, 233 - 242, doi: 10.1007/BF00874457.

Plumb, R.A., 2010: Planetary Waves and the Extratropical Winter Stratosphere. In *The Stratosphere: Dynamics, Transport, and Chemistry* (eds. L.M. Polvani, A.H. Sobel and D.W. Waugh). doi: 10.1002/9781118666630.ch2.

Poli, P., *et al.*, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Clim.*, **29**, 4083 - 4097, doi: 10.1175/JCLI-D-15-0556.1.

Polvani, L.M. and D.W. Waugh, 2004: Upward Wave Activity Flux as a Precursor to Extreme Stratospheric Events and Subsequent Anomalous Surface Weather Regimes. *J. Clim.*, **17**, 3548 - 3554, doi: 10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2.

Polvani, L.M., *et al.*, 2017: Distinguishing stratospheric sudden warmings from ENSO as key drivers of wintertime climate variability over the North Atlantic and Eurasia. *J. Clim.*, **30**, 1959 - 1969 doi: 10.1175/JCLI-D-16-0277.1.

Ramaswamy, V., *et al.*, 2001: Stratospheric temperature trends: Observations and model simulations. *Rev. Geophys.*, **39**, 71 - 122, doi: 10.1029/1999RG000065.

Randel, W.J., 1988: The seasonal evolution of planetary waves in the Southern Hemisphere stratosphere and troposphere. *Q. J. R. Meteorol. Soc.*, **587**, 1385 - 1409, doi: 10.1002/qj.49711448403.

Rao, J., C.I. Garfinkel, and R. Ren, 2019: Modulation of the Northern Winter Stratospheric El Niño–Southern Oscillation Teleconnection by the PDO. *J. Clim.*, **32**, 5761 - 5783, doi: 10.1175/JCLI-D-19-0087.1.

Richter, J.H., K. Matthes, N. Calvo, and L.J. Gray, 2011: Influence of the quasi-biennial oscillation and El Niño–Southern Oscillation on the frequency of sudden stratospheric warmings. *J. Geophys. Res.*, **116**, D20111, doi: 10.1029/2011JD015757.

Rienecker, M.M., *et al.*, 2011: MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Clim.*, **24**, 3624 - 3648, doi: 10.1175/JCLI-D-11-00015.1.

Saha, S., *et al.*, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Am. Meteorol. Soc.*, **91**, 1015 - 1057, doi: 10.1175/2010BAMS3001.1.

Saha, S., *et al.*, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185 - 2208, doi: 10.1175/JCLI-D-12-00823.1.

Scherhag, R., 1952: Die Explosionsartige Stratosphärenerwärmung des Spätwinters 1951/52. *Ber. Deut. Wetterdienst* **38**, 51 - 63.

Seviour, W.J.M., D.M. Mitchell, and L.J. Gray, 2013: A practical method to identify displaced and split stratospheric polar vortex events. *Geophys. Res. Lett.*, **40**, 5268 - 5273, doi:10.1002/grl.50927.

Sigmond, M., J.F. Scinocca, V.V. Kharin, and T.G. Shepherd, 2013: Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nat. Geosci.*, **6**, 1 - 5, doi:10.1038/ngeo1698.

Smith, K.L. and P.J. Kushner, 2012: Linear interference and the initiation of extratropical stratosphere- troposphere interactions. *J. Geophys. Res.*, **117**, doi:10.1029/2012JD017587.

Son, S.-W., *et al.*, 2010: The Impact of Stratospheric Ozone on Southern Hemisphere Circulation Change: A Multimodel Assessment. *J. Geophys. Res.*, **115**, D00M07, doi:10.1029/2010JD014271.

Son, S.-W., *et al.*, 2018: Tropospheric jet response to Antarctic ozone depletion: An update with Chemistry-Climate Model Initiative (CCMI) models. *Environ. Res. Lett.*, **13**, 054024, doi:10.1088/1748-9326/aabf21.

Song, K. and S.-W. Son, 2018: Revisiting the ENSO-SSW relationship. *J. Clim.*, **31**, 2133 - 2143, doi:10.1175/JCLI-D-17-0078.1.

Song, Y. and W.A. Robinson, 2004: Dynamical Mechanisms for Stratospheric Influences on the Troposphere. *J. Atmos. Sci.*, **61**, 1711 - 1725, doi:10.1175/1520-0469(2004)061<1711:DMFSIO>2.0.CO;2.

Steiner, A.K. *et al.*, 2020: Consistency and structural uncertainty of multi-mission GPS radio occultation records. *Atmos. Meas. Tech.* **13**, 2547 - 2575, doi:10.5194/amt-13-2547-2020.

Sun, L. and W.A. Robinson, 2009: Downward influence of stratospheric final warming events in an idealized model. *Geophys. Res. Lett.*, **36**, L03819, doi:10.1029/2008GL036624.

Taguchi, M., 2015: Changes in Frequency of Major Stratospheric Sudden Warmings with El Niño/Southern Oscillation and Quasi-Biennial Oscillation. *J. Meteorol. Soc. Japan. Ser. II*, **93**, 99 - 115, doi:10.2151/jmsj.2015-007.

Taguchi, M. and D.L. Hartmann, 2006: Increased Occurrence of Stratospheric Sudden Warmings during El Niño as Simulated by WACCM, *J. Clim.*, **19**, 324 - 332, doi:10.1175/JCLI3655.1.

Thiéblemont, R., *et al.*, 2019: Drivers and Surface Signal of Interannual Variability of Boreal Stratospheric Final Warmings. *J. Geophys. Res. Atmos.*, **124**, 5400 - 5417, doi:10.1029/2018JD029852.

Thompson, D.W.J. and J.M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297 - 1300, doi:10.1029/98GL00950.

Thompson, D.W.J. and J.M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000 - 1016, doi:10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2.

Thompson, D.W.J. and S. Solomon, 2002: Interpretation of recent Southern Hemisphere Climate Change. *Science*, **296**, 895 - 899, doi:10.1126/science.1069270.

Thompson, D.W.J., et al., 2011: Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nat. Geosci.*, **4**, 741 - 749, doi:10.1038/ngeo1296.

Tripathi, O.P., A. Charlton-Perez, M. Sigmond, and F. Vitart, 2015: Enhanced long-range forecast skill in boreal winter following stratospheric strong vortex conditions. *Environ. Res. Lett.*, **10**, 104007, doi:10.1088/1748-9326/10/10/104007.

Uppala, S.M., *et al.*, 2005: The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.*, **131**, 2961 - 3012, doi:10.1256/qj.04.176.

Waugh, D.W., 1997: Elliptical diagnostics of stratospheric polar vortices. *Q. J. R. Meteorol. Soc.*, **123**, 1725 - 1748, doi:10.1002/qj.49712354213.

Waugh, D.W. and Randel, W.J., 1999: Climatology of Arctic and Antarctic Polar Vortices Using Elliptical Diagnostics. *J. Atmos. Sci.*, **56**, 1594 - 1613, doi:10.1175/1520-0469(1999)056<1594:COAAAP>2.0.CO;2.

Waugh, D.W., A.H. Sobel, and L.M. Polvani, 2017: What Is the Polar Vortex and How Does It Influence Weather? *Bull. Am. Meteorol. Soc.*, **98**, 37 - 44, doi:10.1175/BAMS-D-15-00212.1.

Weinberger, I., C.I. Garfinkel, I.P. White, and L.D. Oman, 2019: The salience of nonlinearities in the boreal winter response to ENSO: Arctic stratosphere and Europe. *Clim. Dyn.*, **53**, 4591 - 4610, doi:10.1007/s00382-019-04805-1.

White, I., *et al.*, 2019: The downward influence of sudden stratospheric warmings: Association with tropospheric precursors. *J. Clim.*, **32**, 85 - 108, doi:10.1175/JCLI-D-18-0053.1.

Woo, S.-H., M.-K. Sung, S.-W. Son, and J.-S. Kug, 2015: Connection between weak stratospheric vortex events and the Pacific Decadal Oscillation. *Clim. Dyn.*, **45**, 3481 - 3492, doi: 10.1007/s00382- 015-2551-z.

Woollings, T., A. Hannachi, B. Hoskins, and A. Turner, 2010: A regime view of the North Atlantic oscillation and its response to anthropogenic forcing. *J. Clim.*, **23**, 1291 - 1307, doi:10.1175/2009JCLI3087.1.

Woollings, T., *et al.*, 2018: Blocking and its Response to Climate Change. *Curr. Clim. Chang. Reports*, **4**, 287 - 300, doi:10.1007/s40641-018-0108-z.

Yoden, S., M. Taguchi, and Y. Naito, 2002: Numerical Studies on Time Variations of the Troposphere- Stratosphere Coupled System. *J. Meteorol. Soc. Japan*, **80**, 811 - 830, doi:10.2151/jmsj.80.811.

Appendix A: Detection and classification of major SSW events

The onset dates of SSW events identified independently for each reanalysis data sets are listed in **Table A6.4**. Then, for the common dates whose identification is described in *Section 6.2*, events are classified as to whether they are splits or displacements according a method adapted from *Seviour et al.* (2013) (**Table A6.5**), the Shibata method (**Table A6.6**) and the method of *Lehtonen and Karpechko* (2016) (**Table A6.7**). These methods are described in more detail in *Section 6.4.2*.

Table A6.4: *Identification of major SSW events in reanalyses.The criterion for the detection is a reversal of zonal-mean zonal wind at 60° N and 10 hPa (see Section 6.4.1 for more details). Cases where the reanalysis deviates from the "common" events are highlighted in bold. Events that do not show a positive meridional temperature gradient at the same level within 5 days of the zonal wind reversal are highlighted in green.*

common	NCEP R1	CFSR	ERA-40	ERA-Interim	JRA-25	JRA-55	MERRA	MERRA-2	NCEP-R2
30-Jan-58	30-Jan-58		**31-Jan-58**			30-Jan-58			
—	**30-Nov-58**		—			—			
17-Jan-60	**16-Jan-60**		17-Jan-60			17-Jan-60			
29-Jan-63	****		**28-Jan-63**			**30-Jan-63**			
—	**23-Mar-65**		—			—			
17-Dec-65	**8-Dec-65**		**16-Dec-65**			**18-Dec-65**			
23-Feb-66	**24-Feb-66**		23-Feb-66			23-Feb-66			
7-Jan-68	****		7-Jan-68			7-Jan-68			
28-Nov-68	**27-Nov-68**		28-Nov-68			**29-Nov-68**			
13-Mar-69	13-Mar-69		13-Mar-69			****			
2-Jan-70	2-Jan-70		2-Jan-70			2-Jan-70			
18-Jan-71	17-Jan-71		18-Jan-71			18-Jan-71			
20-Mar-71	20-Mar-71		20-Mar-71			20-Mar-71			
31-Jan-73	**2-Feb-73**		31-Jan-73			31-Jan-73			
9-Jan-77	****		9-Jan-77			9-Jan-77			
22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79		22-Feb-79
29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80
—	—	—	—	—	**6-Feb-81**	**6-Feb-81**	—	—	—
4-Mar-81	****	**3-Mar-81**	4-Mar-81	4-Mar-81	4-Mar-81	4-Mar-81	4-Mar-81	****	****
4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81
24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84
1-Jan-85	**2-Jan-85**	1-Jan-85	1-Jan-85	1-Jan-85	1-Jan-85	1-Jan-85	1-Jan-85	1-Jan-85	**31-Dec-84**
23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87
8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87
14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88
21-Feb-89	**22-Feb-89**	21-Feb-89	21-Feb-89	21-Feb-89	21-Feb-89	21-Feb-89	21-Feb-89	21-Feb-89	**22-Feb-89**
—	—	—	—	—	—	—	—	—	**5-Feb-95**
15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98
26-Feb-99	**25-Feb-99**	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99
20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00
11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	**12-Feb-01**
31-Dec-01	**2-Jan-02**	**30-Dec-01**	31-Dec-01	**30-Dec-01**	31-Dec-01	31-Dec-01	**30-Dec-01**	**30-Dec-01**	**1-Jan-02**
—	—	**17-Feb-02**	**18-Feb-02**	—	—	—	—	**17-Feb-02**	—
18-Jan-03	18-Jan-03	18-Jan-03		18-Jan-03	18-Jan-03	18-Jan-03	18-Jan-03	18-Jan-03	18-Jan-03
5-Jan-04	**7-Jan-04**	5-Jan-04		5-Jan-04	**6-Jan-04**	5-Jan-04	5-Jan-04	5-Jan-04	**6-Jan-04**
21-Jan-06	21-Jan-06	21-Jan-06		21-Jan-06	21-Jan-06	21-Jan-06	21-Jan-06	21-Jan-06	21-Jan-06
24-Feb-07	24-Feb-07	24-Feb-07		24-Feb-07	24-Feb-07	24-Feb-07	24-Feb-07	24-Feb-07	24-Feb-07
22-Feb-08	22-Feb-08	22-Feb-08		22-Feb-08	22-Feb-08	22-Feb-08	22-Feb-08	22-Feb-08	22-Feb-08
24-Jan-09	24-Jan-09	24-Jan-09		24-Jan-09	24-Jan-09	24-Jan-09	24-Jan-09	24-Jan-09	24-Jan-09
9-Feb-10	9-Feb-10	9-Feb-10		9-Feb-10	9-Feb-10	9-Feb-10	9-Feb-10	9-Feb-10	9-Feb-10
24-Mar-10	24-Mar-10	24-Mar-10		24-Mar-10	24-Mar-10	24-Mar-10	24-Mar-10	24-Mar-10	24-Mar-10

Table A6.5: *Classification of SSW events into splits and displacements adapted from the method described in Seviour et al. (2013). D and S denote displacement and split events, respectively. U denotes unclassifiable events. Bold text highlights disagreement from the "common" classification. Asterisks indicate that there was substantial disagreement on the classification of the 15-Dec-98, 20-Mar-00, and 09-Feb-10 events.*

Shared Dates	common	NCEP R1	CFSR	ERA-40	ERA-Interim	JRA-25	JRA-55	MERRA	MERRA-2	NCEP-R2
30-Jan-58	D	D		D			D			
17-Jan-60	S	S		**D**			S			
29-Jan-63	S	S		S			S			
17-Dec-65	D	D		D			D			
23-Feb-66	D	D		D			**U**			
07-Jan-68	S	S		S			S			
28-Nov-68	D	D		D			D			
13-Mar-69	S	**U**		S			S			
02-Jan-70	S	S		S			S			
18-Jan-71	S	S		S			S			
20-Mar-71	D	D		D			D			
31-Jan-73	S	S		S			S			
09-Jan-77	S	S		S			S			
22-Feb-79	S	S	S	S	S	S	S	S		S
29-Feb-80	D	D	D	D	D	D	D	D	D	D
04-Mar-81	D	D	D	D	D	D	D	D	D	D
04-Dec-81	U	U	U	U	U	U	U	U	U	U
24-Feb-84	D	D	D	D	D	D	D	D	D	D
01-Jan-85	S	S	S	S	S	S	S	S	S	S
23-Jan-87	D	D	D	D	D	D	D	D	D	D
08-Dec-87	S	S	S	S	S	S	S	S	S	S
14-Mar-88	S	S	S	S	S	S	S	S	S	S
21-Feb-89	S	**D**	S	S	S	S	S	S	S	S
15-Dec-98	D*	D	**U**	D	D	**U**	D	D	**U**	D
26-Feb-99	S	S	S	S	S	S	S	S	S	S
20-Mar-00	U*	U	**D**	U	**D**	U	**D**	U	U	U
11-Feb-01	S	S	S	S	S	S	S	S	S	S
31-Dec-01	S	S	S	S	S	S	S	S	S	S
18-Jan-03	S	S	S		S	S	S	S	S	S
05-Jan-04	D	D	D		D	D	D	D	D	D
21-Jan-06	D	D	D		D	D	D	D	D	D
24-Feb-07	D	D	D		D	D	D	D	D	D
22-Feb-08	D	D	D		D	D	D	D	D	D
24-Jan-09	S	S	S		S	S	S	S	S	S
09-Feb-10	U*	**S**	U		**D**	**D**	**S**	U	**D**	**S**
24-Mar-10	D	D	D		D	D	D	D	D	D

Table A6.6: *Classification of major SSW events into splits and displacements using the Shibata technique (Ayarzaguena et al., 2019). S and D denote split and displacement events, respectively.*

Shared Dates	common	NCEP R1	CFSR	ERA-40	ERA-Interim	JRA-25	JRA-55	MERRA	MERRA-2	NCEP-R2
30-Jan-58	S	S		S			S			
17-Jan-60	D	D		D			**S**			
29-Jan-63	S	S		S			S			
17-Dec-65	D	**S**		D			D			
23-Feb-66	D	D		D			D			
07-Jan-68	S	S		S			S			
28-Nov-68	D	D		D			D			
13-Mar-69	D	D		D			D			
02-Jan-70	D	D		D			D			
18-Jan-71	S	S		S			S			
20-Mar-71	D	D		D			D			
31-Jan-73	S	S		S			S			
09-Jan-77	D	D		D			D			
22-Feb-79	S	S	S	S	S	S	S	S		S
29-Feb-80	D	D	D	D	D	D	D	D	D	D
04-Mar-81	D	D	D	D	D	D	D	D	D	D
04-Dec-81	D	D	D	D	D	D	D	D	D	**S**
24-Feb-84	D	D	D	D	D	D	D	D	D	D
01-Jan-85	S	S	S	**D**	S	**D**	S	S	S	S
23-Jan-87	D	D	D	D	D	D	D	D	D	D
08-Dec-87	S	**D**	S	S	**D**	S	S	S	S	S
14-Mar-88	S	S	S	S	S	S	S	S	S	S
21-Feb-89	S	S	S	S	S	S	S	S	S	S
15-Dec-98	S	S	S	S	S	S	S	S	S	S
26-Feb-99	S	S	S	S	S	S	S	S	S	S
20-Mar-00	D	D	D	D	D	D	D	D	**S**	D
11-Feb-01	D	D	D	D	D	D	D	D	**S**	D
31-Dec-01	D	D	D	D	D	D	D	D	D	D
18-Jan-03	S	S	S		S	S	S	S	S	**D**
05-Jan-04	D	D	D		D	D	D	D	D	D
21-Jan-06	D	D	D		D	D	D	D	D	D
24-Feb-07	D	D	D		D	D	D	D	D	D
22-Feb-08	D	D	D		D	D	D	D	D	D
24-Jan-09	S	S	S		S	S	S	S	S	S
09-Feb-10	S	S	S		S	S	S	S	S	S
24-Mar-10	D	D	D		D	D	D	D	D	D

Table A6.7: *Classification of major SSW events into splits and displacements using the Lehtonen and Karpechko (2016) method. S and D denote split and displacement events, respectively.*

Shared Dates	common	NCEP R1	CFSR	ERA-40	ERA-Interim	JRA-25	JRA-55	MERRA	MERRA-2	NCEP-R2
30-Jan-58	S	S		S			S			
17-Jan-60	D	D		D			**S**			
29-Jan-63	D	D		**S**			D			
17-Dec-65	D	D		D			D			
23-Feb-66	S	S		S			S			
7-Jan-68	S	S		S			S			
28-Nov-68	D	D		D			D			
13-Mar-69	D	D		D			D			
2-Jan-70	D	D		D			**S**			
18-Jan-71	S	S		S			S			
20-Mar-71	D	D		D			D			
31-Jan-73	S	S		S			S			
9-Jan-77	S	S		S			S			
22-Feb-79	S	S	S	S	S	S	S	S		S
29-Feb-80	D	D	D	D	D	D	D	D	D	D
4-Mar-81	D	D	D	D	D	D	D	D	D	D
4-Dec-81	D	D	D	D	D	D	D	D	D	D
24-Feb-84	D	D	D	D	D	D	D	D	D	D
1-Jan-85	S	S	S	S	S	S	S	S	S	S
23-Jan-87	D	D	D	D	D	D	D	D	D	D
8-Dec-87	S	S	S	S	S	S	S	S	S	S
14-Mar-88	S	S	S	S	S	S	S	S	S	S
21-Feb-89	S	S	S	S	S	S	S	S	S	S
15-Dec-98	D	D	D	D	D	D	D	D	D	D
26-Feb-99	S	S	S	S	S	S	S	S	S	S
20-Mar-00	D	D	D	D	D	D	D	D	D	D
11-Feb-01	S	S	**D**	S	**D**	S	S	S	S	S
31-Dec-01	D	D	D	D	D	D	D	D	D	D
18-Jan-03	S	S	S	S	S	S	S	S	S	**D**
5-Jan-04	D	D	D		D	D	D	D	D	D
21-Jan-06	D	D	D		D	D	D	D	D	D
24-Feb-07	D	D	D		D	D	D	D	D	D
22-Feb-08	D	D	D		D	D	D	D	D	D
24-Jan-09	S	S	S		S	S	S	S	S	S
9-Feb-10	S	S	S		S	S	S	S	S	S
24-Mar-10	D	D	D		D	D	D	D	D	D

Major abbreviations and terms

20CR v2/v2c	20th Century Reanalysis of NOAA and CIRES
AMIP	Atmospheric Model Intercomparison Project
CESM	Community Earth System Model
CFSR	Climate Forecast System Reanalysis of the NCEP
CFSv2	Climate Forecast System version 2
DOE	Department of Energy
ECHAM	ECMWF-HAMburg model
EMAC	ECHAM/MESSy Atmospheric Chemistry model
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño-Southern Oscillation
EP (Flux)	Eliassen-Palm Flux
ERA-20C	ECMWF 20th century reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA-Interim	ECMWF interim reanalysis
HF	Heat flux
JRA-55	Japanese 55-year Reanalysis
JRA-55C	Japanese 55-year Reanalysis assimilating Conventional observations only
JSPS	Japan Society for the Promotion of Science
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2
MESSy	Modular Earth Submodel System
MSLP	mean sea-level pressure
NAM	Northern Annular Mode
NAO	North Atlantic Oscillation
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NCEP-CPC	National Centers for Environmental Prediction, Climate Prediction Cente
NCEP R2	Reanalysis 2 of the NCEP and DOE
NCEP R1	Reanalysis 1 of the NCEP and NCAR
NDJF	*November-December-January-February*
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
PNJ	polar night jet
QBO	Quasi-biennial Oscillation
REM	Reanalysis ensemble mean
SAM	Southern Annular Mode
SFW	Stratospheric Final Warming
SH	Southern Hemisphere
SPARC	Stratosphere-troposphere Processes And their Role in Climate
SSW	Sudden Stratospheric Warming
S-RIP	SPARC Reanalysis Intercomparison Project
WMO	World Meteorological Organization
W1	Wavenumber 1
W2	Wavenumber 2

Chapter 7: Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS)

Chapter lead authors

Cameron R. Homeyer	School of Meteorology, University of Oklahoma	USA
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA

Co-authors

Luis F. Millán	Jet Propulsion Laboratory, California Institute of Technology	USA
Alexander C. Boothe	(1) School of Meteorology, University of Oklahoma (2) Meteorologist, National Weather Service	USA
Tao Xian	(1) School of Meteorology, University of Oklahoma (2) School of Earth and Space Sciences, University of Science and Technology of China	USA China
Mark A. Olsen	NOAA/Oceanic and Atmospheric Research/Weather Program Office *previously at: Goddard Earth Sciences Technology and Research, Morgan State University*	USA
Michael J. Schwartz	Jet Propulsion Laboratory, California Institute of Technology	USA
Michaela I. Hegglin	Department of Meteorology, University of Reading	UK
Zachary D. Lawrence	(1) Cooperative Institute for Research in Environmental Sciences (CIRES), Univ. of Colorado (2) NOAA Physical Sciences Laboratory (PSL) (3) NorthWest Research Associates	USA
Krzysztof Wargan	(1) Science Systems and Applications Inc. (2) Global Modeling and Assimilation Office, NASA Goddard Space Flight Center	USA

Abstract. The ExUTLS is an important region for understanding the impacts of and feedbacks to anthropogenically forced climate change. Modern reanalyses provide output at vertical resolution that facilitates detailed examination of the ExUTLS and the myriad dynamical, chemical, and physical processes that occur in this layer. This chapter compares diagnostics of many ExUTLS processes in modern reanalysis datasets. The diagnostics include characterization of the tropopause based on different definitions (including multiple tropopauses, vertical structure, comparison of temperature-gradient based tropopause characteristics with radiosonde observations, *etc.*); UTLS jet characteristics and long-term changes; atmospheric transport from trajectory model calculations; and diagnostics of mixing and stratosphere-troposphere exchange (STE). In addition, assimilated UTLS ozone from recent reanalyses is evaluated and compared with satellite observations. Overall results highlight the importance of using high-resolution (particularly in the vertical) reanalyses on their native grids to capture many ExUTLS processes, including tropopause structure and evolution. Most of the diagnostics evaluated show the MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 reanalyses to be suitable for UTLS studies with some limitations; in particular, CFSR/CFSv2 does not agree well with the other reanalyses for several of the diagnostics. While useful information on trends in the tropopause and jet characteristics was obtained, great caution is urged in conducting trend studies from reanalyses, and agreement among several reanalyses is one of the key elements for assessing robustness of the ExUTLS trends shown here.

Contents

7.1 Introduction 267
7.2 Reanalyses and general approach 267
7.3 The extratropical tropopause 268
7.3.1 Lapse rate tropopause altitudes 269
7.3.2 Dynamical tropopause altitudes 270
7.3.3 Multiple tropopauses 273
7.4 Jet streams 277
7.4.1 Jet characterization for UTLS studies 277
7.4.2 Climatology of UTLS jets in reanalyses 279
7.4.3 Trends in UTLS jets in reanalyses 280
7.5 Transport and mixing 282
7.5.1 Stratosphere-troposphere exchange 282
7.5.2 Mixing and transport barriers 284
7.5.3 Mass flux across 380 K isentropic surface 287
7.6 UTLS ozone 290
7.6.1 Ozone in tropopause-relative coordinates 291
7.6.2 Ozone in equivalent latitude coordinates 293
7.6.3 Ozone in jet-relative coordinates 294
7.6.4 Ozone mini-holes 298
7.7 Summary and recommendations 299
References 303
Major abbreviations and terms 307

7.1 Introduction

The ExUTLS is a layer of large dynamical, chemical, and physical variability in the atmosphere. It is the transition between the often turbulent, well-mixed troposphere and the relatively quiescent and stratified stratosphere (*e.g.*, *Gettelman et al.*, 2011). Discontinuities in vertical temperature gradients, trace gas concentrations, and occurrence of clouds exist in this layer and are generally centered near the extratropical tropopause. Dynamical processes that lead to mixing between the upper troposphere (UT) and lower stratosphere (LS) can lead to changes in the chemical characteristics of this layer and, ultimately, its radiative forcing. Namely, several key trace gases that are also greenhouse gases have dominant sources that are confined to either the troposphere or stratosphere: H_2O is prevalent in the troposphere and Earth's radiative forcing is most sensitive to changes in its concentration in the LS (*e.g.*, *Solomon et al.*, 2010; *Forster and Shine*, 1999). O_3 is prevalent in the stratosphere and Earth's radiative forcing is most sensitive to changes in its concentration in the UT (*e.g.*, *Lacis et al.*, 1990). *Riese et al.* (2012) showed that the radiative effects of both O_3 and H_2O are sensitive to mixing processes in the UTLS.

The separation between the tropical UTLS (often referred to as the "Tropical Tropopause Layer" or TTL) and the ExUTLS is often based on the location of the subtropical jets in each hemisphere or the tropopause break (the sharp discontinuity in lapse rate tropopause altitude between tropics and extratropics); that separation is thus the region where the isentropes slope sharply downward with increasing latitude. This chapter focuses on the ExUTLS as defined in this manner (*i.e.*, poleward of the subtropical jet and tropopause break), while *Chapter 8* focuses on the TTL. Processes related to tropical width and monsoon evolution occur at the interface between these two regions, and, because of their close ties to tropical circulations, are discussed primarily in *Chapter 8*. Because the UTLS is strongly coupled to the troposphere below and the stratosphere above, the altitude region we focus on here extends from below the high-latitude tropopauses to above the tropical tropopauses, thus, very roughly, from about 300 hPa to 70 hPa.

The definition of the tropopause (outlined further below) is a necessary element of any UTLS study. Its location defines the transition from troposphere to stratosphere (and, as a result, the depth and altitude location of the UTLS region) and enables further analysis of topics such as transport, composition, dynamics, and their collective impacts on radiation and climate. For example, trace gas profiles and stratosphere-troposphere exchange (STE) are commonly evaluated in a tropopause-relative altitude framework (*e.g.*, *Pan et al.*, 2010; *Hegglin et al.*, 2006; *Hoor et al.*, 2004). Assessing the accuracy of tropopause altitudes and evaluating appropriate methods of defining the tropopause for various applications are thus important focuses of recent and ongoing research using reanalyses.

STE is often assessed in reanalyses by coupling their three-dimensional wind fields with a trajectory model or by driving chemical transport models (CTMs) with reanalysis output and including passive tracers. Many ExUTLS studies focus on individual, large-scale STE processes that are resolved in the reanalyses, such as Rossby wave breaking (RWB; *e.g.*, *Kunz et al.*, 2015; *Homeyer and Bowman*, 2013; *Song et al.*, 2011; *Sprenger et al.*, 2007; *Hitchman and Huesmann*, 2007), stratospheric intrusions or tropopause folds (*e.g.*, *Knowland et al.*, 2017; *Škerlak et al.*, 2015; *Sprenger et al.*, 2003), and extratropical cyclones (*e.g.*, *Jaeglé et al.*, 2017; *Reutter et al.*, 2015). Some studies point to long-term changes in these STE processes, which are important to consider in the context of a changing climate because of their impacts on the radiation budget through changes in the distribution of UTLS water vapor, ozone, and additional greenhouse gases (*e.g.*, *Orbe et al., 2018*; *Zeng et al.*, 2010; *Hegglin and Shepherd*, 2009). Smaller scale phenomena (*e.g.*, shearing instabilities, gravity wave breaking associated with tropopause-penetrating convection) have received increasing attention in ExUTLS research (*e.g.*, *Kunkel et al.*, 2019; *Homeyer et al.*, 2017; *Wang et al.*, 2016a), but these processes will not be evaluated in reanalyses until grid resolution meets the scales necessary to resolve such processes. However, tropopause altitudes from reanalyses (and other global models) are often used in observational studies of such phenomena.

The ExUTLS is particularly important because changes in radiatively active trace gases in the region are important drivers of climate variability and change. *Chapter 4* and *Davis et al.* (2017) did not focus much on the UTLS, but did show zonal mean ozone evaluations that indicated persistent biases in the UTLS, as well as deficiencies in their ability to capture the seasonal cycle of ozone. We include some further evaluations of UTLS ozone here, particularly in dynamical coordinate (EqL, tropopause, jet-relative) frameworks. *Chapter 4* and *Davis et al.* (2017) showed water vapour in the reanalyses evaluated here to be generally unsuitable for scientific use, so we do not further evaluate reanalysis water vapour in the UTLS.

7.2 Reanalyses and general approach

In situ observations of composition and transport events are historically sparse. As a result, reanalyses are the primary source of input for climatological studies of the ExUTLS. UTLS processes in general involve very strong dynamical and chemical gradients, and thus require fine resolution to properly resolve. Older reanalyses (such as NCEP-NCAR R1, NCEP-NCAR R2, and ERA-40) not only have inadequate resolution in both the horizontal and vertical, but also are usually not available on model levels. *Manney et al.* (2017; see *Sections 7.3.3* and *7.4.2* below) showed that even the latest generation high-resolution reanalyses do not capture tropopause and UTLS jet structure when interpolated to a standard pressure grid.

Because of the strong dependence on resolution, we use only the latest generation high-resolution reanalyses in this chapter; these comprise MERRA, MERRA-2, ERA-Interim, CFSR/CFSv2, and JRA-55. For most diagnostics (exceptions will be noted) the reanalyses are used on (or, in the case of spectral models, near) the full model horizontal resolution and on the model vertical grid. For detailed information on model grids and configurations, please see *Chapter 2* and *Fujiwara et al.* (2017); we briefly summarize the most relevant aspects here. The horizontal resolution for these reanalyses is 0.5×0.67, 0.5×0.625, 0.75×0.75, 0.5×0.5, and approximately 0.56×0.56 degrees, respectively; the JRA-55 dataset is provided on the Gaussian grid corresponding to the model horizontal resolution. The vertical resolution of these reanalyses in the ExUTLS ranges from about 0.8 km to about 1.5 km, varying with reanalysis and altitude (Figure 2.1 of *Chapter 2*; Figure 3 of *Fujiwara et al.*, 2017): CFSR/CFSv2 vertical spacing increases gradually from about 0.6 km at 7 km altitude to about 0.95 km at 20 km altitude; ERA-Interim and JRA-55 have very similar vertical grids, increasing smoothly from about 0.6 km at 7 km altitude to about 1.4 km at 20 km altitude; MERRA and MERRA-2 increase rapidly from about 0.6 km at 7 km altitude to about 1.2 km at 9 km altitude and remain nearly constant above that up to above 20 km altitude. Details of the assimilation systems and data inputs for reanalyses are given in *Fujiwara et al.* (2017) and *Chapters 2* and *4*.

Most of the comparisons shown here are done starting in 1979 or 1980, and the end dates vary from 2010 to 2015. As noted in *Chapter 3* and *Long et al.* (2017) (as well as several other chapters), some diagnostics show significant changes in reanalysis agreement over the 30 - 40-year periods studied here, generally in relation to large changes in the reanalysis data inputs. Except where specifically noted, most of the diagnostics in this chapter do not show strong sensitivity to the exact time period analyzed, or to such changes in the reanalysis inputs. A few diagnostics are computed for shorter time periods to compare with observational datasets with more limited records.

The reanalyses used have very different treatments of, and inputs for, ozone that affect the UTLS, as described in detail in *Fujiwara et al.* (2017), *Davis et al.* (2017), and *Chapter 4*. Of particular relevance to this chapter is that MERRA-2 assimilates Aura MLS profile and Ozone Monitoring Instrument (OMI) total column ozone after October 2004, and ERA-Interim assimilates these ozone data in 2008 and the near-real-time (NRT) MLS ozone data starting in mid-2009. *Davis et al.* (2017) and *Chapter 4* show results that suggest persistent biases in assimilated ozone in the UTLS, which are generally not well understood given increasing uncertainties in ozone observations in this region.

7.3 The extratropical tropopause

Tropopause altitudes are critical to ExUTLS transport studies, especially those leading to STE. Due to their global and continuous coverage, reanalyses have often been used to identify the tropopause for observational and modeling-based transport studies. Differences in assimilation and model design (*e.g.*, grid resolution) between reanalyses result in differing tropopause altitudes. Incorrect tropopause altitudes can lead to significant biases in transport estimates owing to the typically strong gradients in trace gases at the tropopause level. Thus, it is important to evaluate the accuracy of reanalysis tropopause altitudes and identify the similarities and differences between reanalyses to inform their uses in ExUTLS studies.

Many previous studies have employed unique methods to determine the altitudes of the tropopause. In the ExUTLS these methods can be summarized into 5 general types:

1. Temperature lapse rate: this approach identifies changes in vertical temperature gradients (or lapse rates) to distinguish the well-mixed troposphere from the stably stratified stratosphere. The most common lapse-rate tropopause definition is that outlined by the WMO (*World Meteorological Organization*, 1957). Issues related to its application with model output are discussed in *Homeyer et al.* (2010).

2. Potential vorticity (PV) isosurface: often referred to as the 'dynamical tropopause', this approach depends largely on atmospheric stability but enables unique tracking of air mass history since PV is quasi-conserved over time periods of several days. It is most commonly used in transport analyses. Recent efforts have also used a PV gradient approach to identify the tropopause, especially for studies that evaluate isentropic transport (*e.g.*, *Kunz et al.*, 2011b).

3. Chemical tropopause: this approach identifies changes in atmospheric composition with height. In observations, ozone is typically used as it is often uniformly low in the troposphere but increases rapidly with altitude in the lower stratosphere (*e.g.*, *Bethan et al.*, 1996). In models, an artificial tracer with sources at the lower boundary is typically used to identify the chemical transition associated with the tropopause (*e.g.*, *Prather et al.*, 2011).

4. Stability transition: this approach is similar to the temperature lapse rate in that it identifies the sharp change in static stability between troposphere and stratosphere, but it depends on the Brunt-Vaisälä frequency. Some studies identify the stability transition using curve-fitting techniques (*e.g.*, *Homeyer et al.*, 2010) and others simply search for the LS stability maximum in combination with the temperature lapse rate definition (*e.g.*, *Gettelman and Wang*, 2015).

5. Lagrangian tropopause: this approach uses a trajectory model to determine the fraction of particles in a given volume that have recently (within the prior 15 - 60 days) been located within the planetary boundary layer and is similar to the Eulerian artificial tracer approach for the chemical tropopause. The time period used for trajectory calculation is fixed for this method and is commonly 30 days (*e.g.*, *Berthet et al.*, 2007).

Regardless of the method used to identify the extratropical tropopause, its altitude commonly spans a range from ~8 km to ~13 km (350 hPa > p > 150 hPa). Temperatures of the extratropical tropopause are typically between 205 K and 225 K.

Several of the diagnostics and evaluations shown below use tropopauses calculated within the JETPAC (JEt and Tropopause Products for Analysis and Characterization) software package (*Schwartz et al.*, 2015; *Manney et al.*, 2014, 2011). In JETPAC, the dynamical tropopause is defined by PV values in the extratropics from 2.0 to 6.0 potetntial vorticity units units (PVU), joined with the 380 K PV contour in the tropics; the PV values cover the range that has been widely used (*e.g.*, *Schoeberl*, 2004; *Highwood et al.*, 2000). The primary lapse rate tropopause is defined using the WMO definition. Multiple lapse rate tropopauses are then identified if *dT/dz* drops below $-2\,\mathrm{K\,km^{-1}}$ after (that is, at a higher altitude) remaining below that threshold for at least 2 km above the primary lapse rate tropopause, and then rises above $-2\,\mathrm{K\,km^{-1}}$ again; *Randel et al.* (2007a) showed that this criterion results in a climatology of multiple tropopauses in (relatively coarse-resolution) meteorological analyses comparable to that from high-resolution measurements.

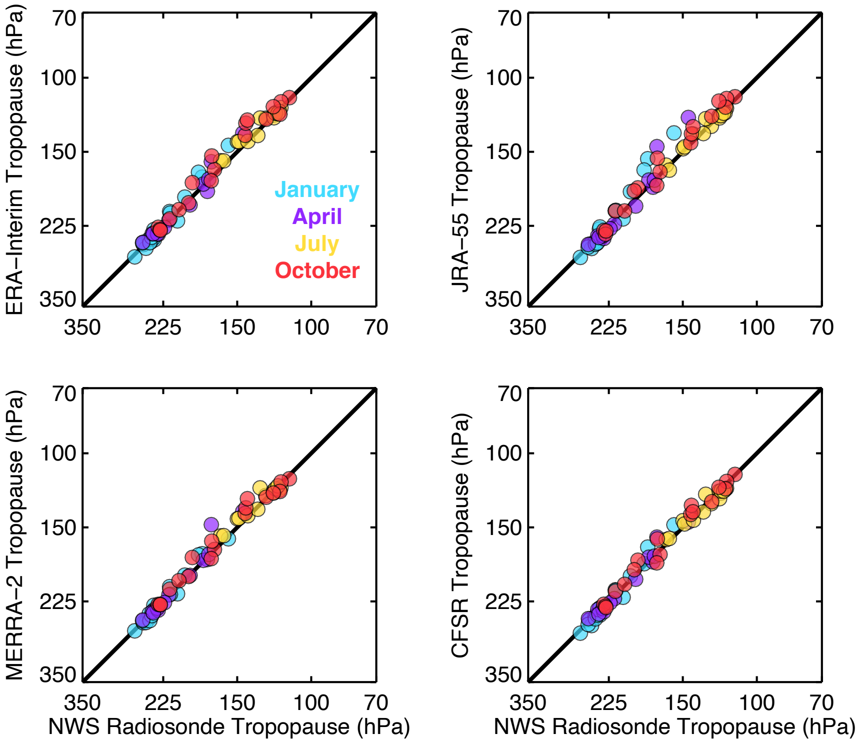


***Figure 7.1:** Scatterplots of monthly mean WMO lapse-rate tropopause pressure from reanalyses and NWS radiosondes over the CONUS. Results for ERA-Interim, JRA-55, MERRA-2, and CFSR are shown from January 2001 to October 2001 in 3-month increments. The thick black lines are 1-to-1 lines. The comparisons are only for NWS stations with a continuous record of observations throughout each month (typically ~15 out of ~50 stations).*

7.3.1 Lapse rate tropopause altitudes

The uncertainty (*i.e.*, error) of the tropopause altitude calculated using numerical model output such as that from a reanalysis (based on comparisons with radiosonde observations) is typically comparable to the vertical resolution of the model in the UTLS (*e.g.*, *Xian and Homeyer*, 2019; *Solomon et al.*, 2016; *Homeyer*, 2014; *Homeyer et al.*, 2010). For ERA-Interim, JRA-55, and CFSR/CFSv2, the expected uncertainty in the extratropical tropopause altitude is therefore ~800 m, while it is ~1000 m in MERRA-2. For more information on differences in grid resolution, see *Section 7.2*, *Chapter 2* and Figure 3 of *Fujiwara et al.* (2017). In addition to the variance among reanalysis, it is important to access the accuracy of tropopause altitudes in the reanalyses. Such an assessment is typically done by determining the bias in tropopause altitude through comparisons of the reanalysis tropopause altitudes with those computed from high-resolution radiosonde observations.

Figure 7.1 shows comparisons of monthly mean tropopause altitudes from four modern reanalyses (ERA-Interim, JRA-55, MERRA-2, and CFSR) with those computed using operational high-resolution National Weather Service (NWS) radiosondes in the Contiguous United States (CONUS). These comparisons are valid for four months from a single year, but the results are comparable to those from alternative months within all years examined (2001 - 2010; not shown). In particular, this comparison reveals that reanalysis tropopause altitudes are largely unbiased, while some unique biases can be found throughout the year. Namely, there is some evidence of a slight high bias for tropopause altitudes near ~150 hPa during northern hemisphere (NH) winter (DJF) and spring (MAM) in each reanalysis, with the largest such biases commonly found in JRA-55. *Xian and Homeyer* (2019) show similar comparisons for instantaneous tropopause identifications using a global set of long-term radiosonde observations and find that MERRA-2 tropopauses are most often biased high, while the remaining reanalyses are most often biased low. Errors in instantaneous tropopause altitudes in all reanalyses are greatest in the subtropics (*i.e.*, the transition between the ExUTLS and TTL). Globally, the average instantaneous tropopause error (root-mean-sqare, RMS, difference) is ~1 km in all modern reanalyses.

To examine differences in reanalysis tropopause altitudes at a larger scale, global comparisons of monthly mean tropopause altitudes from several combinations of the ERA-Interim, JRA-55, CFSR, and MERRA-2 reanalyses are shown in **Figure 7.2**. These comparisons correspond to the same time period as that in **Figure 7.1** and reveal that reanalyses that share a common vertical grid agree on the location of the tropopause. Most reanalysis tropopause comparisons closely follow each other in NH winter and spring, but deviate in the ~225 hPa to ~100 hPa altitude range in NH summer (JJA) and fall (SON). These differences occur near the common location of the tropopause break within the southern hemisphere (SH) (not shown). As is true for the analysis presented in **Figure 7.1**, these results are found consistently throughout the reanalysis record.

Long-term changes in the extratropical tropopause temperature and/or altitude are an indication of climate change and are relevant to the assessment of transport and other ExUTLS processes. *Xian and Homeyer* (2019) evaluate long-term changes in the WMO lapse-rate tropopause altitude during the period 1981 to 2015 and show that ERA-Interim, JRA-55, and MERRA-2 indicate similar changes in magnitude, pattern, and sign, while CFSR provides a substantially different picture. They also show that trends in ERA-Interim, JRA-55, and MERRA-2 are broadly consistent with radiosonde observations, but the extent of patterns is not yet known given the relatively poor global coverage of such observations. **Figure 7.3** shows these 35-year trends from ERA-Interim, JRA-55, MERRA-2, and CFSR in a tropopause break-relative latitude coordinate, which enables assessment of tropopause altitude changes within tropical and extratropical reservoirs separately.. Trends are mostly positive (*i.e.*, tropopause altitudes are increasing) for the three reanalyses in agreement, with the

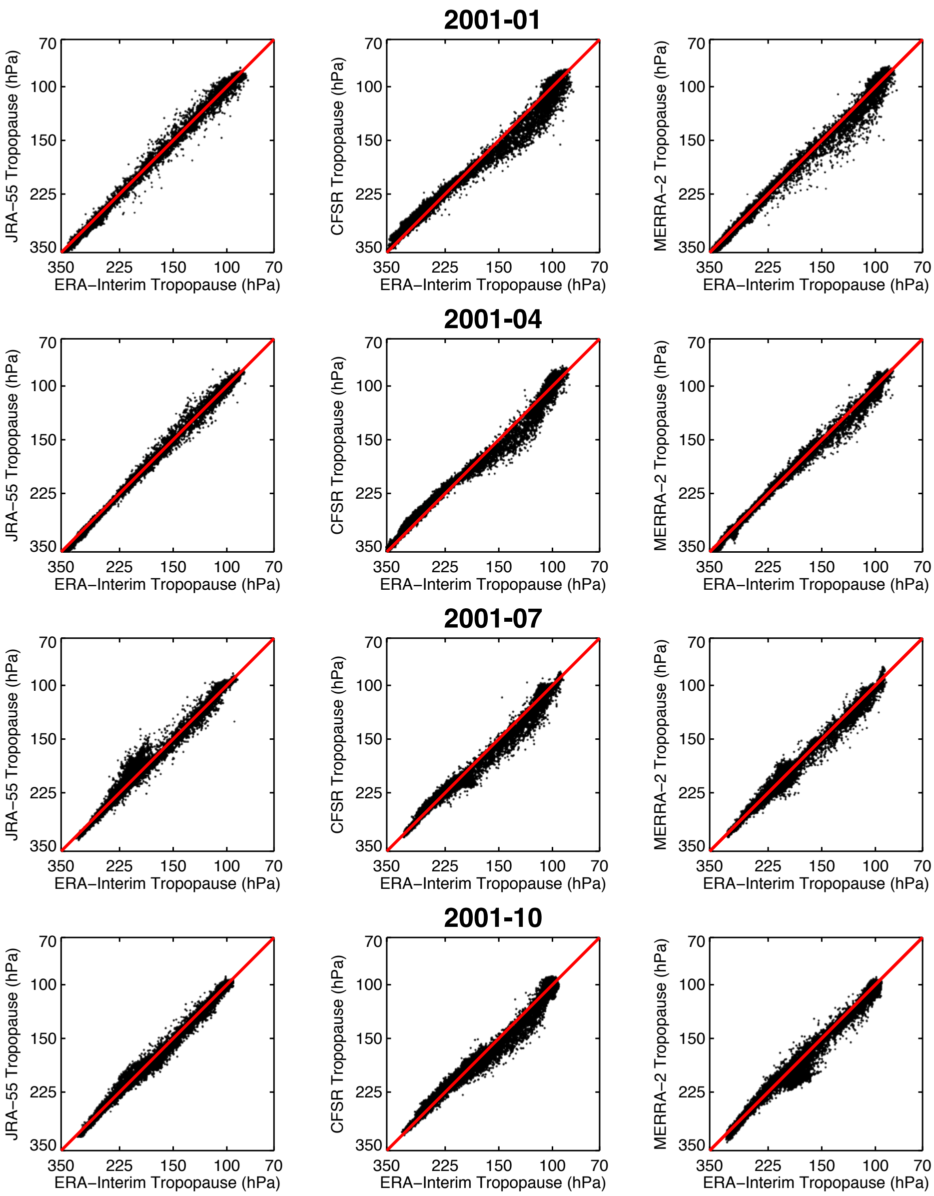


Figure 7.2: *Scatterplots comparing global monthly mean WMO lapse-rate tropopause pressures from multiple reanalyses. Comparisons are given left to right between JRA-55 and ERA-Interim, CFSR and ERA-Interim, and MERRA-2 and ERA-Interim, respectively. Results for January 2001 to October 2001 in 3-month increments are provided from top to bottom. The red lines are 1-to-1 lines.*

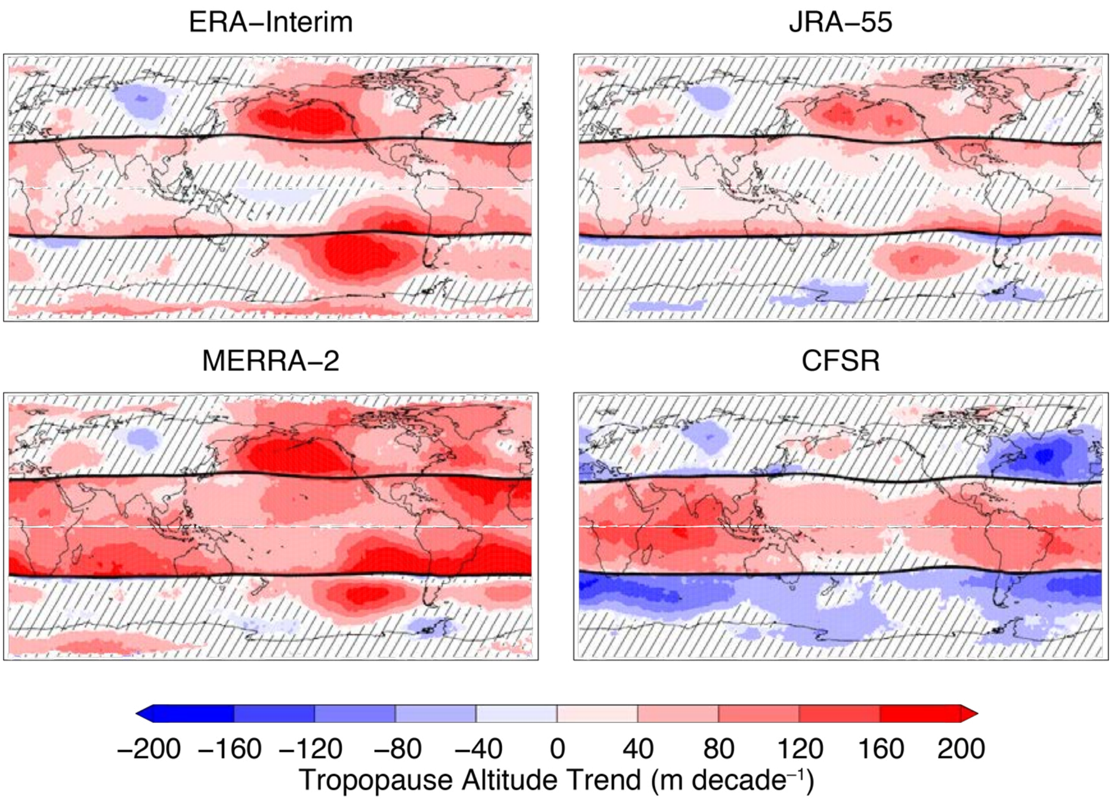


Figure 7.3: *Maps of 35-yr (1981 - 2015) trends in the WMO lapse-rate tropopause altitude in a tropopause break-relative latitude coordinate from four modern reanalyses: ERA-Interim, JRA-55, MERRA-2, and CFSR/CFSv2. Color-filled regions are statistically significant at the 99 % level. Thick black lines show the annual-mean tropopause break latitudes for each reanalysis. Figure 5 from Xian and Homeyer (2019).*

greatest changes found within the tropics and extratropical Pacific. It is not yet clear how these patterns are affected by changes in UTLS dynamics or regional variations in tropospheric heating.

7.3.2 Dynamical tropopause altitudes

The dynamical tropopause is typically defined by a three-dimensional contour of PV in the extratropics, joined with an isentropic surface, usually 380 K, in the tropics (*i.e.*, where the PV contour lies at higher potential temperature or is ill-defined). The most appropriate PV value to use depends on the focus of the study and on latitude, with higher values (*e.g.*, 4.5 PVU) typically lying near the lapse rate tropopause at higher latitudes. Furthermore, *Kunz et al.* (2011a) found that the barrier to isentropic transport was at higher PV (up to about 5.5 PVU) on higher potential temperature levels. Here we have compared the dynamical tropopauses using 1.5, 2.0, 3.5, 4.5, and 6.0 PVU in the extratropics joined with the 380 K isentropic surface in the tropics (the tropopause calculations are from JETPAC, described by *Manney et al.*, 2011). Qualitatively similar results were found for each of the tropopause values, with lower values usually showing slightly larger differences between the reanalyses. The examples shown below are for 2.0 PVU, one of the most commonly used values for a PV-based tropopause. To assess differences that could be related to the large increases in input data sources and changes from TOVS to ATOVS in the 1998 through 2002 time frame (see, *e.g.*, *Chapters 2* and *3*, and *Fujiwara et al.*, 2017; *Long et al.*, 2017), we examined climatologies for two separate 8-year periods, 1986 - 1993 and 2003 - 2010; the results showed no substantial differences in patterns of reanalysis agreement between the periods, so we show only the full comparison period here. We compare MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 for 1980 - 2014; the tropopauses are calculated using the data on the full model grid but averaged in 2-degree bins for the comparisons.

Figure 7.4 shows climatological dynamical tropopause altitude maps for DJF and JJA, showing the reanalysis ensemble mean (REM, the average of the four reanalyses used) and the differences of each of the reanalyses from the REM (reanalysis - REM, so positive values indicate that the reanalysis tropopause altitude is higher than that of the REM). In both seasons, the differences from the REM are generally less than 0.2 km over most of the globe, with some regions near 30 °N and 30 °S (near the latitude of the break in the lapse rate tropopause (LRT), where the dynamical tropopause height also drops sharply) showing larger differences (magnitudes up to about 1.5 km). The relatively large differences in the tropopause break region primarily arise from small differences in the latitude location of the sharp gradient in

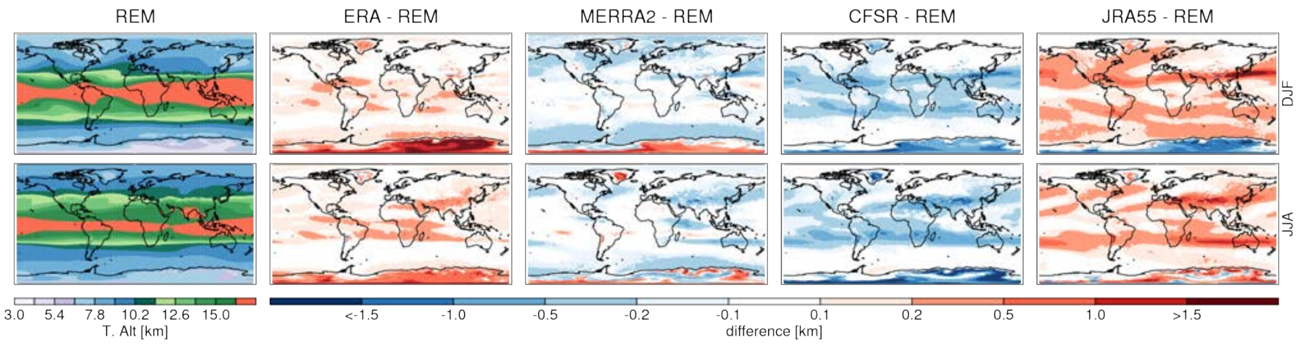


***Figure 7.4:** Climatological (1980-2014) 2 PVU tropopause altitudes for DJF (top row) and JJA (bottom), from the REM for the four reanalyses, and (second through rightmost panels) the difference of each reanalysis from the REM (reanalysis – REM) for that climatological period. Adapted from Millán et al. 2021.*

tropopause altitudes. These differences are most prominent in CFSR/CFSv2 and JRA-55 in the regions where the subtropical jets are climatologically strongest (see, *e.g.*, *Manney et al.*, 2014), *i.e.*, in the NH over Africa and Asia in DJF, and in the same longitude region in both hemispheres in JJA. In general ERA-Interim and JRA-55 have higher, and MERRA-2 and CFSR/CFSv2 lower, dynamical tropopauses than the REM, except over Antarctica in DJF, where MERRA-2 (JRA-55) is higher (lower). CFSR/CFSv2 shows very low and MERRA-2 very high dynamical tropopauses over Greenland in JJA, with values up to about 1.5 km from the REM. The exception to the good agreement among the reanalyses is over large portions of Antarctica, where ERA-Interim and CFSR/CFSv2 show positive and negative differences, respectively, from the REM in both seasons, with magnitudes of over 3 km for ERA-Interim in DJF and CFSR/CFSv2 in JJA. In DJF, MERRA-2 and JRA-55 show opposite-signed differences from the REM over Antarctica than they do over the rest of the globe, with smaller differences that vary in sign locally in JJA. Because both LRT and dynamical tropopauses are often somewhat ill-defined in the polar regions, especially during winter, and conventional data inputs (*e.g.*, high -resolution radiosonde temperature profiles that help capture the vertical structure) to the reanalyses are sparser (especially over Antarctica), larger disagreements in the Antarctic are not surprising. In general, there are no dramatic differences in the agreement among reanalyses between the early and late periods, though there is a small decrease in the range of differences from the REM in the Antarctic (not shown). The lack of difference before to after the TOVS/ATOVS transition mentioned above (not shown) indicates that the increases in data inputs that so profoundly affect reanalysis differences in Antarctic temperature values (*e.g.*, *Lawrence et al.*, 2018; *Long et al.*, 2017) do not have a strong influence on the PV values demarking the tropopause. Similarly, *Xian and Homeyer* (2019) found no apparent (or significant) trends, steps, or discontinuities in the LRT time series in the Antarctic associated with changes in data inputs.

Figure 7.5 shows frequency distributions of each of the four reanalyses versus the REM for DJF and JJA. All of the reanalyses cluster strongly around the one-to-one line with the REM, as expected given the generally good agreement seen in the climatological maps. The largest departures from the

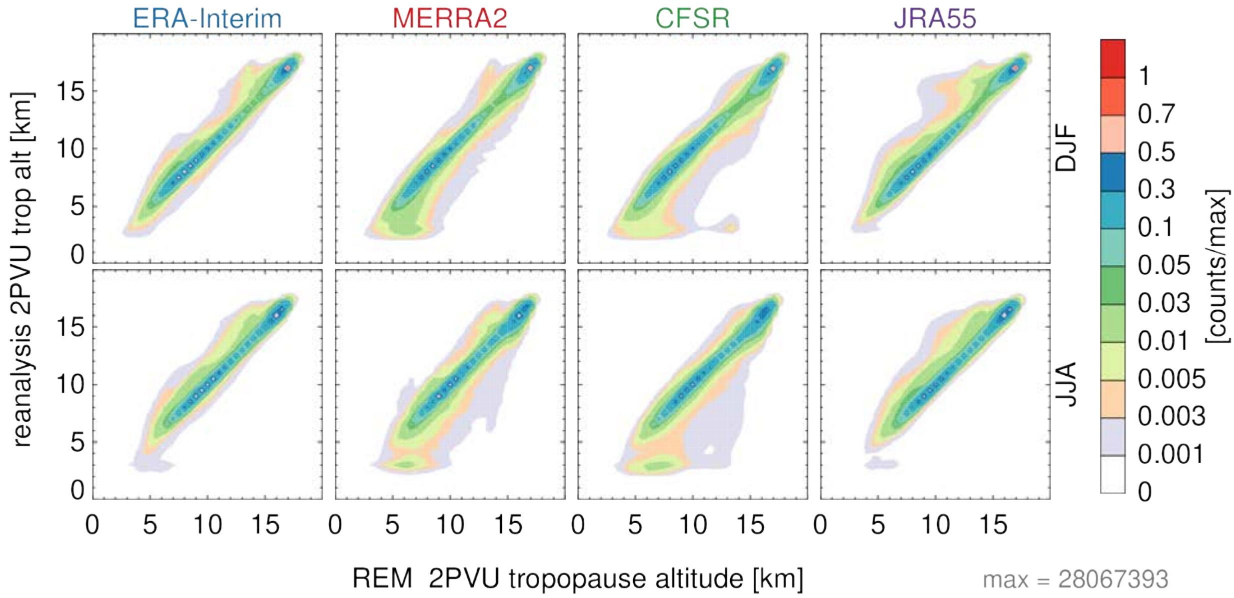


***Figure 7.5:** Density plots of REM tropopause altitude (x-axis) versus tropopause altitude for each of four reanalyses (y-axis), for DJF (top) and JJA (bottom), for 1980-2014. Adapted from Millán et al. 2021.*

REM are seen for low REM tropopause altitudes, below about 8 km, where the distributions for each of the reanalyses become quite wide, indicating considerable uncorrelated variability among the reanalyses in the lowest tropopause altitude values.

7.3.3 Multiple tropopauses

Manney et al. (2017) examined the climatology of multiple lapse rate tropopauses identified using the JETPAC tools, as described above. Regions with multiple tropopause altitudes occurring in the altitude layer between approximately 10 km and 20 km have been linked with poleward transport of tropical upper tropospheric air into the extratropical lower stratosphere (*e.g.*, *Homeyer et al.*, 2011; *Pan et al.*, 2009; *Randel et al.*, 2007b) and with poleward RWB (*e.g.*, *Homeyer and Bowman*, 2013).

Figure 7.6 shows comparisons of zonally averaged multiple tropopause frequency distributions over the seasonal cycle from multiple reanalyses. As described in detail by *Manney et al.* (2017), the frequency distributions (here and in other JETPAC-based frequency distribution plots shown later) are normalized by dividing by the total number of points that would "fill" each bin, thus in this case the number of grid points and days in each bin (latitudes × longitudes × years). We compute differences of MERRA, ERA-Interim, JRA-55, and CFSR/CFSv2 from MERRA-2, the most recent of the reanalyses used. Differences are fairly large among the reanalyses and depend strongly on vertical resolution and grid spacing in the UTLS, which vary significantly among the reanalyses (see *Chapter 2*, Figure 2.1, and *Fujiwara et al., 2017*, Figure 3). In the NH, smallest differences are seen between MERRA and MERRA-2, which share a vertical grid in addition to being different versions of the same data assimilation system/model. JRA-55 generally shows fewer multiple tropopauses than MERRA-2 and CFSR/CFSv2 generally shows more. CFSR/CFSv2 shows many more multiple tropopauses in the tropics than any of the other reanalyses. In the SH winter and spring, MERRA, ERA-Interim show substantially lower multiple tropopause frequencies than MERRA-2 in mid-latitudes and substantially higher multiple tropopause frequencies in the south polar region; the pattern is similar in JRA-55 but muted because of the overall lower frequencies.

Manney et al. (2017) showed that these differences are relatively zonally symmetric, especially in the SH. Vertical cross-sections of multiple tropopauses (*Manney et al.*, 2017) indicate that MERRA-2 shows more sharply peaked secondary tropopause altitudes, leading to a layered pattern of differences with the other reanalyses; these differences were larger in the period before the TOVS/ATOVS transition, suggesting that they are related to differences in the temperature structure related to reanalysis input changes (as shown in zonal means by, *e.g.*, *Long et al.*, 2017, also see *Chapter 3*). *Manney et al.* (2017) also evaluated multiple tropopauses in JRA-55C versus JRA-55 and found large differences (up to about 30 %) in SH middle to high latitudes, with high-latitude multiple tropopauses clustered

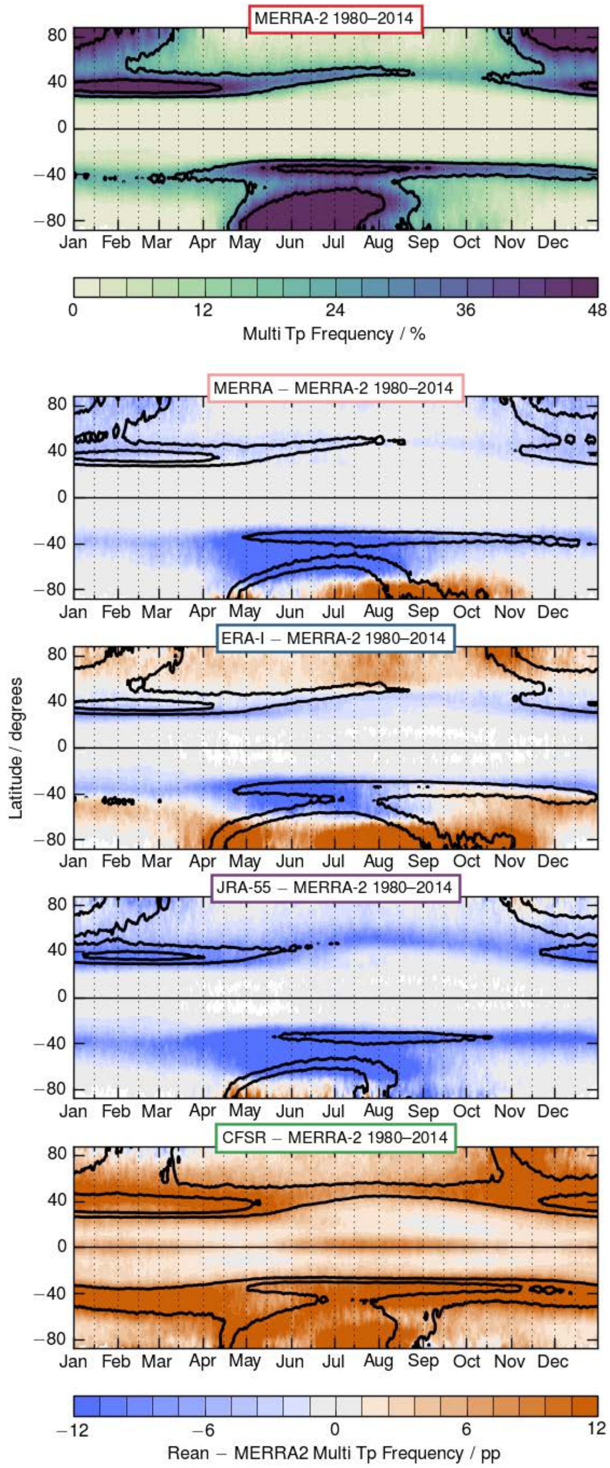


Figure 7.6: *Climatological seasonal cycle of zonally averaged frequency distributions of multiple tropopauses from MERRA-2 (top), and differences between those frequency distributions and the other reanalyses (remaining rows). Black overlays show frequency contours of 24 % and 48 % for the reanalysis in each panel. The differences are expressed in "percentage points" (see Manney et al., 2017) to indicate that they are the absolute differences of values initially expressed as a percent. Adapted from Manney et al., 2017.*

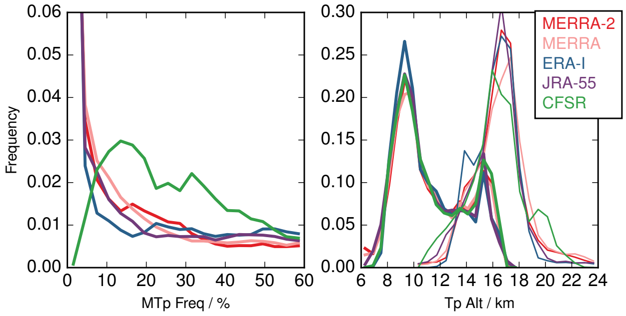


Figure 7.7: *Summary of globally / seasonally averaged frequency distributions of multiple tropopause frequencies (left) and altitudes (right; primary tropopause thick lines, secondary tropopause thin lines) in five reanalyses. Adapted from Manney et al., 2017.*

in different longitude regions. In addition, *Manney et al.* (2017) showed that CFSR/CFSv2 interpolated to pressure levels (which significantly degrades the vertical resolution) does very poorly at representing multiple tropopauses, with much lower frequency distributions (see *Section 7.4.2* and **Figure 7.11** below).

Global annual mean multiple tropopause frequency differences are summarized in **Figure 7.7.** In this broad average, the multiple tropopause frequency distributions agree fairly well among the reanalyses except for CFSR/CFSv2, which shows many fewer instances of low multiple tropopause frequencies and many more of high ones than the other reanalyses, reflecting the patterns (especially the large frequencies in low latitudes) seen in **Figure 7.6**. The primary tropopause altitudes show good agreement on average for all of the reanalyses. The same is mostly the case for secondary tropopause altitudes, except that ERA-Interim shows a secondary peak near 14 km and CFSR/CFSv2 shows a lower frequency of peaks near 16 - 18 km and a secondary peak near 20 km.

Figures 7.8 and **7.9** show 2005 - 2015 climatologies of dynamical fields from the MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 reanalyses at the measurement locations of the Aura Microwave Limb Sounder (MLS), along with MLS ozone in multiple tropopause regions as identified in JETPAC using each of the reanalyses' temperature profiles at the MLS measurement locations; the differences of these profiles from the REM of those four reanalyses are also shown. The profiles are screened and interpolated to MLS locations as described by *Schwartz et al.* (2015), wherein multiple tropopauses associated with the extreme thermal structure under the winter polar vortices are screened out. Generally similar results are seen in other seasons. In both hemispheres, there are typically larger differences among the reanalyses in multiple tropopause than in single tropopause regions. JRA-55 stands out as usually having larger differences from the REM than the other reanalyses, especially in the NH. For many of the fields, MERRA-2 and ERA-Interim are closer to the REM, with JRA-55 and CFSR/CFSv2 showing opposite extremes. The temperature differences from the REM in the UTLS range up to about 0.8 K in single tropopause regions and nearly 2 K in double tropopause regions. As can be deduced from the ozone profiles (which show the same ozone data, so the differences arise solely from identification of different profiles as having single or double tropopauses in different reanalyses), the reanalyses show significant differences in which profiles are identified as having double tropopauses, and these differences have implications for using reanalysis dynamical fields in analysis of trace gas observations. Reanalysis ozone in multiple tropopause regions is discussed below in *Section 7.6.1.*

The tropopause inversion layer (TIL), a region of enhanced static stability just above the primary LRT (*e.g.*, *Birner et al.*, 2006), is clearly seen in all of the reanalyses evaluated here (*Schwartz et al.,* 2015), also noted good representation of the TIL in MERRA and in operational GMAO analyses of that generation), in contrast to the poor representation in older reanalyses evaluated by *Birner et al.* (2006); this is in agreement with the findings of *Pilch Kedzierski et al.* (2016), who showed that ECMWF operational analyses and ERA-Interim substantially improved the TIL representation over earlier reanalyses. Indeed, several other recent studies have shown the latest generation of reanalyses to be useful for studying TIL variability and evolution (*e.g.*, *Wargan and Coy*, 2016; *Wang et al.*, 2016b; *Gettelman and Wang*, 2015). The TIL is generally strongest in CFSR/CFSv2 and weakest in JRA-55, but in each case shown still appears to be somewhat weaker than seen in high-resolution data such as GNSS-RO, as was found by *Pilch Kedzierski et al.* (2016) and as is consistent with the climate model results of *Hegglin et al.* (2010). The distance above the primary tropopause appears to be similar to that seen in GNSS-RO data (GNSS-RO analyses are shown by *Wang et al., 2016b*; *Pilch Kedzierski et al.*, 2016; *Hegglin et al.*, 2010) for the reanalyses evaluated here. Our evaluations indicate that differences in TIL representation among the reanalyses cannot be attributed solely to differences in the vertical resolution, since MERRA and MERRA-2 have somewhat coarser vertical resolution in this region than the other three reanalyses evaluated here (*Chapter 2*, *Section 7.2,* and *Fujiwara et al.*, 2017), yet do not show the weakest TIL; this is consistent with other studies (*e.g.*, as shown for climate models by *Hegglin et al.,* 2010) and suggests a dependence on differences in the data assimilation procedures, which is also supported by the studies cited above.

Xian and Homeyer (2019) evaluate long-term changes in double tropopause frequencies in radiosondes and modern reanalyses, focusing on profiles where multiple WMO tropopauses are identified at 20 km and below (to isolate multiple tropopauses indicative of STE between the tropical UT and extratropical lowe stratosphere). For these studies, the WMO definition is used, wherein additional tropopauses (*i.e.*, secondary and above) are identified using the same criteria as that for the first tropopause if the lapse rate exceeds 3 K km^{-1} in a layer at least 1 km deep above a previous identification. *Xian and Homeyer* (2019) find good agreement for global patterns of double tropopause events between the observations and reanalyses, but an under-representation of the frequency of events. **Figure 7.10** shows trends in monthly double tropopause frequency in radiosondes and reanalyses

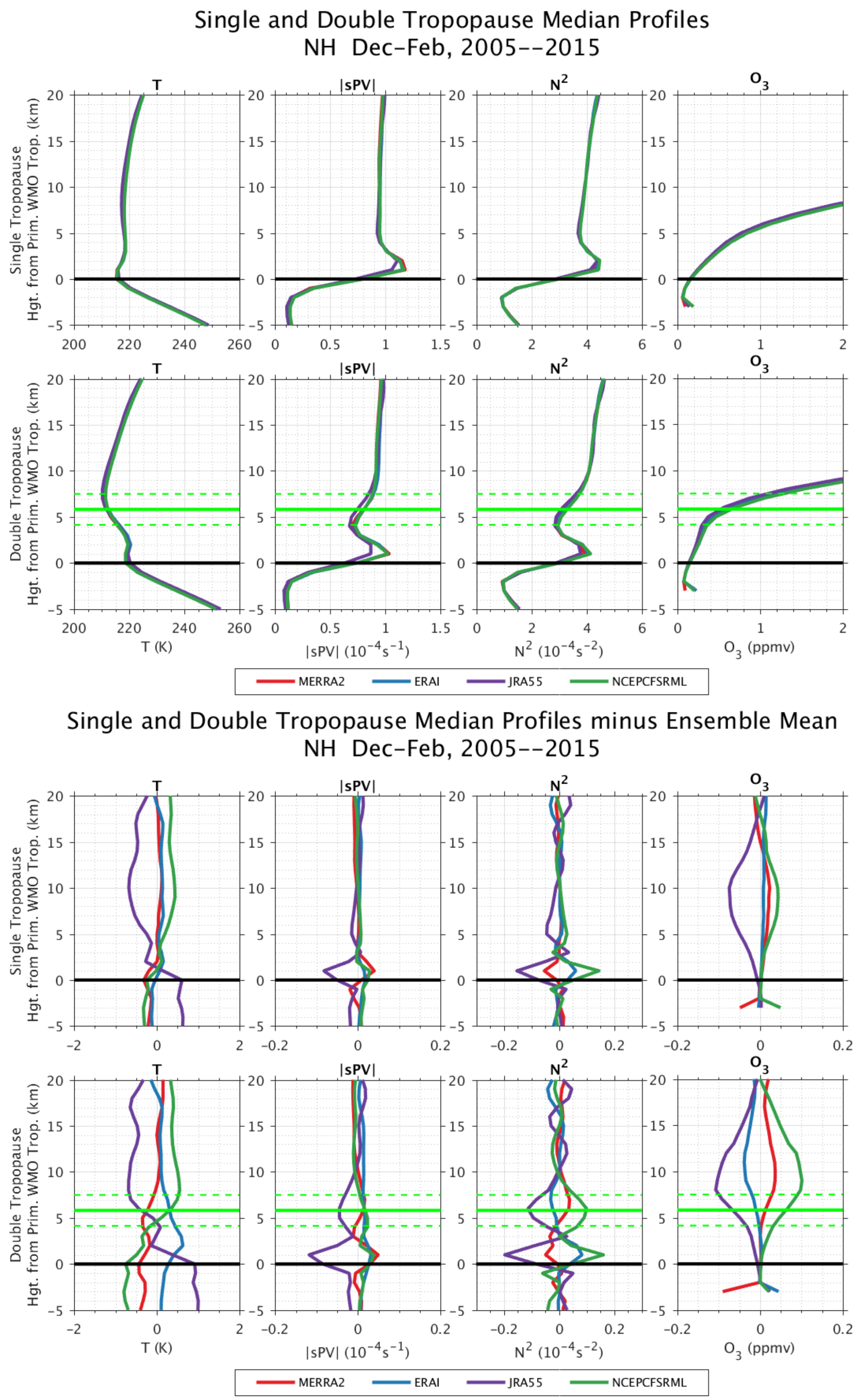


Figure 7.8: *(Top) Climatological (left to right) temperature, scaled PV, N^2, and MLS ozone distributions from MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 in regions with and without multiple lapse rate tropopauses, for the NH in DJF in 2005 through 2015. All fields are interpolated to the MLS measurement locations before averaging. (Bottom) Differences of those profiles from the REM of the four reanalyses. The ozone shown is from MLS in all cases – the differences thus arise solely from the identification of different profiles as having single and double tropopauses.*

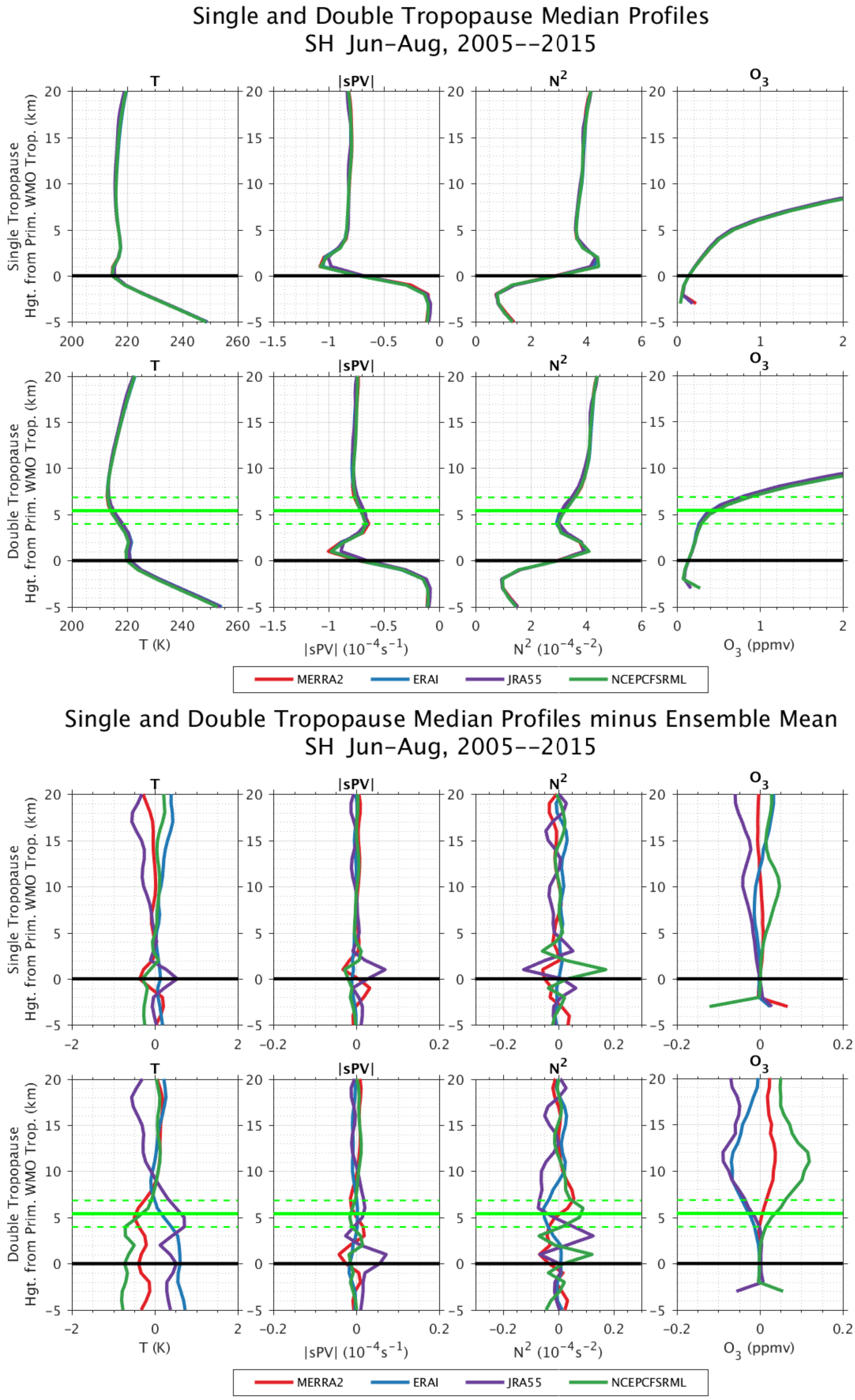


Figure 7.9: *As in* ***Figure 7.8*** *but for the SH in JJA.*

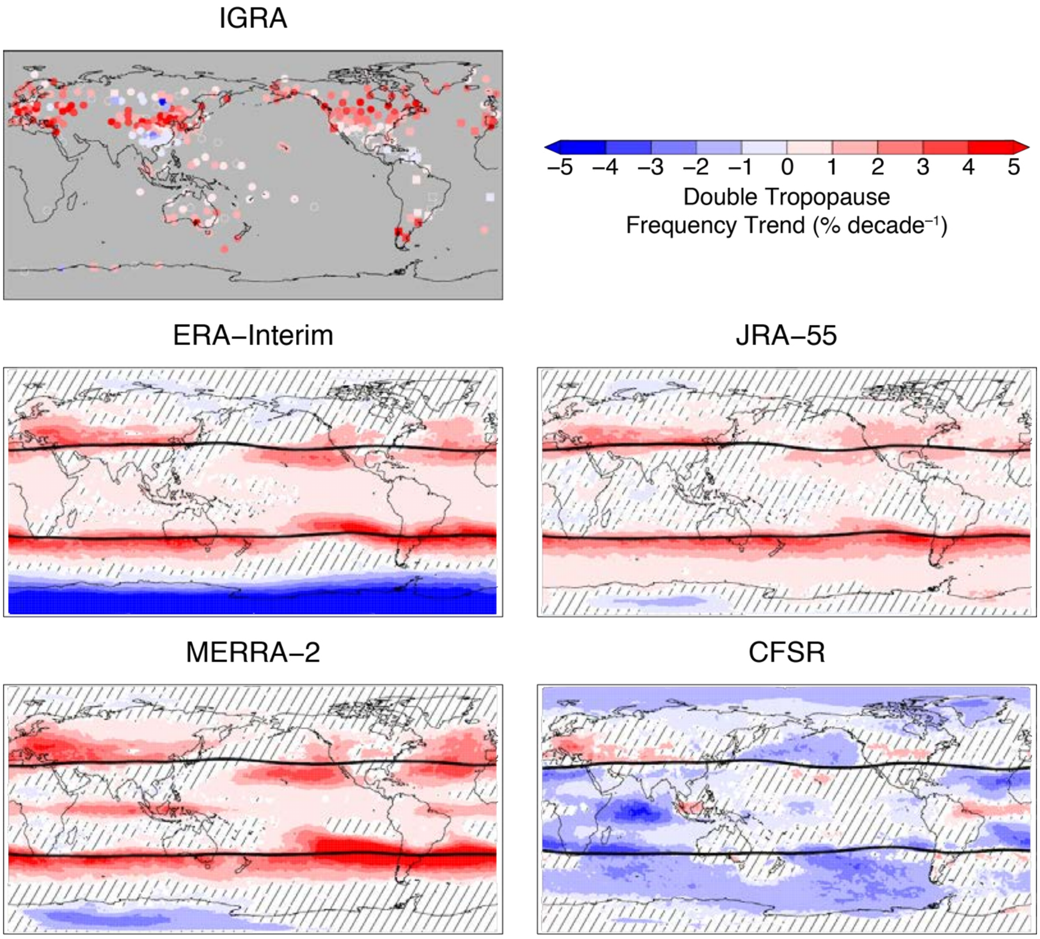


Figure 7.10: *Maps of 35-yr (1981 - 2015) trends in the WMO lapse-rate tropopause double tropopause frequency from global radiosonde observations and four modern reanalyses: ERA-Interim, JRA-55, MERRA-2, and CFSR/CFSv2. Color-filled regions are statistically significant at the 99 % level. Thick black lines show the annual-mean tropopause break latitudes for each reanalysis. From Xian and Homeyer (2019).*

over the period from 1981 to 2015 and reveals that double tropopause events are increasing globally throughout the subtropics. CFSR/CFSv2 is the only reanalysis that is broadly inconsistent with the observations and differs considerably from the remaining reanalyses. Tropopause break-relative analysis of double tropopause trends (not shown here) demonstrates that increases in double tropopause frequency are largely poleward of instantaneous tropopause break locations in the reanalyses (*i.e.*, in the extratropics). While changes in double tropopause frequency over this time period may be related to changes in the Hadley cell circulations (see *Chapter 8*), the relationship between the two has not been thoroughly investigated.

7.4 Jet streams

7.4.1 Jet characterization for UTLS studies

The UT jet streams are a key component of the atmospheric circulation and closely linked with weather and climate phenomena such as storm tracks, precipitation, and extreme events (*Mann et al.*, 2017; *Harnik et al.*, 2016; *Koch et al.*, 2006, and references therein). The UT jets and the tropopause are themselves sensitive to climate change and ozone depletion (*Waugh et al.*, 2015; *Grise et al.*, 2013; *McLandress et al.*, 2011; *WMO*, 2011; *Lorenz and DeWeaver,* 2007, and references therein), as well as to natural modes of variability such as El Nino Southern Oscillation (ENSO) and the Quasi-biennial oscillation (QBO) (*Maney et al.*, 2021; *Olsen et al.*, 2016; *Lin et al.*, 2014, 2015; *Hudson*, 2012, and references therein). The upper tropospheric jets (as well as the tropopauses) are important drivers of composition variability in the UTLS, acting as transport barriers and controlling STE and long-range transport. Assessing UTLS composition and its relationships to the dynamics of the tropopauses and UTLS jets is an important outstanding problem (*e.g.*, *Hegglin et al.*, 2016). As noted by *Manney and Hegglin* (2018a), many of the critical characteristics of jets cannot be directly observed, so reanalyses are one of our most important tools for understanding UTLS jets and their impact on composition. In the following sections we use JETPAC to characterize and compare UTLS jets in the most recent suite of reanalyses. The jet characterization from JETPAC is as described by *Manney et al.* (2011, 2014, 2017) and *Manney and Hegglin* (2018a); briefly:

A UT jet is identified wherever there is a wind speed

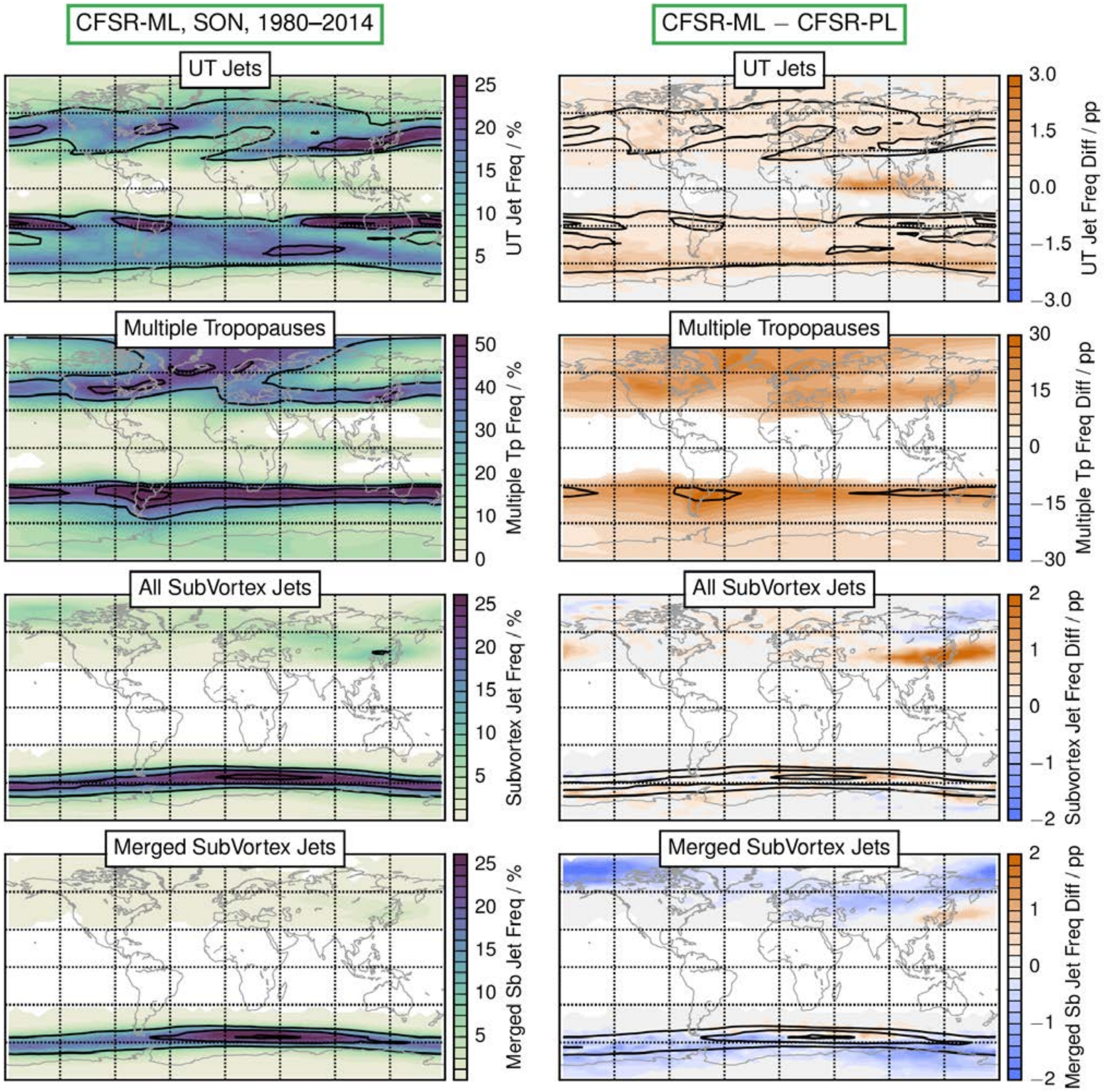


Figure 7.11: *Comparison of CFSR/CFSv2 on (left) model levels and pressure levels, showing climatological (1980-2014) SON frequency distributions of (top to bottom) upper tropospheric jets, multiple tropopauses, all subvortex jets, and merged subvortex jets. The difference between model and pressure levels is shown in the right column. Black overlays show frequency contours (10, 20, and 30 % on jet plots; 30, 45, and 60 % on multiple tropopause plots). From Manney et al., 2017.*

maximum greater than 40 m s^{-1}; the boundaries of the jet region are the points surrounding that (in both horizontal and vertical directions) where the wind speed drops below 30 m s^{-1}. When more than one maximum above 40 m s^{-1} appears within a given 30 m s^{-1} contour, they are defined as separate cores if the latitude distance between them is greater than 15 degrees or the decrease in wind speed between them is greater than 25 m s^{-1}. These parameters were tuned to approximate as closely as feasible the choices that would be made by visual inspection.

The UT jets may be further characterized as "subtropical" (STJ; thought of as primarily radiatively driven) or "polar" (PJ; also referred to as "eddy-driven") (see *Manney et al.*, 2014, and references therein, for detailed discussion of the spectrum of jet characteristics). *Manney et al.* (2011, 2014) used a simple latitude criterion (appropriate for climatological studies) to identify STJ and PJ. A more robust definition is needed for regional and case studies. We follow *Manney and Hegglin* (2018a) here, and define the STJ as the most equatorward westerly jet for which the WMO tropopause altitude at the equatorward edge of the jet is greater than 13.0 km and that tropopause altitude drops by at least 2.0 km from the equatorward to the poleward side of the jet. This definition identifies the jet across which the "tropopause break" occurs. The PJ is then defined as the strongest jet poleward of the STJ, or poleward of

40 degrees latitude if no STJ was identified.

The subvortex jet core is identified at each reanalysis model level as the most poleward maximum in westerly wind speed that exceeds 30 m s^{-1}, and the locations of the 30 m s^{-1} contour crossings poleward and equatorward of this define the boundaries of the subvortex jet region. The bottom of the subvortex jet often extends down to the level of the tops of the upper tropospheric jets. To distinguish between the two in such cases, we first identify the subvortex jet at levels down to a pressure near 300 hPa. We then work down from the level nearest 80 hPa to identify the lowest altitude at which the wind speed of the jet is still decreasing with decreasing altitude; this is defined as the bottom of the subvortex jet. Those cases where the subvortex jet joins with the top of a UT jet are referred to as "merged" jets here.

The tropopause definitions in JETPAC were discussed in *Section 7.3*.

7.4.2 Climatology of UTLS jets in reanalyses

Manney et al. (2017) described a comprehensive comparison of UTLS jets and multiple tropopauses. An important aspect of this study was to assess differences between alternate products from individual reanalysis systems. Of particular note is a comparison of CFSR/CVSv2 data on model and pressure levels that highlights the importance of vertical grid spacing for both tropopause and UTLS jet characteristics (*e.g.*, **Figure 7.11** shows such a comparison for SON; other seasons show generally similar differences, albeit somewhat smaller in the equinox seasons). The pressure level data substantially underestimate upper tropospheric jet and multiple tropopause frequencies primarily because of the coarser spacing of the levels, while they overestimate merged jet frequencies for the same reason (because relatively shallow layers where neither upper tropospheric nor subvortex jets exist are missed).

Figure 7.12 shows differences in the climatological (1980 through 2014) zonally averaged annual cycle of upper tropospheric jets frequencies, with other reanalyses compared to MERRA-2; see *Manney et al.* (2017) for a discussion of significant regional differences among the reanalyses. Generally good qualitative agreement is seen among the reanalyses for large-scale climatological features in UTLS jet distributions, but quantitative differences are sufficient that they could have significant consequences for transport and variability studies. Most of the differences in distributions of UTLS jets were found to be consistent with differences in assimilation model grids and resolution – for example, ERA-Interim (with coarsest native horizontal resolution) typically shows a significant low bias in upper tropospheric jet frequencies with respect to MERRA-2, and JRA-55 a more modest one, while CFSR/CFSv2 (with finest native horizontal

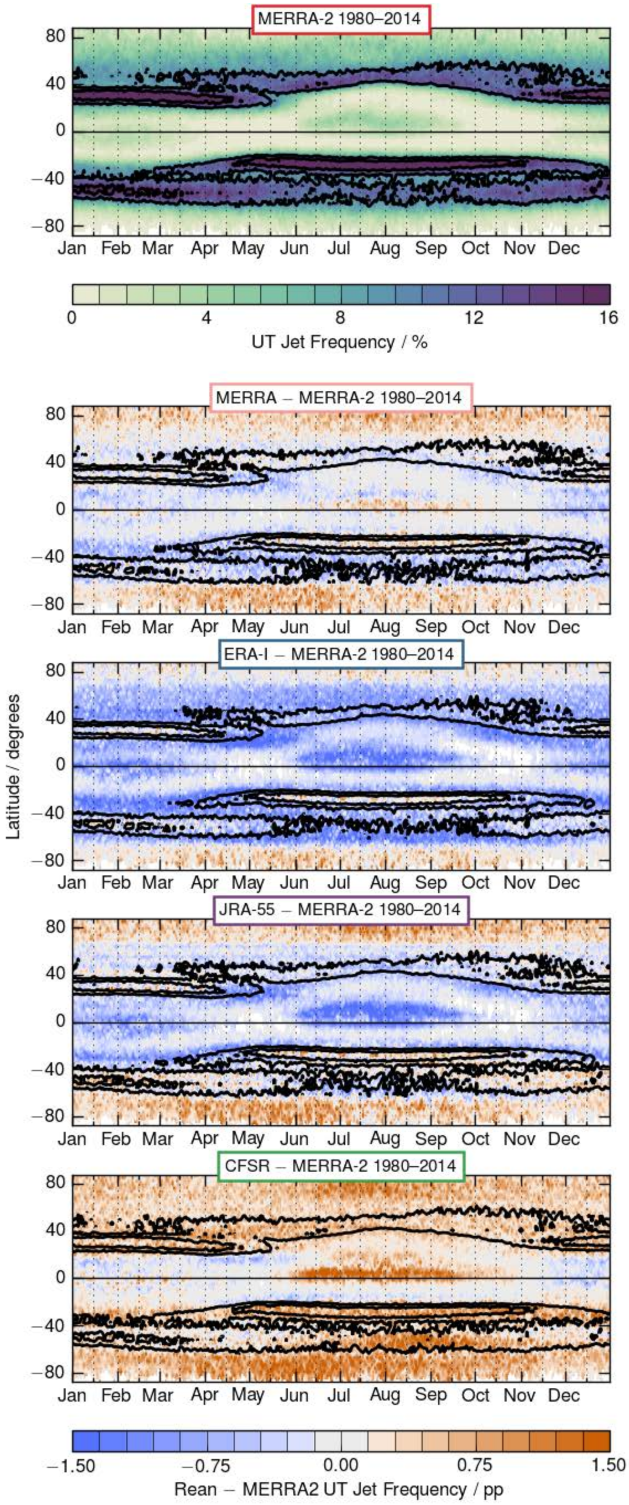


Figure 7.12: *Climatological (1980 - 2014) annual cycle of zonally averaged frequency distributions of upper tropospheric jets from MERRA-2 (top), and differences between those frequency distributions and the other reanalyses (remaining rows). Black overlays show frequency contours of 10 % and 15 % from the corresponding reanalysis in each panel. Adapted from Manney et al., 2017.*

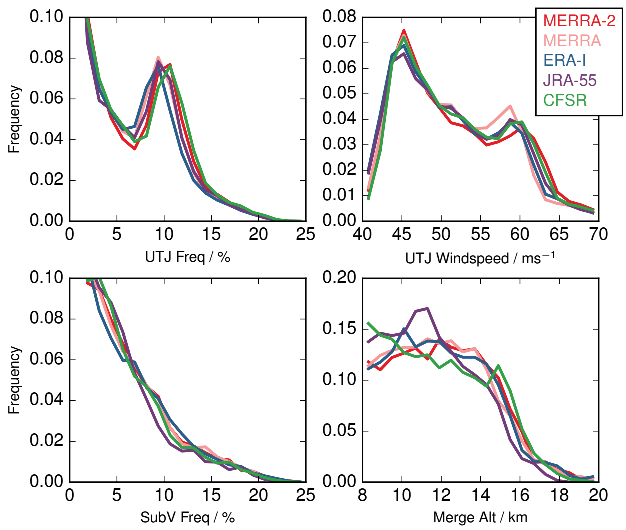


Figure 7.13: *Summary of globally / sesonally averaged frequency distributions for (a) upper tropospheric jet frequency distributions and wind speeds and (b) subvortex jets in five reanalyses. Adapted from Maney et al., 2017.*

resolution) shows a high bias with respect to MERRA-2. Agreement between the subvortex jets characterized using model-level data was also generally good, with ERA-Interim showing slightly higher and JRA-55 slightly lower maximum subvortex jet frequencies than MERRA-2 in NH winter (see *Manney et al.*, 2017). Because the subvortex jets are identified on individual levels, whether a reanalysis shows higher or lower frequencies is dependent on a complex interplay of horizontal resolution, vertical resolution, and vertical grid level locations, as well as potentially other model differences.

Figure 7.13 shows a top level summary of UT jet frequencies and windspeeds, as well as subvortex jet frequencies and merge altitudes. When averaged globally and seasonally, UT jets show very good agreement among the reanalyses in frequency and windspeed. The subvortex jet frequencies also show very good agreement, but with slightly larger differences in the altitude at which subvortex and UT jets merge. As noted above and by *Manney et al.* (2017), the merge altitude is very sensitive not only to the vertical resolution, but also to the specific altitudes of the model levels.

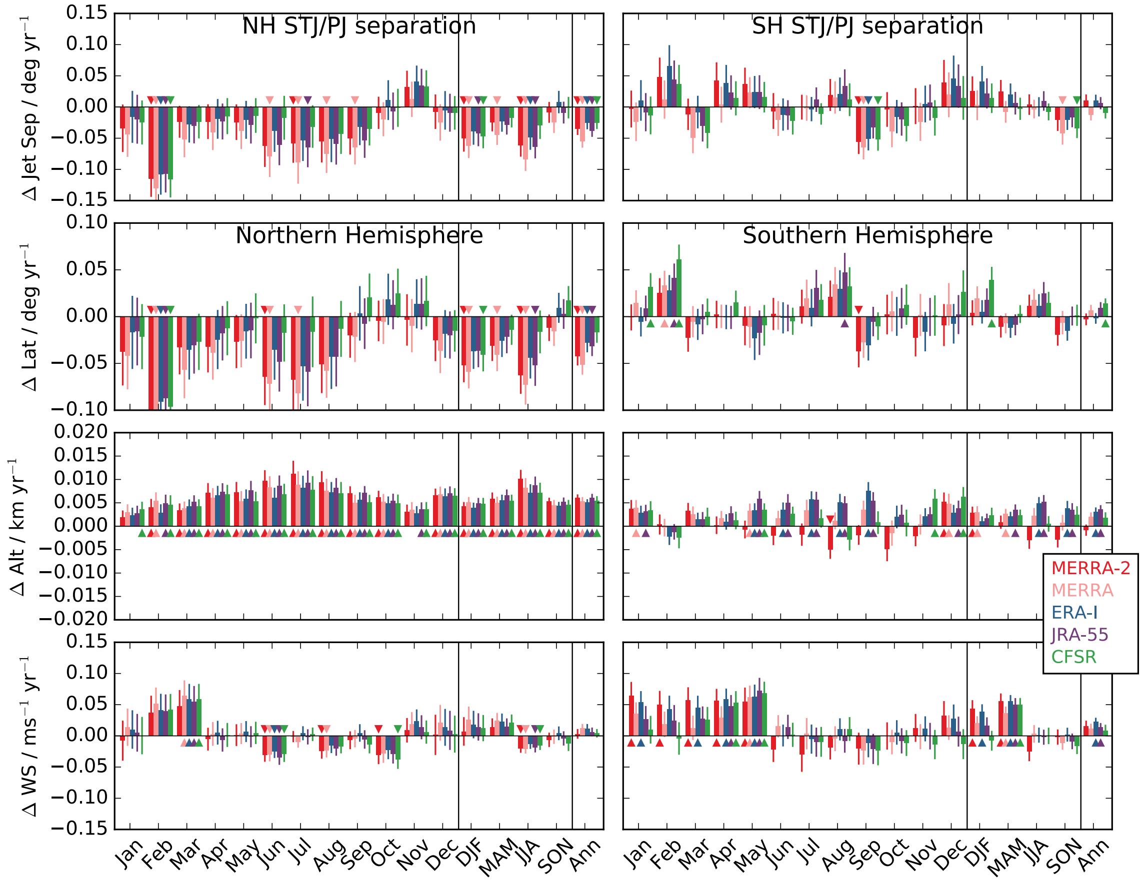


Figure 7.14: *Bar charts of zonally averaged NH and SH polar jet and polar/subtropical jet separation trends as a function month and season, showing five reanalyses. The bars show the slopes of the fits, and the error bars (centered about the top of the bars) show the 1-σ uncertainty in that slope. Note that the absolute value of latitude is used, so positive slopes (bars extending upward from the zero line) indicate a poleward shift in both hemispheres. The zero line in each case indicates no trend in the quantity shown. Triangles indicate cases where a permutation analysis (see Manney and Hegglin, 2018a) shows the slope to be significant at the 95 % confidence level. (From Manney and Hegglin, 2018a). © 2018 American Meteorological Society, used with permission.*

7.4.3 Trends in UTLS jets in reanalyses

Manney and Hegglin (2018a) (see *Manney and Hegglin*, 2018b, for figure labeling corrections) examined variability and trends in UT jets by using the JETPAC fields to calculate the monthly and seasonal mean latitude, altitude, and wind speed of the jets, both averaged over all longitudes and by longitude region. Trends in STJ latitude are used as a measure of tropical width, and the results for this diagnostic, as well as other jet and tropopause based diagnostics, are discussed in *Chapter 8* of this report. **Figure 7.14** shows monthly, seasonal, and annual zonally averaged trends in PJs from five reanalyses, and **Figure 7.15** shows PJ trends as a function of longitude for DJF. For the most part, the NH polar jet shows a relatively robust equatorward and upward shift, with good consistency among the reanalyses. However, there are some times (*e.g.*, October to November in the zonal mean) and regions (*e.g.*, over the north Atlantic in DJF) that show poleward shifts or inconsistent shifts. Zonal mean trends are less consistent among reanalyses in the SH, and smaller in the zonal mean; however, a robust poleward shift of the SH PJ is seen in DJF (*Manney and Hegglin*, 2018a) except in the eastern to central Pacific. In general, the trends vary strongly with both longitude and season (see *Manney and Hegglin*, 2018a, for a detailed summary of all regional and seasonal trends and their significance), and in many cases the trends are not robust, either because they are not statistically significant or because they do not agree among all the reanalyses. As discussed in detail in *Chapter 8*, there are only a few regions / seasons with robust tropical widening, and also some with robust tropical narrowing. Agreement among the reanalyses in the trend direction is a necessary (but not sufficient) condition to consider a trend robust. *Manney and Hegglin* (2018a) found several cases where one or more reanalyses showed a statistically significant trend that was opposite in sign to that from other reanalyses. In particular, there are several cases in the SH when either MERRA-2 or CFSR/CFSv2 shows opposite behavior to the other reanalyses (*e.g.*, **Figure 7.15**, lower right panel).

Although some of the reanalyses do show better or worse agreement in assessment of trends, because the attribution of trends in reanalyses can be so strongly dependent on possible changes in the input data, we

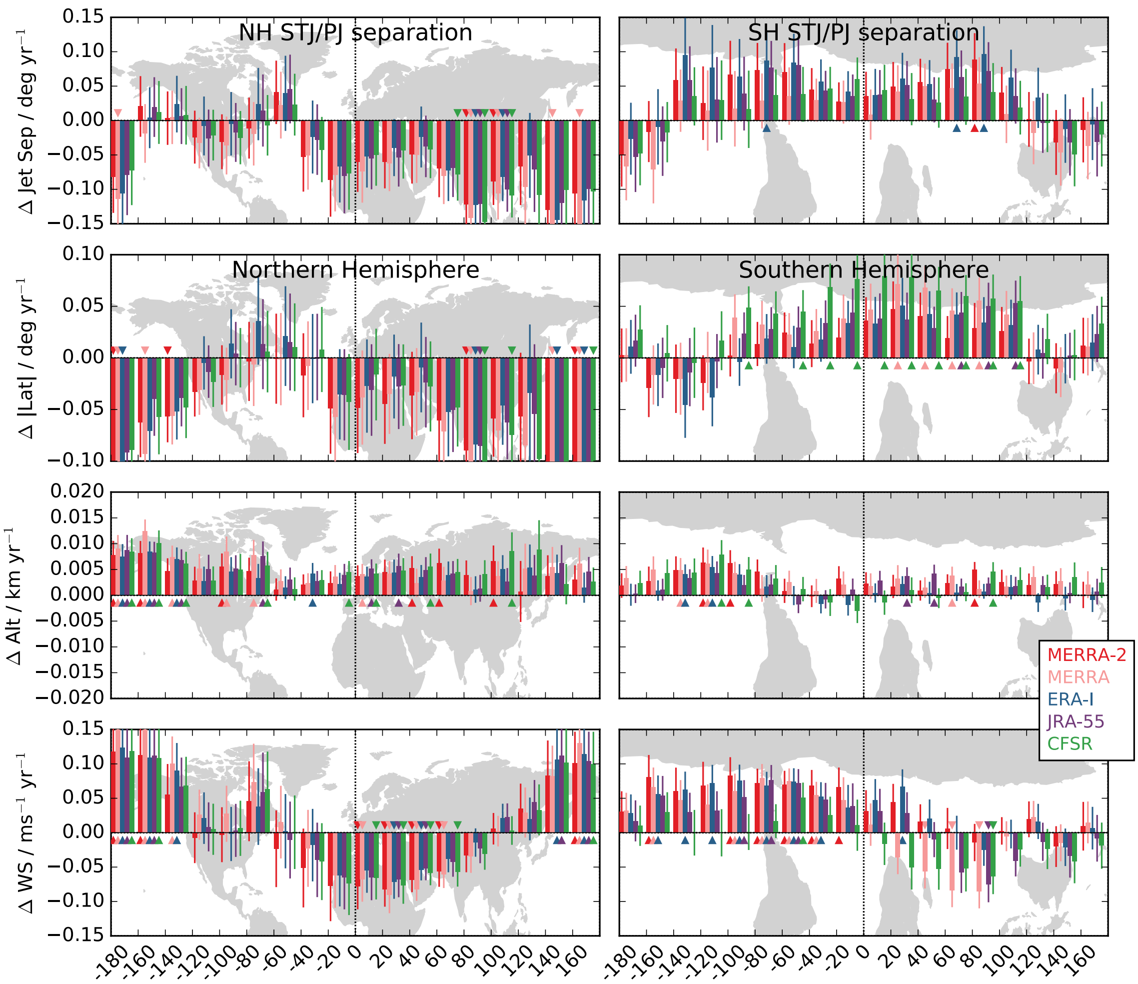


Figure 7.15: *Bar charts of global polar jet and polar/subtropical jet separation trends as a function of longitude in 20-degree bins, showing five reanalyses for DJF. Layout is as in* ***Fig. 7.14****. (From Manney and Hegglin, 2018a). © 2018 American Meteorological Society, used with permission.*

recommend extreme caution in attempting to evaluate trends from reanalysis data. However, because many diagnostics (such as jet core locations) cannot be obtained from observational data, consistency among the reanalyses is an important condition for concluding that an apparent trend *may* be robust. Furthermore, given the sensitivity to horizontal and vertical resolution of jet characteristics demonstrated in *Section 7.4.2*, and large regional and seasonal variability, it is recommended that trend studies should use data on the model grids when possible and account for regional and seasonal variability.

7.5 Transport and mixing

Transport and mixing in the ExUTLS have important impacts on the chemical and radiative characteristics of this layer. In particular, transport that involves exchanges of air across the tropopause (STE) most often leads to the greatest impacts. As discussed in the Introduction, only large-scale transport processes are resolved in reanalyses, so diagnostics used to compare transport and mixing are limited to such scales here.

7.5.1 Stratosphere-troposphere exchange

STE is commonly examined using trajectory-based (*i.e.*, Lagrangian) methods. Such trajectories are driven by horizontal winds and either kinematic vertical velocity (*i.e.*, omega) or diabatic heating rates for the vertical component. Trajectory paths are compared to a representation of the tropopause (commonly an iso-surface of PV, but alternatively the lapse-rate tropopause) and those that cross this surface are identified as either troposphere-to-stratosphere transport (TST) or stratosphere-to-troposphere transport (STT). Eulerian methods to calculate STE provide a complementary bulk transport diagnosis and may differ considerably from Lagrangian methods.

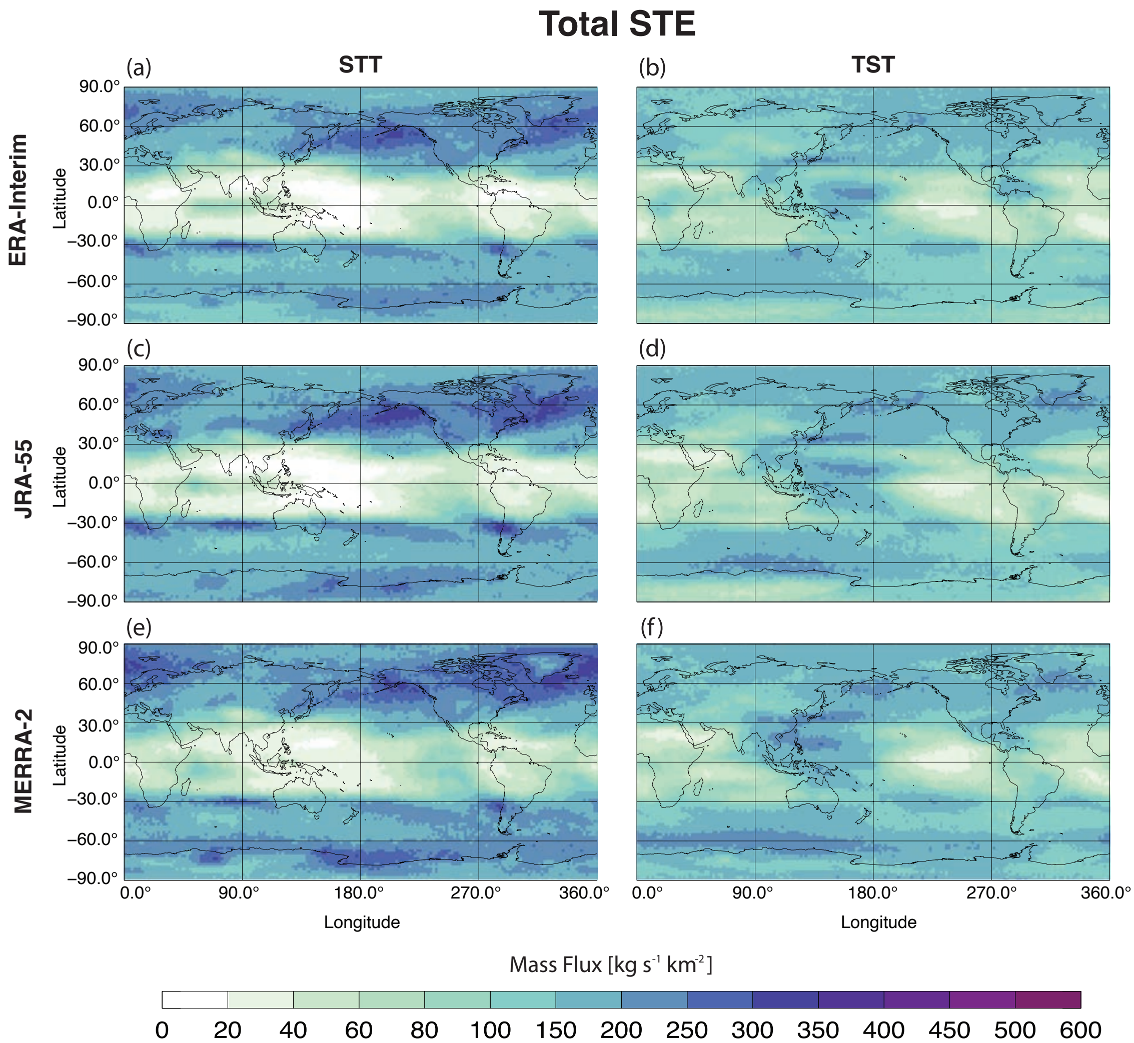


Figure 7.16: *A modified version of Figure 4 from Boothe and Homeyer (2017): global distributions of annual-mean (left) STT and (right) TST from (top-to-bottom) ERA-Interim, JRA-55, and MERRA-2. These STE estimates were calculated using a trajectory model. This figure is modified from that in Boothe and Homeyer (2017) by excluding results for MERRA and revising the analysis to use MERRA-2 wind fields from the "ASM" collection (instead of the "ANA" fields used in the original; see Chapter 2 for discussion of ASM versus ANA MERRA-2 fields).*

Operational forecast model analyses and reanalyses have been used for STE studies over the past few decades (*e.g.*, *Škerlak et al.*, 2014; *Sprenger and Wernli*, 2003; *Seo and Bowman*, 2002; *Wernli and Bourqui*, 2002; *Stohl*, 2001; *Appenzeller et al.*, 1996). Some studies have been regional, focused on single transport processes, or limited to short time periods (*i.e.*, a few years or less). Others have evaluated global transport over longer time periods. Despite the common use of trajectories in these studies, differences in trajectory integration times, conditions applied to categorize an individual particle's path as irreversible exchange, and the input wind fields have led to significant differences in estimates of STE. Few studies have conducted STE calculations using multiple wind fields (*e.g.*, those from more than one reanalysis).

Boothe and Homeyer (2017) conducted trajectory calculations driven by the 3D wind fields of four modern reanalyses (ERA-Interim, JRA-55, MERRA, and MERRA-2) over a 15-year period (1996 - 2010) to determine global STE. In particular, forward and backward trajectories were computed each day for a global 3D lapse-rate tropopause-relative grid of particles, each having constant mass. Trajectories that crossed the tropopause during 1 day downstream and remained in their destination reservoir (*i.e.*, stratosphere for TST, troposphere for STT) for at least 4 out of 5 days downstream were flagged as likely irreversible transport. These particles were also required to have been in their parent reservoir (*i.e.*, troposphere for TST, stratosphere for STT) for at least 4 out of 5 days upstream to be kept for STE analyses. For complete details on the trajectory model used and STE identification methods, the reader is referred to *Boothe and Homeyer* (2017).

amounts of TST, STT, and net STE (TST-STT) occur primarily in the extratropics.

Analysis of the long-term variability of TST and STT was also found in *Boothe and Homeyer* (2017) to differ considerably among the reanalyses. In particular, for ERA-Interim and JRA-55, TST was found to increase in the tropics and STT was found to increase in the extratropics during the 15-year study period, while MERRA and MERRA-2 showed the opposite behavior. MERRA also showed large increases in TST in the extratropics, while the remaining analyses showed little change in this component of STE. **Figure 7.17** shows these results from *Boothe and Homeyer* (2017).

While the objective of *Boothe and Homeyer* (2017) was to compare STE in the reanalyses, questions remain on the source of the differences found. The authors did show that differences in STE amounts are accompanied by consistent differences in the frequency of exchange events. The authors also hypothesize that differences in the dynamics (both horizontal and vertical motion), tropopause altitudes, assimilated datasets, and model grids may contribute to the observed differences in STE. Evidence for systematic differences in vertical motion and tropopause altitude was given in *Boothe and Homeyer* (2017) and also in **Figures 7.1** and **7.2** of this report.

One caveat of the *Boothe and Homeyer* (2017) study is that the MERRA and MERRA-2 wind fields used were not those recommended for transport studies by the NASA team. Guidelines were released after the *Boothe and Homeyer* (2017) study to specify that ASM fields should be

Findings from *Boothe and Homeyer* (2017) include important differences in the magnitudes, geographic locations, annual cycles, and long-term changes and variability of STE between the reanalyses. The authors separate STE into three regions (tropics, subtropics, and extratropics) and two directions (TST and STT) to further evaluate the similarities and differences in STE among the reanalyses. **Figure 7.16** shows comparisons of the geographic distributions of annual mean TST and STT from three of the four reanalyses (modified from *Boothe and Homeyer,* 2017). These distributions highlight some of the important differences found in the locations of TST and STT maxima, especially in the tropics. Despite these differences, *Boothe and Homeyer* (2017) show that the annual cycles of TST and STT are similar among the reanalyses and that differences in the

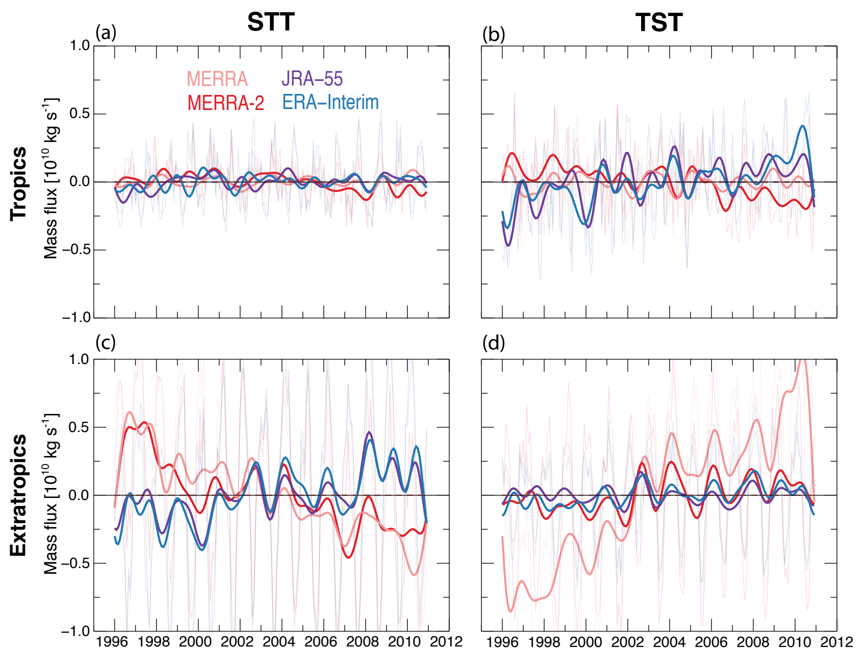


Figure 7.17: *Figure 14 from Boothe and Homeyer (2017): 15-year timeseries of long-term mean relative (left) STT and (right) TST in the (top) tropics and (bottom) extratropics in ERA-Interim, JRA-55, MERRA-2, and MERRA. These STE estimates were calculated using a trajectory model.*

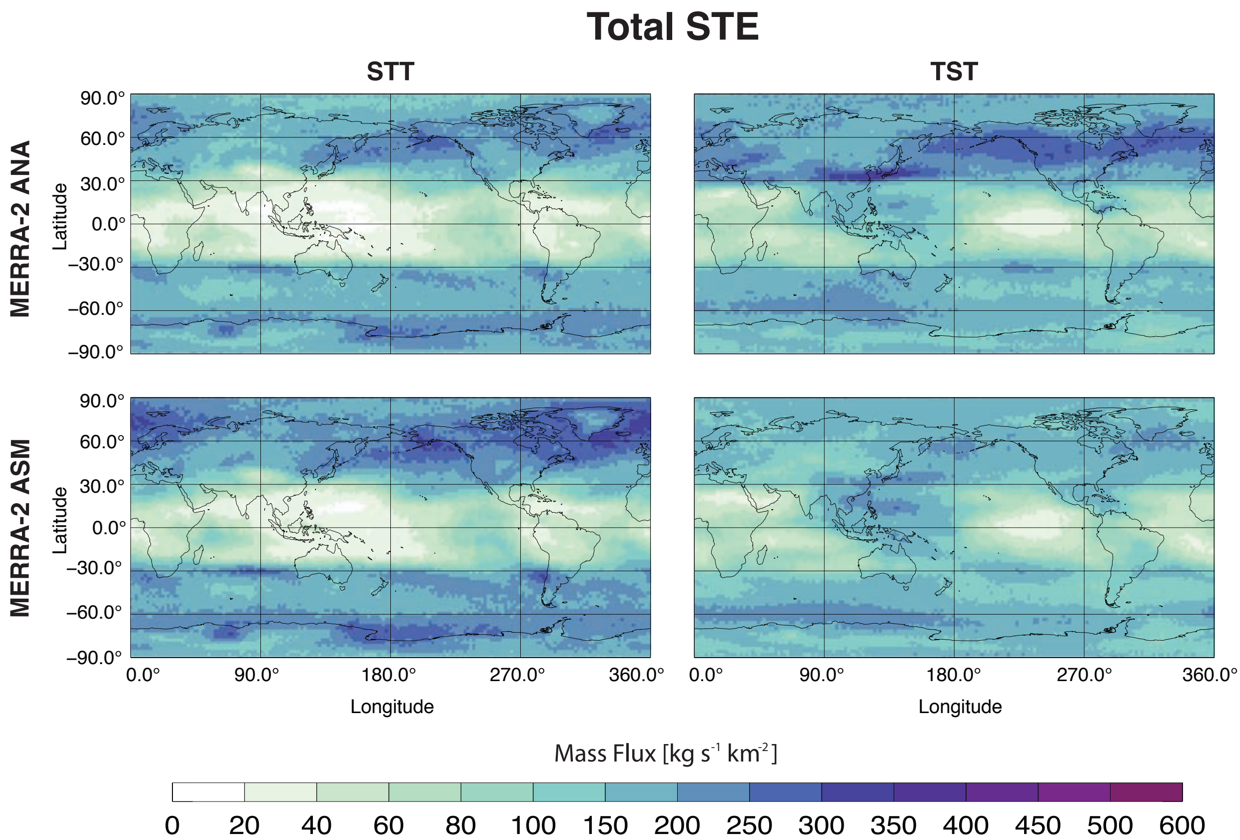


Figure 7.18: *A comparison of (top) MERRA-2 ANA panels from Figure 4 of Boothe and Homeyer (2017) and (bottom) the revised MERRA-2 ASM analysis. Global distributions of annual-mean (left) STT and (right) TST. These STE estimates were calculated using a trajectory model.*

used instead of ANA fields. Thus, an update to the *Boothe and Homeyer* (2017) analysis was completed to determine differences in transport calculated using 3D winds from these two products. **Figure 7.18** compares geographic distributions of global mean STT and TST from the MERRA-2 ANA and ASM analyses. While some slight differences in patterns are observed, the biggest change in STE results from using ASM fields instead of ANA is that STT increases and TST decreases, leading MERRA-2 transport patterns and magnitudes to be more similar to those of ERA-Interim and JRA-55. Time series analyses shown in *Boothe and Homeyer* (2017) were also revisited, but no significant changes in the results were found (*i.e.*, long-term variability and changes are consistent in the ANA and ASM analyses; not shown).

7.5.2 Mixing and transport barriers

PV-based diagnostics in equivalent latitude (EqL) coordinates can provide information on mixing and transport barriers. In particular, PV gradients indicate the strength of transport barriers such as the stratospheric polar vortex or the tropopause (*e.g.*, *Manney and Lawrence*, 2016; *Kunz et al.*, 2011a; *Mahlman*, 1997; *McIntyre and Palmer*, 1983, and references therein). Effective diffusivity (K_{eff}) is also commonly used to assess the location and strength of mixing and transport barriers in stratospheric and UTLS studies (*e.g.*, *Abalos et al.*, 2016; *Allen and Nakamura*, 2001, 2003; *Haynes and Shuckburgh*, 2000a,b, and references therein). Here we show comparisons of PV gradients and K_{eff} as a function of equivalent latitude and time for 2005 through 2015 to assess potential differences in the representation of mixing and transport barriers in reanalyses. For this analysis, K_{eff} is calculated directly from the PV fields (as described by, *e.g.*, *Santee et al.*, 2011; *Manney et al.*, 2009, and references therein), with PV used on the native model levels, as described by *Lawrence et al.* (2018). As noted by *Lawrence et al.* (2018), some caution is required in using PV fields from different reanalyses as they are derived from the reanalysis fields provided in different ways. The calculation of PV gradients and (especially) K_{eff} depends on horizontal resolution. To the extent that overall biases represent the ability of the reanalyses to resolve small-scale mixing processes, these differences may be helpful in evaluating the use of different reanalyses in transport studies. However, scaling K_{eff} (which is typically used as a qualitative measure of mixing and transport barriers) to a similar range allows more quantitative comparison of the locations, times, and relative strength of mixing regions and transport barriers. We thus scale K_{eff} by subtracting the global climatology for 2005 - 2015 for each reanalysis from the daily values and dividing by the standard deviation of that climatological mean. The time period

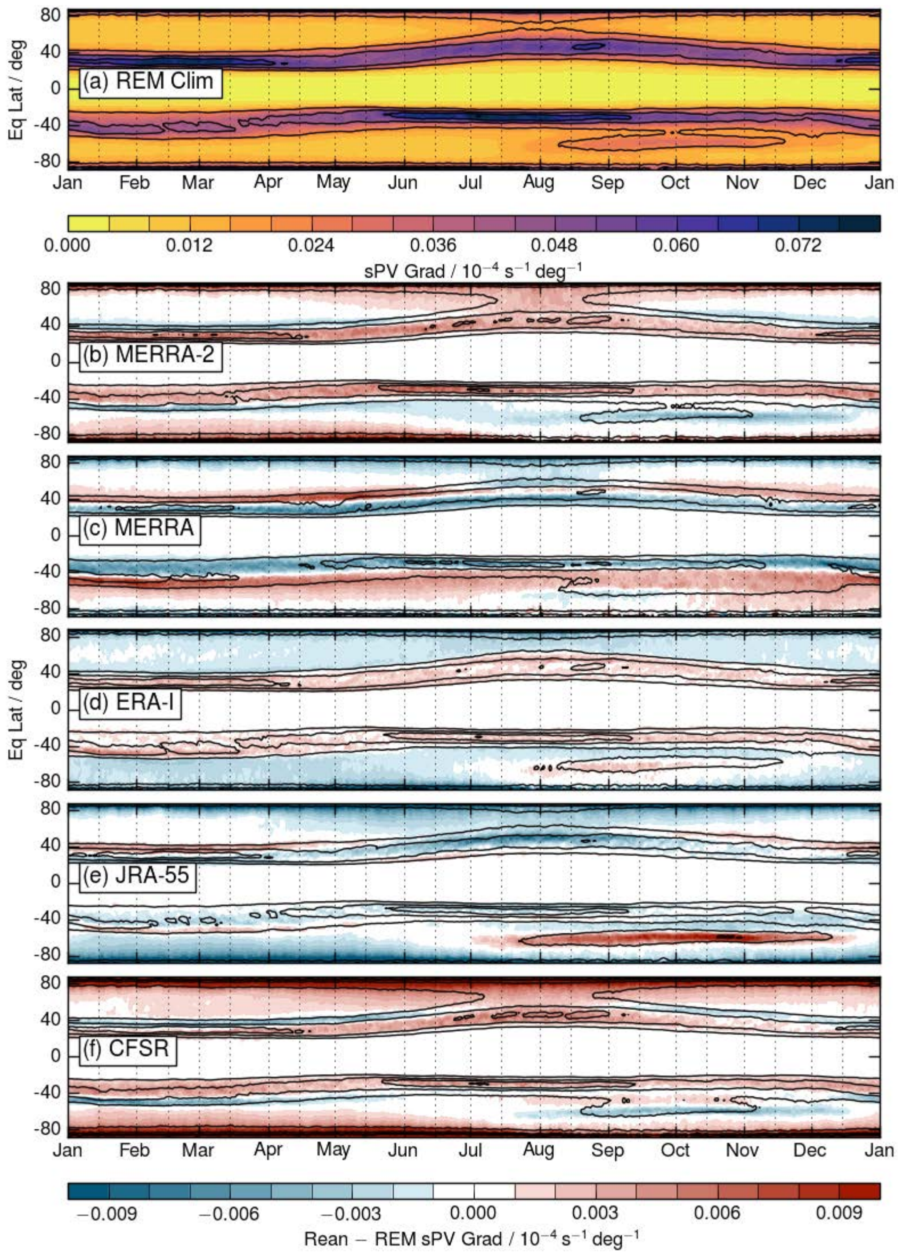


Figure 7.19: *Climatological (2005 - 2015) annual time series of (Top) sPV gradients at 350 K as a function of equivalent latitude from a reanalysis ensemble mean (REM, including MERRA-2, MERRA, ERA-Interim, JRA-55, and CFSR/CFSv2 reanalyses), and (following rows) the difference of each reanalysis from the REM. The black overlays show the same selected contours, from the REM on the top panel, and each of the reanalyses on the following panels.*

2005 through 2015 was chosen to facilitate comparisons of EqL/time series of assimilated and MLS ozone (see *Section 7.6* below).

Figures 7.19 and **7.20** show climatological scaled PV (sPV) gradients as a function of EqL on the 350 K and 390 K isentropic surfaces, respectively. Plots of sPV in the same format (not shown; also see *Millán et al.*, 2021) indicate that, while biases exist between the reanalyses' sPV fields in the UTLS, they are typically less than about 10 % except near the equator (where sPV values themselves are very low). The sPV gradients at 350 K are largely tropospheric in character, with strongest gradients along the UT subtropical jets, whereas the sPV gradients at 390 K are largely stratospheric, with strongest gradients along the polar vortex edges in winter. In the SH, enhanced gradients near 60 ° S EqL demark the lowest extension of the subvortex. At 390 K, the top of the subtropical jet is apparent in each hemisphere (more clearly in the NH) as enhanced sPV gradients near 30 ° S EqL from about November through May in the NH and May through August in the SH, times with the strongest sPV gradients along the UT subtropical jet at lower levels.

The differences among reanalysis sPV gradients are generally modest, on the order of 10 % (much larger differences at the highest EqLs are likely due to noise in the sPV fields there and are not physically meaningful). The

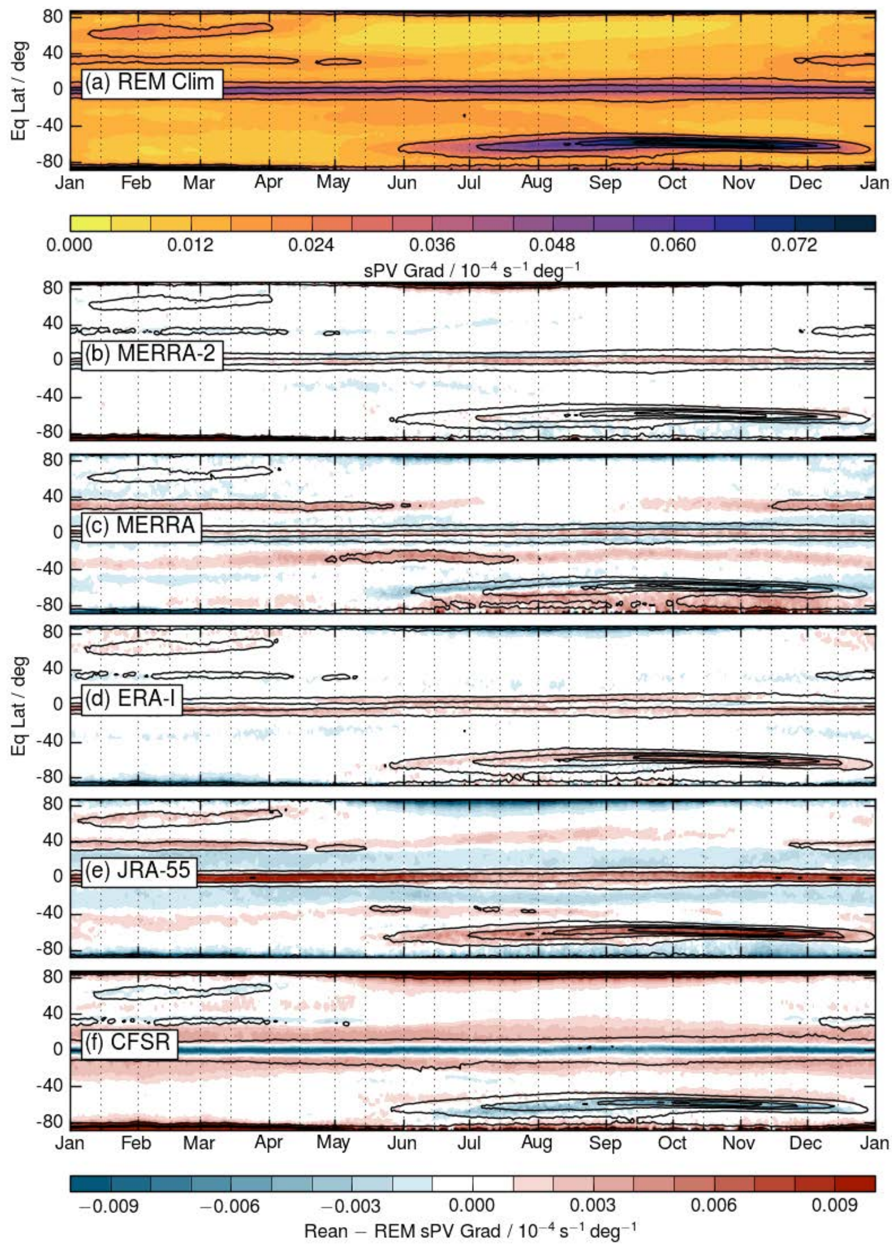


***Figure 7.20:** As in **Fig. 7.19** but at 390 K.*

differences at 350 K along the NH subtropical jet tend to have a dipole structure in latitude, suggesting small differences in the location of the strongest sPV gradients; the patterns of biases with respect to the jets appear to be largely consistent throughout the annual cycle. MERRA-2 and (especially) CFSR/CFSv2 350 K sPV gradients are generally stronger than those in the REM, while those in the other reanalyses tend to be weaker, which may be related to the higher horizontal resolution of those reanalyses.

Differences in sPV gradients among the reanalyses at 390 K are still generally modest, and are largest at the locations of the winter stratospheric subvortex jet. Near the SH subvortex jet, ERA-Interim and JRA-55 show stronger gradients and CFSR/CFSv2 weaker gradients than the REM, while MERRA-2 gradients are close to the REM and MERRA shows a dipole pattern suggestive of a slightly more poleward transport barrier. A similar pattern is seen near the NH subvortex jet, but the differences are much smaller.

The patterns of climatological K_{eff} (**Figures 7.21** and **7.22**) are consistent with those in the sPV gradients, with low/high values of K_{eff} in regions of high/low sPV gradients. Strong mixing regions are seen at 390 K during and following the breakup of the stratospheric vortices in the lowermost stratosphere (May through October in the NH, November through April in the SH). At 350 K, the transport barriers align with the UT subtropical jets, with relatively strong mixing away from those regions, except from about August through October in the SH when the subvortex jet presents a significant transport barrier.

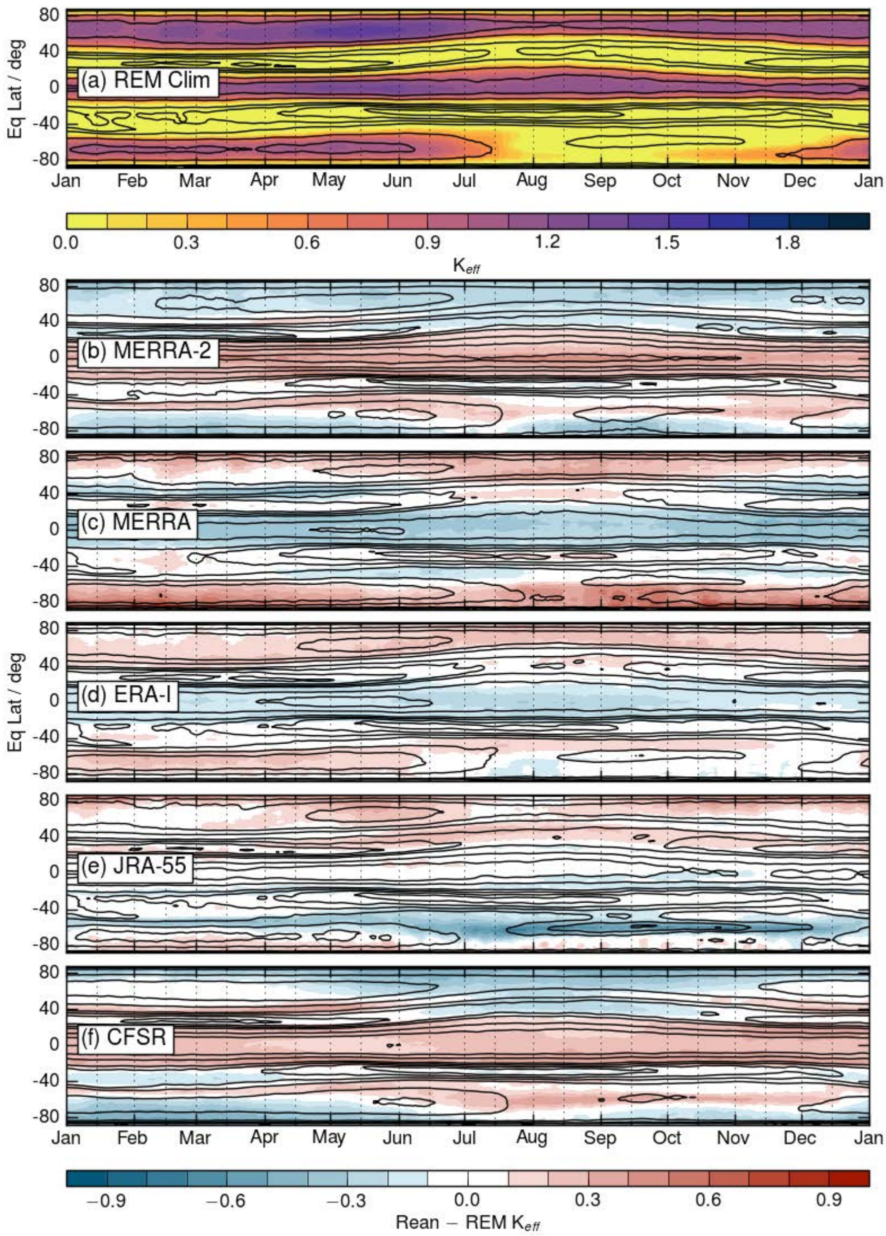


***Figure 7.21:** As in **Fig. 7.19** but for effective diffusivity, scaled as described in the text.*

At 350 K, MERRA, JRA-55, and ERA-Interim tend to have higher values in the strong mixing regions poleward of the subtropical jets, suggesting more mixing. MERRA-2 and CFSR/CFSv2 show higher values at low latitudes, suggesting more mixing in the tropics. At 390 K, the reanalyses show large differences in the transport barrier at the edge of the SH subvortex jet, with MERRA and ERA-Interim showing higher values and MERRA-2 and CFSR/CFSv2 lower values than the REM. While quantitative differences at all UTLS levels in K_{eff} are relatively large, the qualitative seasonal patterns of mixing and transport barriers are captured well in all of the reanalyses.

Figure 7.23 shows a comparison of interannual variability in K_{eff} at 350 K among the reanalyses. Overall patterns are similar to those seen in the climatology, indicating substantial differences in the K_{eff} gradients but good agreement in the timing and location of mixing and transport barriers. Of particular note is a stepwise change at the time of the transition between CFSR and CFSv2, suggesting that overall CFSv2 indicates more mixing than CFSR; a similar stepwise change is seen at other UTLS levels. The other fields evaluated in the EqL/time plane (sPV and its gradients, wind speed, assimilated ozone) show no more than small discontinuities at this time; the large discontinuity in K_{eff} probably arises because that calculations is highly sensitive to "noise" (that is, small scale structure) in the fields, which is likely to have changed across the CFSR/CFSv2 transition.

While relatively large differences in K_{eff} magnitudes and ranges (even when scaled by the global mean and standard deviation) argue against any quantitative use, all of the reanalyses capture well the timing and locations of mixing regions and transport barriers.

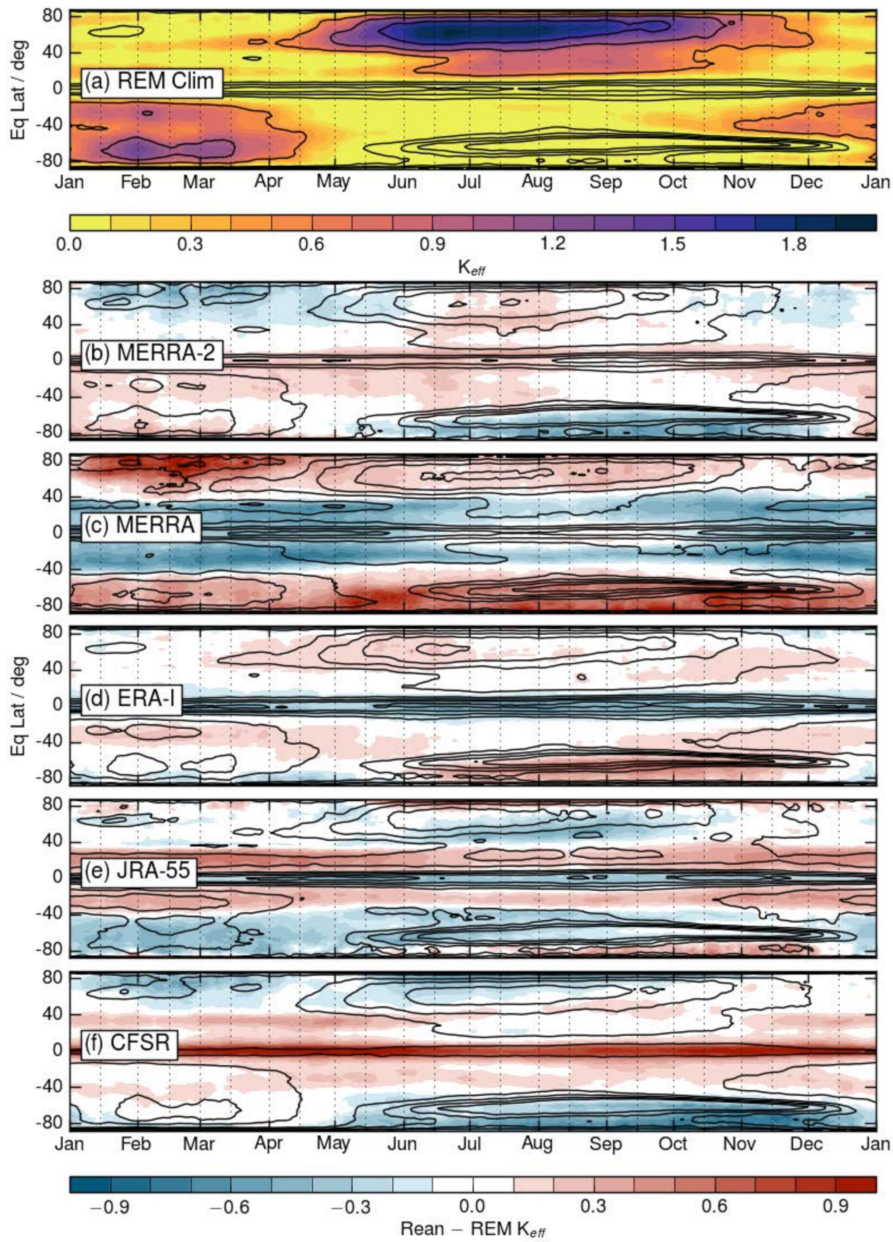


Figure 7.22: *As in* ***Figure 7.20*** *but for effective diffusivity, scaled as described in the text.*

7.5.3 Mass flux across 380 K isentropic surface

The flux of mass across the 380 K potential temperature surface can be used to directly measure TST in the tropics and also as a proxy for the net STT in the extratropics (*e.g.*, *Olsen et al.*, 2013). The 380 K isentrope is assumed to be the lowest potential temperature surface that lies entirely at or above the tropopause for all seasons; it does not intersect the tropopause where isentropic cross-tropopause flux can occur. Thus, the 380 K cross-isentrope transport computed from the net heating rate is used to estimate the net bulk flow through the tropopause in the region below. The hemispherically integrated extratropical flux can also be considered a single-valued quantity related to the stratospheric circulation in that hemisphere. A change in the net extratropical flux of mass must necessarily be caused by some change in the stratospheric circulation. The net flux of a chemical species, such as ozone, across the 380 K potential temperature surface can be interpreted as the convolution of the total air mass flux and the concentration of the species near the surface. (These quantities are not entirely independent since the transport will impact the concentration of the species). Therefore, it is valuable to evaluate and compare the 380 K air mass flux in the reanalyses, particularly since the meteorological fields are frequently used to drive CTMs.

The radiative heating rate information provided for each reanalysis (see *Chapter 2*) is postprocessed to get total daily-mean radiative heating rates. In some reanalyses total heating rates are provided, in others (*e.g.*, JRA-55) all of the physical terms provided are summed to get them. The general procedures used and details for each reanalysis are

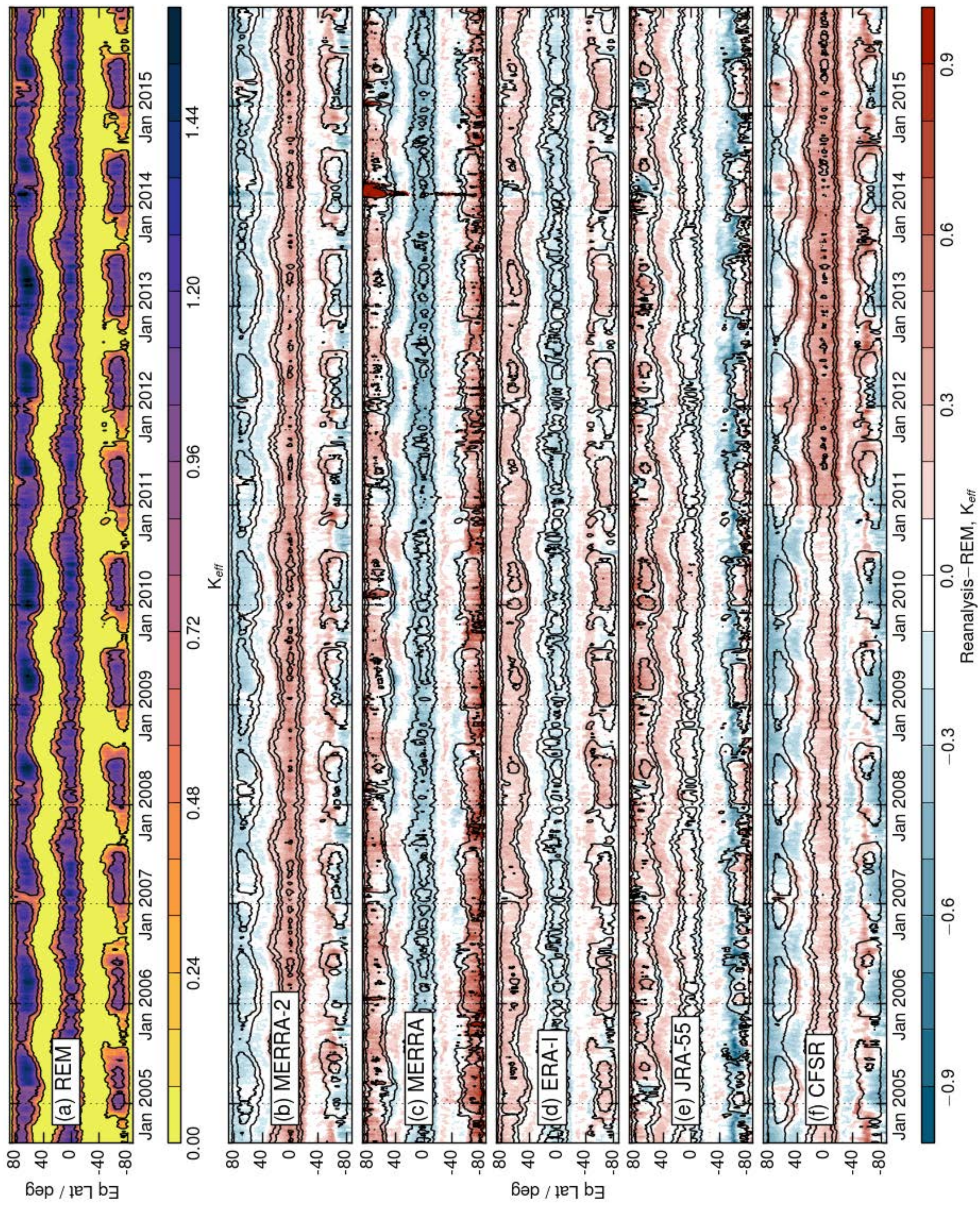


Figure 7.23: *Time series of effective diffusivity (scaled as described in the text) on the 350 K isentropic surface for 2005 through 2015 as a function of equivalent latitude from (top) the REM, and (following rows) the difference of each reanalysis from the REM. The black overlays show the same selected contours from the REM on the top panel and each of the reanalyses on the following panels.*

summarized in the context of zonal means by *Martineau et al.* (2018); here, the fields are used on a 1 (for ERA-Interim and CFSR) or 1.25 (for MERRA, MERRA-2, and JRA-55) degree latitude/longitude grid. ASM fields are used for MERRA and MERRA-2. The flux across the 380 K surface is calculated for each reanalysis using these diabatic heating rates interpolated to the 380 K surface, as follows:

$$F = -\int \left(\frac{\dot{\theta}}{g}\right)\left(\frac{dp}{d\theta}\right) dA \quad (7.1),$$

where p is pressure, θ is potential temperature, $\dot{\theta}$ is the diabatic heating rate, and A is the area (*e.g.*, *Schoeberl*, 2004; *Olsen et al.*, 2004). **Figure 7.24** shows the time series for 1980 through 2010 of the annual net mass flux integrated from 30° to the pole in each hemisphere. In the NH, all of the reanalyses have similar interannual variability and are well correlated at greater than 99% confidence except for CFSR, which is not statistically significantly correlated to the other reanalyses. The relative difference between each reanalysis remains fairly constant throughout the time period, although the MERRA-2 flux shows a slight increasing trend during the last decade not seen in the others. This result appears on the surface as if it might be inconsistent with the results of *Boothe and Homeyer* (2017), but the two calculations cannot be directly compared since the 380 K surface is typically substantially above the extratropical tropopause; moreover, the uncertainties in both calculations are difficult to quantify, and may depend on different ways in which the radiative heating rates are provided for the reanalyses. Thus, understanding this possible discrepancy would require further detailed study. The multi-year mean flux and standard deviation of each time series is shown in **Table 7.1**. The multi-year mean of ERA-Interim is about 10%-20% greater than the other reanalyses excluding CFSR. However, the standard deviation of these time series remains at 3%-4% of each mean, reflecting the high correlation. In contrast, the interannual variability of CFSR is much greater with a standard deviation of 11%.

In the SH, CFSR is generally better correlated with the other reanalyses, but there is an unexplained downward jump around the year 2000 (**Figure 7.24**). The other reanalyses show a smaller apparent discontinuity around the same time; these changes could result from the relationship of temperature changes during the TOVS/ATOVS transition around 1998 to 1999 (see *Chapters 2* and *3*, and *Long et al.*, 2017) being reflected in the diabatic heating rates. The difference of the ERA-Interim flux from the other reanalyses is much smaller in the SH than it is in the NH. Again, excluding CFSR, the standard deviations of the time series are consistent at 4% (**Table 7.1**). Thus, the reanalyses agree that the interannual variability is similar between the hemispheres.

Figure 7.25 shows maps of the 1980-2010 average 380 K air mass flux for each reanalysis. In all cases, the patterns are comparable, with similar locations of maxima and minima. The maximum upwelling tends to occur in a subtropical band from northeastern Africa to southeast Asia just south of 30°N. The minimum upwelling occurs just to the south along the equator from Africa to the Maritime Continent. The maximum extratropical downwelling in the SH occurs in a band between about 45°S and 60°S. In contrast, the maximum downwelling in the NH extratropics occurs in the polar region.

The mean "turn-around" latitudes (where the flux is zero) are consistent between all the reanalyses and are located at about 30°N and 30°S. The differences between the reanalyses occur primarily in the magnitude of the maxima and minima. For example, the maximum upwelling in JRA-55 over India is about 0.9 g cm^{-1} day^{-1} and the maximum in ERA-I at this same location reaches 1.3 g cm^{-1} day^{-1}.

7.6 UTLS ozone

Chapter 4 provided an overview of assimilated ozone in the most recent reanalysis and briefly discussed zonal mean diagnostics of UTLS ozone. They found persistent biases in UTLS ozone, as well as inconsistencies in the reanalyses' representations of the ozone annual cycle. We add here comparisons of diagnostics of ozone distributions

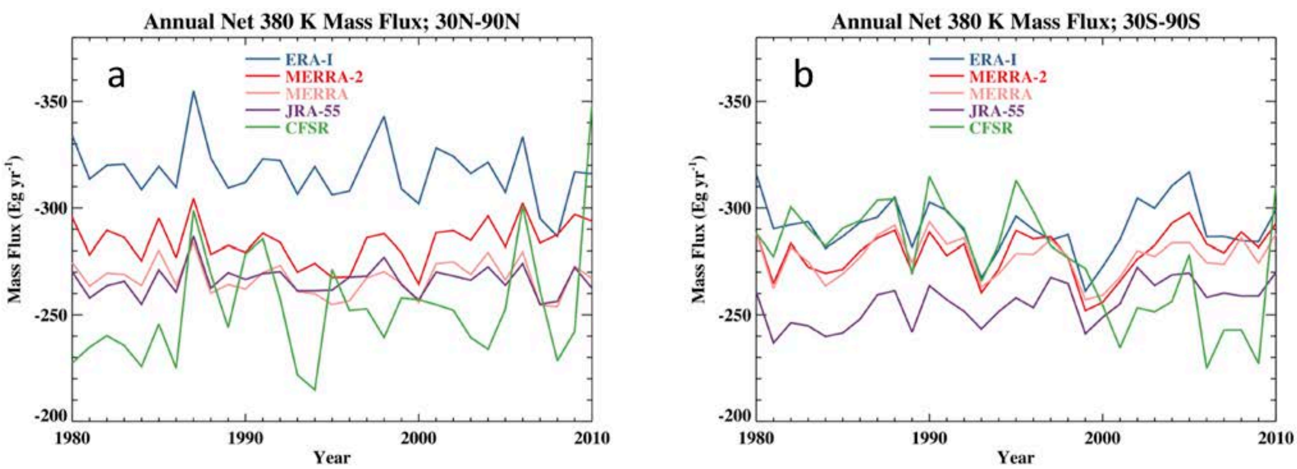


Figure 7.24: *Time series for 1980 through 2010 of the annual net mass flux integrated from 30°-90° latitude in (a) the NH and (b) the SH. Values are in 110^{18} g yr^{-1} and negative values denote a net downward flux.*

and evolution in dynamical coordinates and evaluation of transient dynamically-driven low ozone events. Reanalysis ozone fields are compared with v4 Aura Microwave Limb Sounder (MLS) data (*Livesey et al.*, 2020).

Table 7.1: *The 1980 - 2010 mean 380 K air mass flux and standard deviation (10^{18} g yr^{-1}) for each reanalysis. Standard deviations are given as a percent of the value in parentheses.*

(10^{18}g yr^{-1})	NH Mean Flux	NH Std Dev	SH Mean Flux	SH Std Dev
ERA-I	317	13 (4%)	291	13 (4%)
MERRA-2	284	10 (4%)	279	11 (4%)
MERRA	267	8 (3%)	277	10 (4%)
JRA-55	266	7 (3%)	255	10 (4%)
CFSR	253	28 (11%)	276	26 (9%)

7.6.1 Ozone in tropopause-relative coordinates

Figures 7.26 and **7.27** show reanalysis ozone profiles compared to those from MLS for 2005 through 2015, using the same classification of single and double tropopause regions described in *Section 7.3.3* and *Schwartz et al.* (2015), for each hemisphere's winter season. Generally similar patterns of differences are seen in other seasons. MERRA-2 (which assimilates MLS data throughout the period compared) shows closer agreement with MLS throughout the UTLS and lower stratosphere than the other reanalyses. At altitudes greater than about 5 km above the primary tropopause, ERA-Interim (JRA-55) ozone values become much higher (lower) than those from MLS; however, since the ozone values themselves increase rapidly, above about 10 km above the primary tropopause

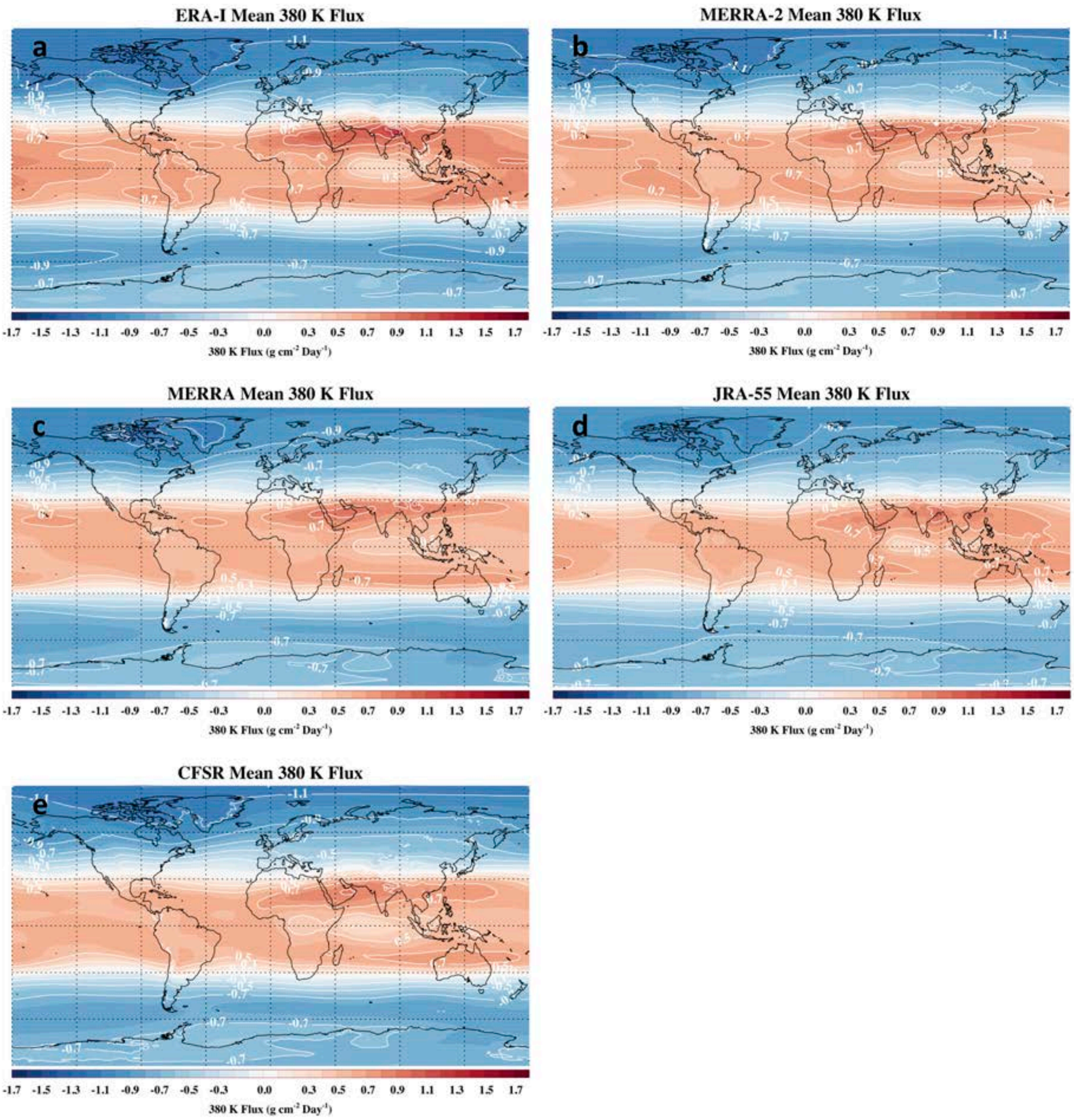


Figure 7.25: *Distribution of the 1980 - 2010 mean air mass flux across the 380 K surface for (a) ERA-Interim, (b) MERRA-2, (c) MERRA, (d) JRA-55, and (e) CFSR. White contours at increments of 0.2 g cm^{-1} day^{-1}. Negative values denote a net downward flux.*

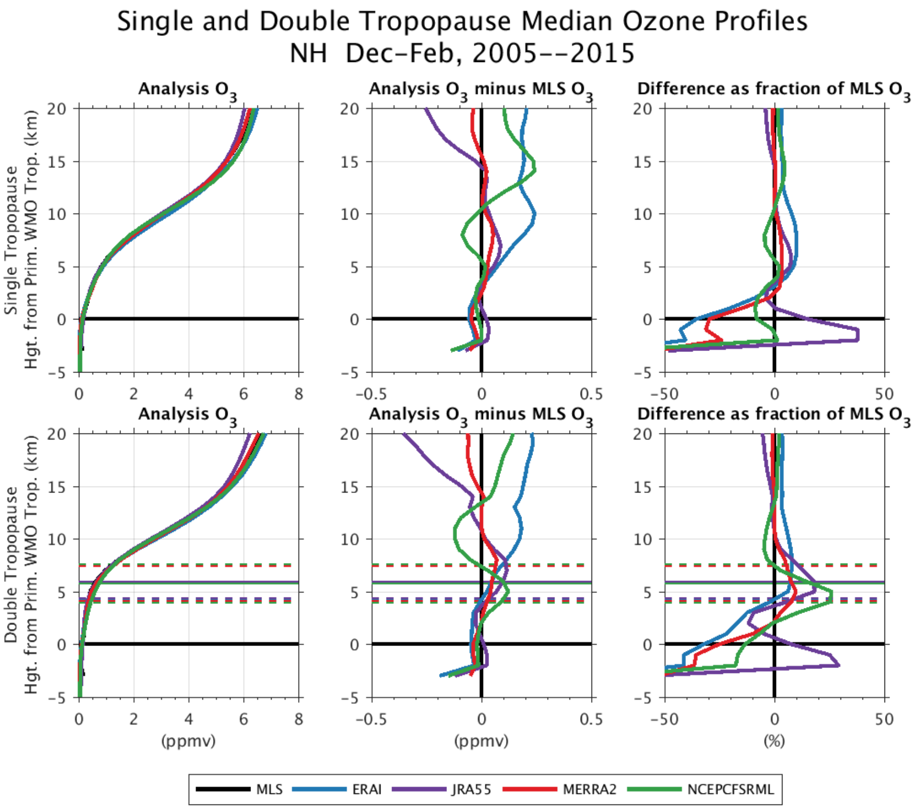


Figure 7.26: *Climatological (2005-2015) ozone profiles from MLS and four reanalyses interpolated to the MLS measurement locations for DJF in the NH in (top) single and (bottom) double tropopause regions, plotted relative to primary tropopause altitude. Left plots show the ozone profiles, center plots the mixing ratio differences from MLS, and right plots the difference from MLS expressed as a percent of the MLS value. Horizontal lines show the mean (solid) and standard deviation (dashed) of the mean secondary tropopause altitude from the primary for each reanalysis.*

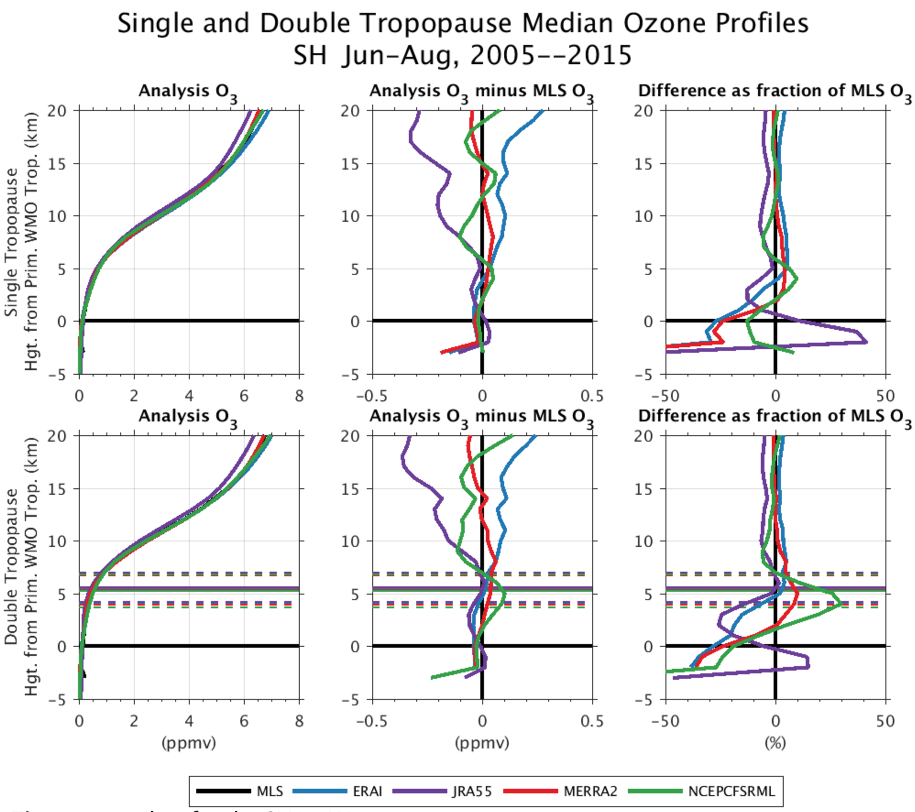


Figure 7.27: *As in* ***Figure 7.26****, but for the SH in JJA.*

the differences are less than 10%. In the region up to about 10km above the primary tropopause, the reanalyses' differences from MLS are up to about ±100ppbv (up to about 10%) in single tropopause regions and about ±200ppbv (20-30%) in double tropopause regions.

7.6.2 Ozone in equivalent latitude coordinates

Chapter 4 shows a brief overview of stratospheric (520K and 850K) and UTLS (350K) ozone as a function of EqL and time compared with MLS values. Here we update and extend this analysis with a focus on the UTLS. *Section 7.5.2* shows comparisons of some diagnostics of mixing and transport barriers for the same coordinate system and time period, which can be useful for interpretation of similarly mapped trace gas fields. **Figures 7.28** and **7.29** compare the climatological (2005-2015) distributions of ozone as a function of EqL over the annual cycle in the reanalyses with that from MLS data at 340K and 390K, respectively. At 340K strong ozone gradients are seen along the transport barrier represented by the tropopause and the subtropical jet (see, *e.g.*, sPV gradients as a function of EqL shown in *Section 7.5.2*), with high ozone in the high-latitude winter and spring arising from descent into the stratospheric subvortex. In the SH, decreasing high-latitude MLS ozone in September through October demarks the lowest extent of chemical loss in the stratospheric vortex. At 390K, the high-ozone values are confined to the polar winter regions (arising from descent in the stratospheric vortex), and a strong signature of chemical ozone loss is seen from September through December (as noted by *Manney et al.*, 2005; *Santee et al.*, 2011, the SH subvortex does not break up until late December to early January).

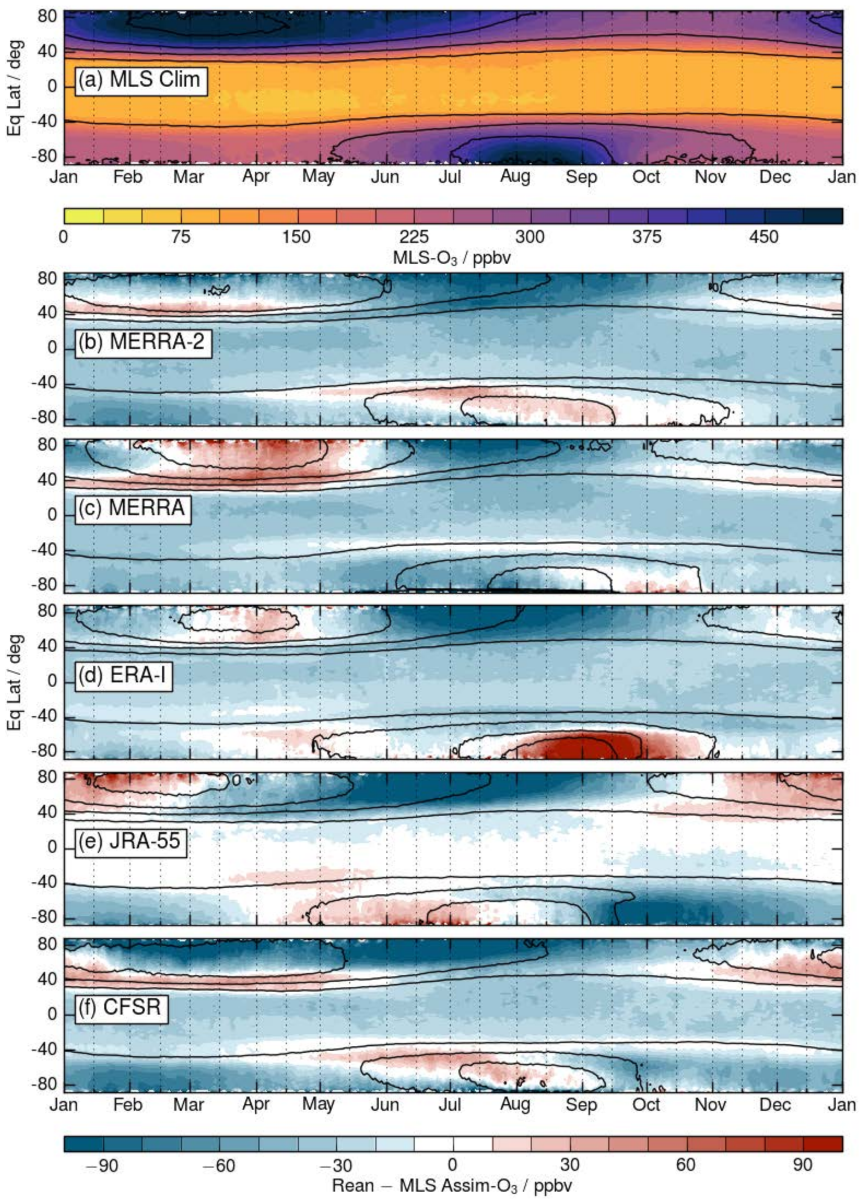


Figure 7.28: *Climatological (2005-2015) annual cycle of ozone as a function of EqL at 340 K, showing (top) v4.2 MLS ozone and (following rows) difference (reanalysis – MLS) of MLS from each of the MERRA-2, MERRA, ERA-Interim, JRA-55, and CFSR/CFSv2 reanalyses. The black overlays show the same selected ozone contours, from MLS on the top panel and each of the reanalyses on the following panels.*

MERRA-2, which assimilates MLS ozone at pressures below about 178hPa (261hPa in the last half of 2015, which is included in this record), shows much smaller differences from MLS than the other reanalyses at 390K and slightly smaller differences at 340K (generally below the level where MLS data are assimilated). At 390K, all of the reanalyses overestimate ozone in the SH spring (that is, they underestimate chemical loss); the differences are smallest for MERRA-2 and largest for ERA-Interim. Large differences are also seen in the NH subvortex, with MERRA and, to a lesser degree, CFSR/CFSv2 underestimating ozone and ERA-Interim overestimating it. All reanalyses tend to underestimate 390K ozone in the SH winter before extensive chemical ozone loss has occurred, though the magnitude of the underestimate in MERRA-2 is smaller. At 340K the reanalyses generally tend to underestimate MLS ozone except in the polar winter to spring – ERA-Interim substantially overestimates SH polar ozone (by over 20%) in August through October, and the other reanalyses overestimate polar winter/spring ozone by around 10%, with varying timing and EqL extent. Despite these differences, all of the reanalyses represent the seasonal cycle well qualitatively (that is, timing and

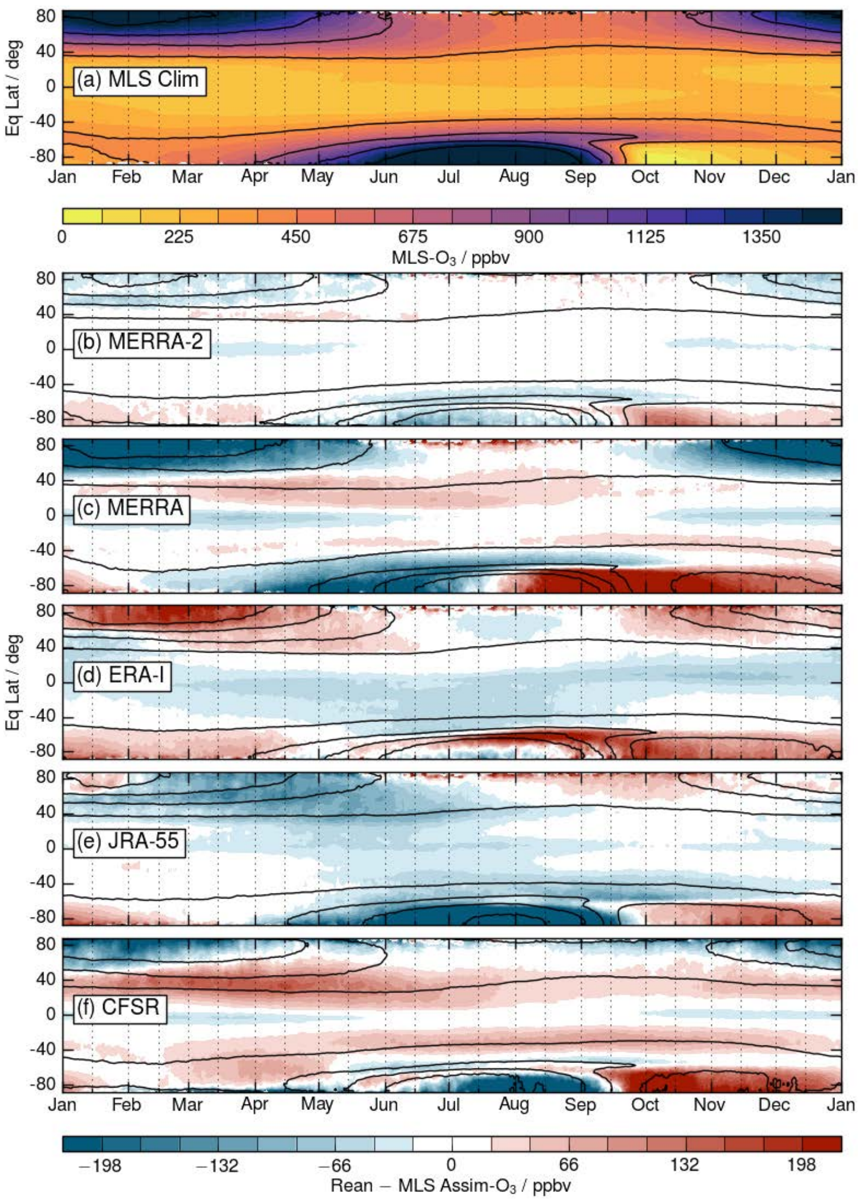


Figure 7.29: *As in* ***Fig. 7.28****, but at 390 K*

approximate magnitude) at both levels.

Figures 7.30 and **7.31** show the time series from 2005 through 2015 of daily MLS ozone as a function of EqL and the differences from the reanalyses. The general features noted in the climatology are also apparent here, and the reanalyses appear to do a reasonable job of capturing interannual variations in that the differences from MLS are not more extreme in extreme years. However, several discontinuities in the time series are worth noting: At the beginning of June 2015, MERRA-2 switched from assimilating v2 to v4 MLS ozone profile data, and the lowest level assimilated changed from about 178 hPa to 261 hPa (for May 2016 and thereafter, not shown in this analysis, this was changed to 215 hPa) (*Wargan et al.*, 2017). A small discontinuity is seen in the MERRA-2 / MLS differences at this time, with slightly better agreement at 340 K and overall slightly more negative differences at 390 K. ERA-Interim also shows several discontinuities related to changes in MLS data assimilated. MLS v2 data were assimilated in 2008, at pressure less than or equal to 215 hPa, and MLS NRT data were assimilated starting in mid-2009 (v2-NRT through 2012 and v3-NRT thereafter). MLS v2.2-NRT ozone data were very limited and were not suitable for scientific use at pressures above 68 hPa (*Lambert et al.*, 2008); during the period when these data were assimilated, the ERA-Interim fields show biases with v4 MLS data that are as large as or larger than those before any MLS data were assimilated. A marked improvement is seen when ERA-Interim began assimilating v3-NRT data, which speaks to the improvements in those retrievals (which allowed these data to be assimilated at pressures as high as 215 hPa), and the biases are similar to those in 2008 when operational MLS ozone data were assimilated down to the same presssure level.

7.6.3 Ozone in jet-relative coordinates

Figures 7.32 and **7.33** show climatological comparisons of assimilated ozone distributions in jet-relative coordinates (see, *e.g.*, *Manney et al.*, 2011) with Aura MLS ozone in the same coordinate system. The assimilated ozone is evaluated both as mapped directly from the native reanalysis grid to jet coordinates (right column in these figures), and as first interpolated (bi-linearly in the horizontal and linearly in time) to the MLS measurement locations and then mapped into jet coordinates (center column in these figures; the latter is restricted to ozone at the same geographic locations, so provides a more fair comparison). The differences in MLS data when mapped to jet coordinates using jet information from each of the reanalyses (left column in these figures) are relatively small (up to about 15 %), suggesting that, at least in the zonal mean climatological view, all of the reanalyses provide jet information that is appropriate for this mapping; this is consistent with the results of *Manney et al.* (2017; see *Section 7.4.2*) showing a consistent climatology of the upper tropospheric jets in all of the most recent reanalyses when analyzed at (or near) their native model resolutions. That the differences between MERRA and MERRA-2 in this mapping are much smaller than those between MERRA-2 and the other reanalyses suggests that the latter arise primarily from the differing grids/resolutions of the reanalyses (those of MERRA and

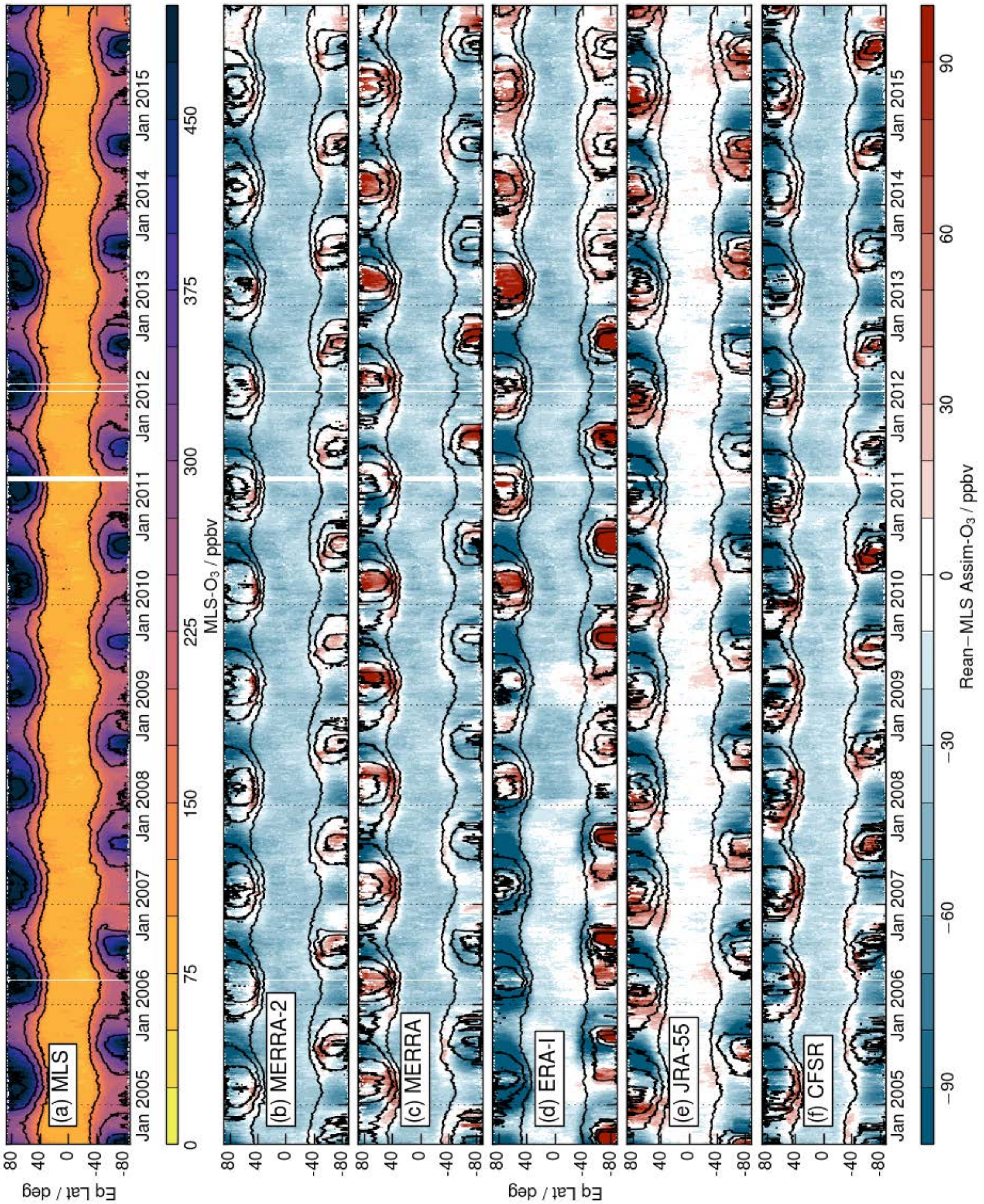


Figure 7.30: *(Top) Time series of 340 K MLS ozone for 2005 through 2015 and (following rows) the difference of each reanalysis from MLS. The black overlays show the same selected ozone contours, from MLS on the top panel and each of the reanalyses on the following panels.*

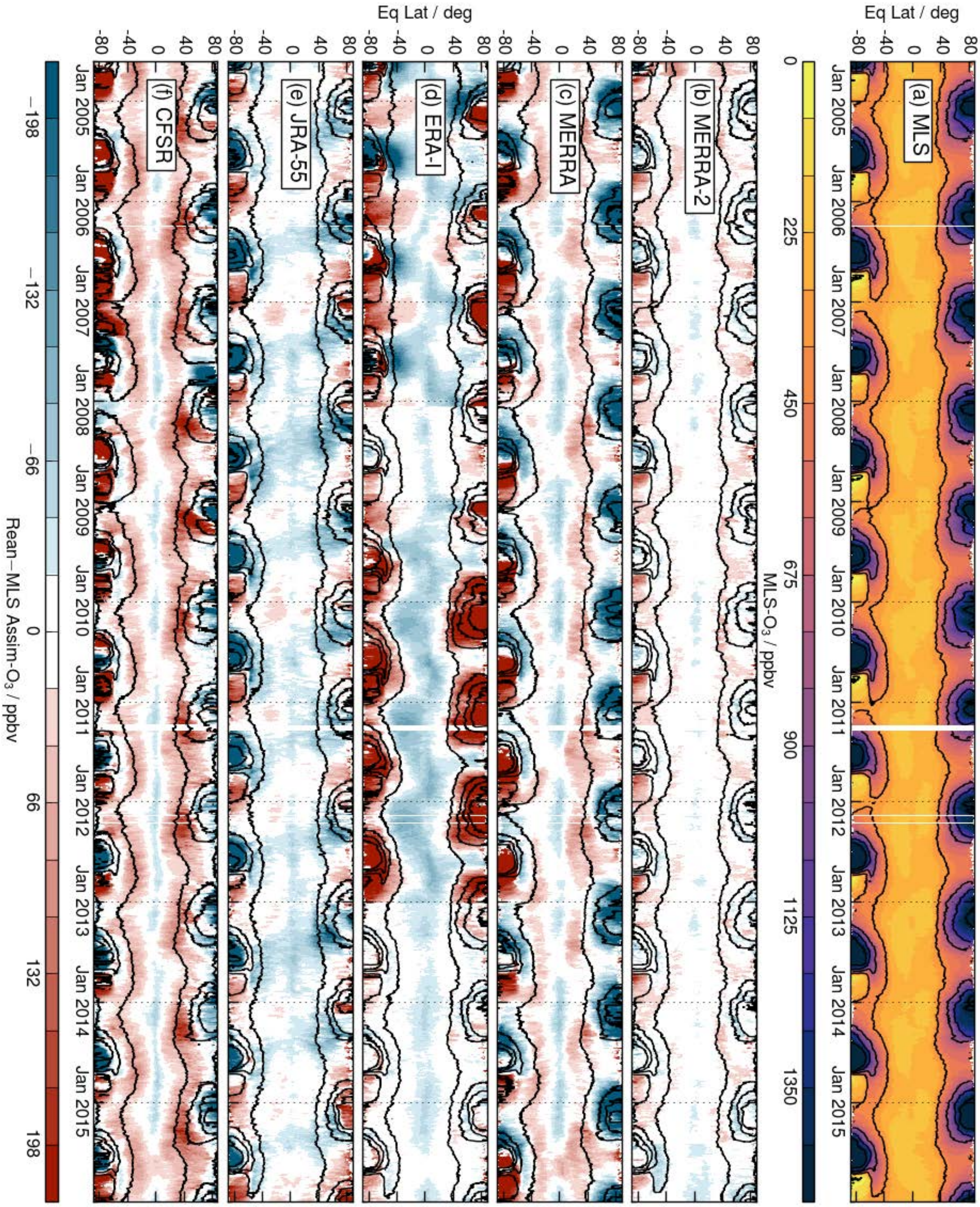


Figure 7.31: *As in **Figure 7.30** but at 390 K.*

MERRA-2 being the most similar in the horizontal and the same in the vertical).

Figure 7.32 shows reasonably good agreement between ozone in all of the reanalyses and MLS near and above the tropopause, with the sign of the differences varying in different reanalyses. Except in the SH in JRA-55, all of the reanalyses show lower ozone than MLS in the region below about 2 km below the tropopause. Since earlier versions of MLS data have shown high biases in this region (*e.g.*, *Hubert et al.*, 2016), it is unclear whether this is primarily due to MLS biases or whether the reanalyses may capture less stratosphere-to-troposphere transport than the MLS measurements indicate. The low bias in reanalysis ozone in the extratropical upper troposphere compared to MLS is consistent with that shown in zonal mean satellite data comparisons in *Chapter 4*. The reanalyses re-mapped from their native grids often, but not always, show similar patterns of differences from MLS data to those re-mapped after being interpolated to the MLS measurement locations. That these differences do sometimes show different qualitative patterns suggests that, even with the dense sampling of MLS data, the satellite sampling can be an important confounding factor in comparisons that are not based on geographically coincident data even when those data are mapped (as they are here) in coordinate systems that match dynamically similar air masses.

Similar results are seen for cross-sections in jet coordinates in other seasons. **Figure 7.33** shows the annual cycle in jet-coordinate MLS and reanalysis ozone as a climatological (1980 - 2015) slice as a function of latitude from the jet and time at the subtropical jet core altitude. The timing of the seasonal cycle is well-defined and agrees well with that from MLS in all of the reanalyses studied. At the level of the jet core (where the ozone fields are strongly

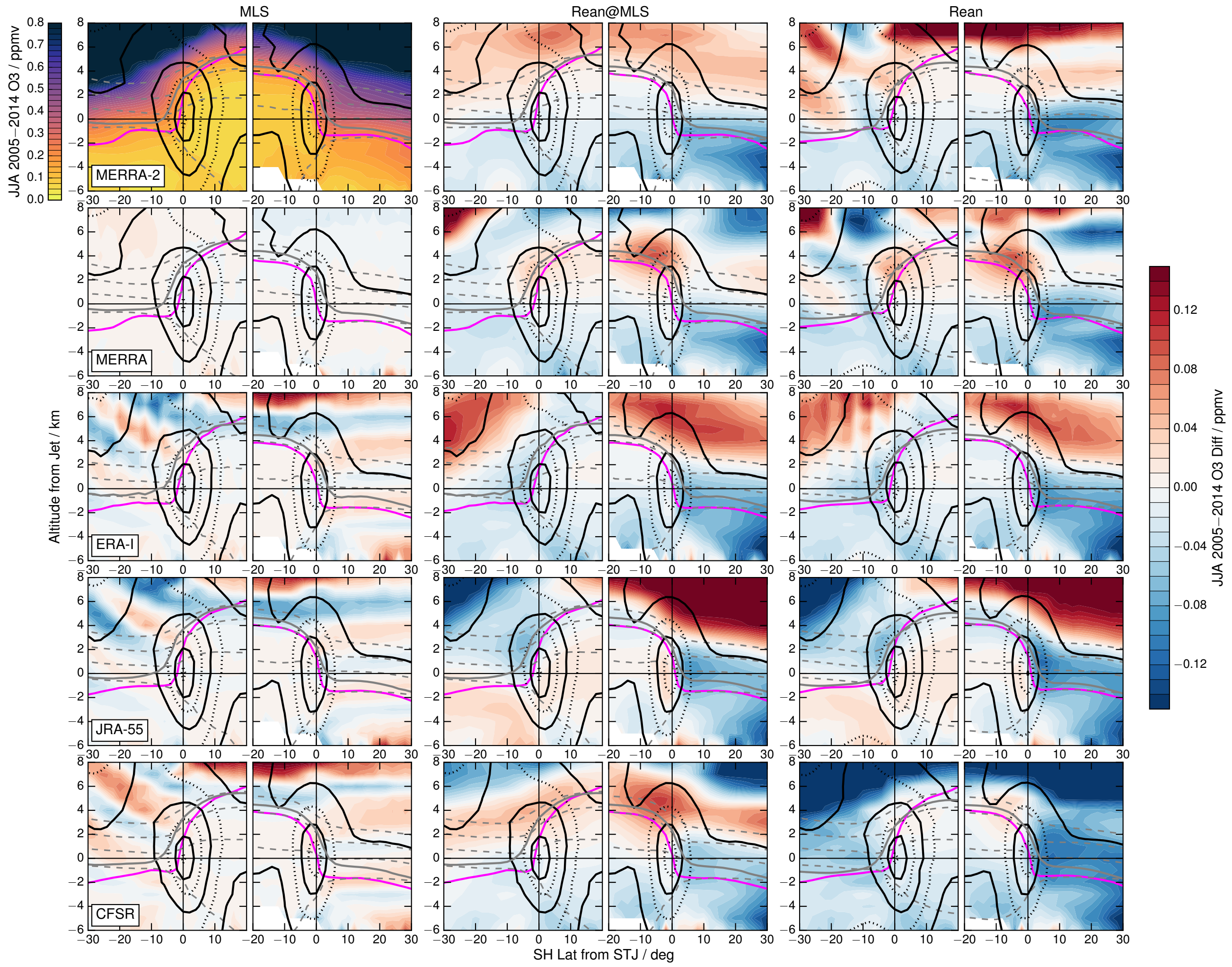


***Figure 7.32:** JJA mean climatological ozone in jet-relative coordinates, for 2005 through 2014. Top left plots shows MLS ozone mapped in jet coordinates using MERRA-2; the remainder of the left column shows the difference between that and MLS ozone mapped to jet coordinates with each of the other reanalyses. The center column shows the difference between each reanalyses' ozone mapped after interpolating to the MLS measurement locations and the MLS ozone mapped with that reanalyses; the right column is similar, except the reanalysis ozone is mapped into jet coordinates directly from its native grid. Overlays show: Windspeeds (black, from 10 to 80 by 10 m s-1, even values dotted), potential temperature (grey dashed, 330 to 390 by 20 K), the 3.5 PVU contour (magenta, negative in SH), and the LRT (grey solid).*

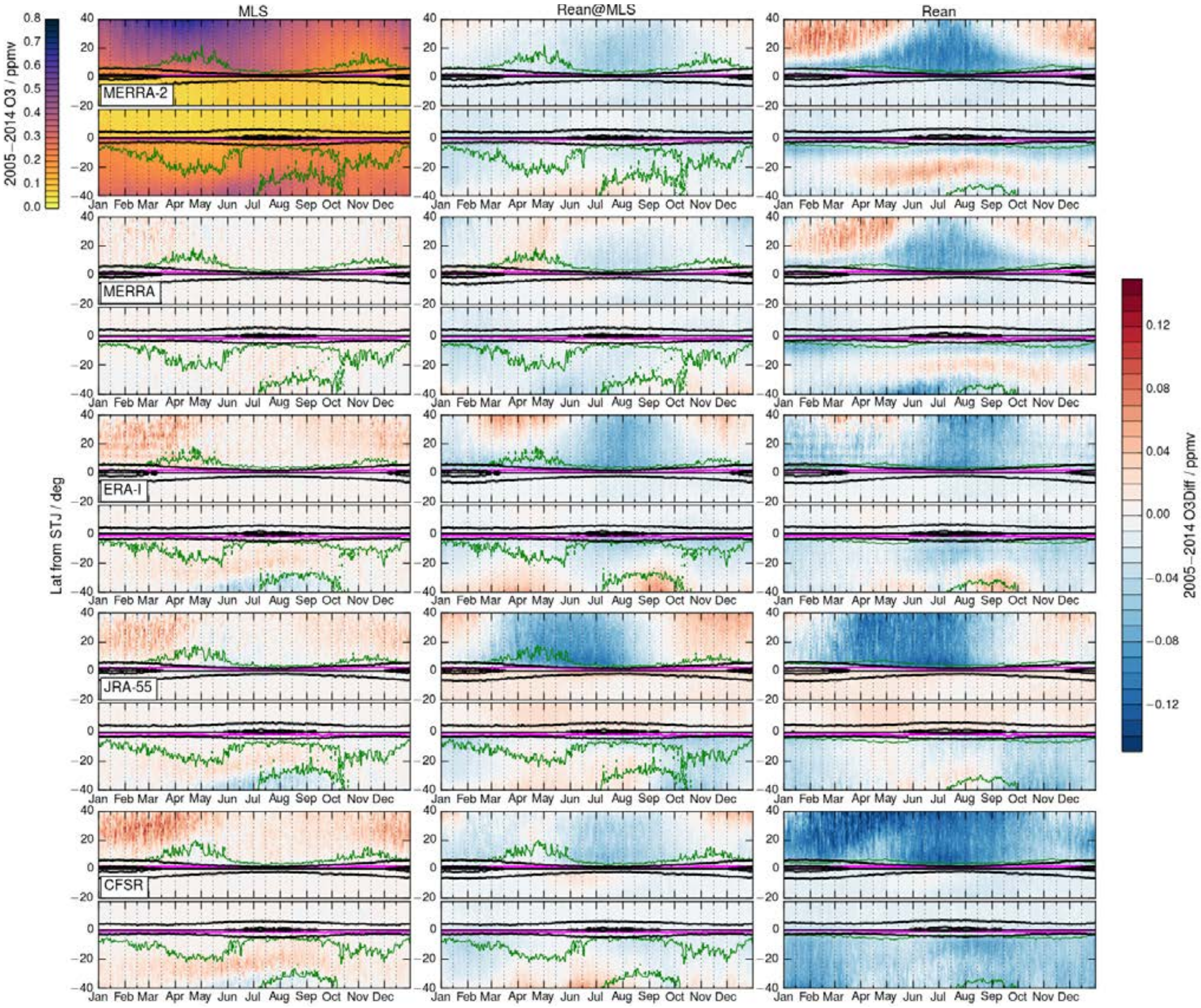


Figure 7.33: *Annual cycle in climatological (2005 - 2014) ozone at the subtropical jet core altitude as a function of time and latitude from the subtropical jet. Columns are as in* ***Figure 7.32****. Overlays show: Windspeeds (black, 40, 60, and 80 m s*$^{-1}$*), the 3.5 PVU contour (magenta, negative in SH), and the LRT (green).*

influenced by the subtropical jet), it is clear that interpolation to MLS locations before mapping into this dynamical coordinate is critical in providing a fair comparison, and thus that sampling effects are substantial. When comparing the reanalyses first interpolated to MLS measurement locations with MLS, the differences themselves show a seasonal cycle that varies among the reanalyses. Those differences are largest (up to about 20 % in the NH spring and summer) in JRA-55 and smallest (below 10 %) in MERRA-2; this is unsurprising since MERRA-2 assimilates MLS ozone and JRA-55 has the crudest ozone assimilation system of the reanalyses studies here.

7.6.4 Ozone mini-holes

Dynamical redistribution of ozone can produce large transient and localized reductions in total column ozone, also known as mini-holes (*e.g., Hood et al.*, 2001; *James,* 1998a,b; *Newman et al.*, 1988). *Millán and Manney* (2017) analyzed the representation of NH mini-hole events from several reanalyses (ERA-Interim, MERRA, MERRA-2, CFSR/CFSv2, and JRA-55) using data from OMI (*Levelt et al.*, 2006) and MLS (*Waters et al.*, 2006). OMI column ozone data allow us to compare their geographical representation while MLS ozone profile data allow us to study their vertical representation. Several definitions of mini-holes exist in the literature (*e.g.*, *Koch et al.*, 2005; *Hood et al.*, 2001; *James*, 1998a). Here, we define mini-hole events as regions where the total column ozone value is less than 25 % below the monthly mean. Further, we only consider as mini-hole events those ozone fluctuations with an area larger than 200,000 km^2.

Millán and Manney (2017) found that the reanalysis fields display the same mini-hole seasonal variability as OMI, with more mini-hole events during winter when the atmosphere is more dynamically active (see **Figure 7.34**). OMI and the reanalysis fields also display similar mini-hole

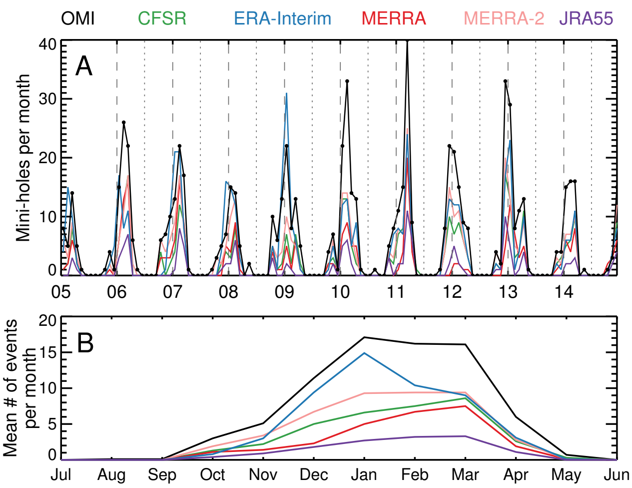


Figure 7.34: *(A) Mini-hole events per month during 2005 - 2014 in the Northern Hemisphere as found in OMI data and reanalysis fields (black, green, blue, red, pink, purple lines represent OMI, CFSR/CFSv2, ERA-Interim, MERRA, MERRA-2, and JRA-55 respectively). Dashed vertical lines indicate the beginning of each January, dotted vertical lines the beginning of each July. (B) Mean number of mini-hole events in a given month (during 2005 - 2014). (From Millan and Manney, 2017.)*

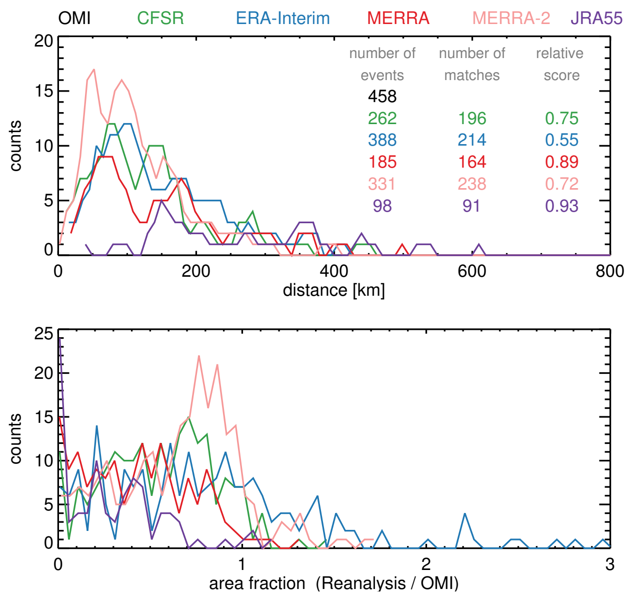


Figure 7.35: *(Top) Histograms of the distance between the mini-hole events found in the reanalysis fields and the ones found in OMI data (Black, green, blue, red, pink, purple lines represent OMI, CFSR, ERA-Interim, MERRA, MERRA-2, and JRA-55 respectively). Also shown is the total number of events as well as the number of matches between the events found in OMI and in the reanalyses. (Bottom) Histograms of the area fraction of mini-hole events. (From Millan and Manney, 2017.)*

geographical distributions, with mini-holes occurring most frequently over the North Atlantic storm tracks. All of the reanalyses studied underestimate the number of mini-hole events, with the underestimation ranging from 34 % less for ERA-Interim up to 83 % less for JRA-55. Further, reanalyses typically underestimate the area of the mini-hole events and most of the time are between 75 km and 300 km away from the events found in OMI (see **Figure 7.35**). Mini-holes found in CFSR/CFSv2, MERRA, MERRA-2 and ERA-Interim reanalyses display an eastward bias with respect to the events found in OMI data. JRA-55 does not show a consistent bias direction, a feature that is most likely related to their crude treatment of ozone (see *Chapters 2* and *4*).

The composite view of the vertical representation of mini-hole events agrees with previously reported mechanisms for dynamical mini-hole formation: Anticyclonic poleward Rossby wave breaking occurs in the UTLS; local uplift of air near the tropopause brings ozone poor air into the column and is accompanied by equatorward advection of polar air in the mid-stratosphere. On average, in the events found in both MLS and the reanalyses, the vertical structure in the reanalyses qualitatively agrees with that in MLS in that about two-thirds of the ozone reduction originates in the UTLS and the rest in the mid-stratosphere. Mini-hole regions do typically show more double tropopauses than in the surrounding air, but the association is not strong because double tropopauses occur most frequently above strong cyclonic circulation systems while mini-holes occur most frequently above anticyclonic systems.

7.7 Summary and recommendations

In this chapter, we evaluate an extensive set of diagnostics that are critical to understanding ExUTLS dynamical and transport processes, including the representation of the extratropical tropopause, UT jet streams, mixing and transport diagnostics, and ozone distributions and evolution. Because representing these processes requires high resolution, we focus on the recent full-input reanalyses, including MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2, and provide some comparisons that demonstrate just how important resolution is. The conventional input JRA-55C reanalysis was also compared for a few diagnostics. Earlier reanalysis (*e.g.*, NCEP-NCAR R1 and NCEP-NCAR R2, ERA-40) are not suitable for detailed UTLS studies because of their coarse resolution, especially in the vertical, and are not evaluated here. We find broadly consistent behavior among modern reanalyses in their representation of the extratropical tropopause, UT jet streams, and transport and mixing diagnostics. Larger differences are found in the representation of ozone in the ExUTLS, thought to be largely because of differences in the treatment of ozone among the reanalyses (*Davis et al.*, 2017) (also see *Chapter 4*). Our key finding and recommendations are given below:

Key findings:

- The reanalyses evaluated here agree well, with each other and with high-resolution radiosonde observations, on the location of the tropopause. CFSR/CFSv2 shows the smalest errors with respect to lapse rate tropopauses in radiosonde data of the analyses evaluated.

- Long-term trends (1981 - 2015) in tropopause altitude are in broad agreement both among the reanalyses and with observations, except for CFSR/CFSv2.

- The representation of multiple lapse-rate tropopause altitudes, which can be an indication of lateral STE events between the tropical UT andextra tropical lower stratosphere (ExLS), is highly dependent on the vertical grid resolution of reanalyses. CFSR/CFSv2 has the highest frequency of multiple tropopauses and the highest ExUTLS resolution of the renalyses evaluated here.

- Using pressure and model-level versions of CFSR/CFSv2, we have shown that the degraded vertical resolution in the pressure level fields makes them unsuitable for identifying tropopause locations, especially for multiple tropopause situations.

- JRA-55C was shown to be unsuitable for identifying multiple tropopauses because of its inability to qualitatively reproduce the distributions in SH high latitudes.

- Despite a general under-representation in multiple tropopause frequency compared to observations, most modern reanalyses reproduce the pattern and sign of observed long-term trends in multiple tropopause frequency.

- The reanalyses show good overall agreement in representation of the climatology of UT jets and of the subvortex jet in the lowermost stratosphere.

- Robust trends in UT jets (latitude, altitude, and windspeed) are limited to particular longitude regions and seasons. Disagreement among the reanalyses is most common for the SH jets; in particular, MERRA-2 and/or CFSR/CFSv2 sometimes differ from the other reanalyses even in the sign of the SH jet latitude trends.

- Lagrangian estimates of STE using full 3D kinematic winds are in broad agreement among the reanalyses, with some important differences in the locations and long-term changes of TST and STT. Transport estimates are sensitive to the choice of vertical coordinate (that is diabatic calculations in isentropic coordinates versus kinematic calculations in isobaric coordinates)

- Mixing diagnostics including effective diffusivity and PV gradients as a function of EqL show generally good agreement in climatological seasonal cycle and interannual variability.

- Mass flux across the 380 K isentropic surface agrees well between MERRA-2, ERA-Interim, and JRA-55, with CFSR/CFSv2 showing inconsistencies in the seasonal cycle.

- Climatological ozone distributions and seasonal cycles show good qualitative agreement; because of the large differences in the ozone products assimilated and the methods of assimilating them, this points to good representation of the dynamics in the UTLS, where ozone changes are primarily driven by dynamical and transport processes.

- Reanalysis ozone mapped in EqL generally reproduces at least qualitatively the interannual variability in MLS observed ozone, but ERA-Interim shows several step function changes that are related to changes in the versions of MLS ozone assimilated; in particular, in mid-2009 through 2012, large biases in ERA-Interim UTLS ozone arise from use of an early version of MLS NRT data.

Recommendations and future work:

- Based on previous work, and additional studies shown here, we recommend only the recent high-resolution reanalyses (MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 are such analyses evaluated herein) as suitable for ExUTLS dynamical and transport studies. The dynamical diagnostics derived from these reanalyses indicate that they are all suitable for use in such studies with some limitations.

- Given the inherent sensitivities of transport diagnostics to the method (*e.g.*, Lagrangian *vs.* Eulerian) and time period used, future reanalyses should incorporate tracers (*e.g.*, stratospheric mean age) for more direct transport comparisons.

- A few diagnostics (*e.g.*, effective diffusivity in CFSR/CFSv2; ozone in ERA-Interim) show substantial discontinuities when assessed over many years, and thus should be used with extreme caution and awareness in any analysis of those diagnostics.

- Despite the above point, further studies of mixing diagnostics (which cannot be compared with direct observations), including trends, comparisons with free-running models, and assessment in relation to trace gas observations, could provide useful information for model and data assimilation system improvement.

- Because many diagnostics of UTLS processes cannot be directly compared with data, using multiple reanalyses and assessing agreement among them should be an important part of ExUTLS studies.

- For diagnostics that cannot be directly compared with data, and in light of similar changes in input data, agreement among the reanalyses should be regarded as a necessary, but by no means a sufficient, condition for robustness of trends.

- As is the case for diagnostics described in other chapters (*e.g.*, *Chapter 10*), differences between the PV fields arising from differing products provided by the reanalysis centers add to the uncertainties in the evaluations. It would be helpful in the future for all reanalysis centers to provide PV on the model grids.

- The results from reanalyses assimilating MLS ozone (which has relatively high vertical resolution compared to other ozone profilers currently used) show promise for future improvements, and more attention to consistently assimilating high-resolution ozone observations in future reanalyses would be extremely beneficial to understanding the processes controlling ozone in this region, where it is of such great importance to the radiative balance.

- Future work is needed to better elucidate the role of various elements of model design in producing observed differences in tropopause location and characteristics (*e.g.*, through idealized simulations with the core models of each reanalysis).

- In the future, the accuracy of tropopause identifications in reanalyses should improve as the vertical grid spacing decreases. These diagnostics should be evaluated in forthcoming reanalyses (most immediately, in ERA5) and the impacts of these improvements on estimates of STE and their long-term changes should be explored.

- The accuracy of transport estimates from reanalyses is largely unknown, since global estimates of transport from observing systems are not available and the outcomes are sensitive to the input fields and methods used. Comparison of transport calculations using reanalysis wind fields and trace gas observations is one path to examine the accuracy of transport in reanalyses.

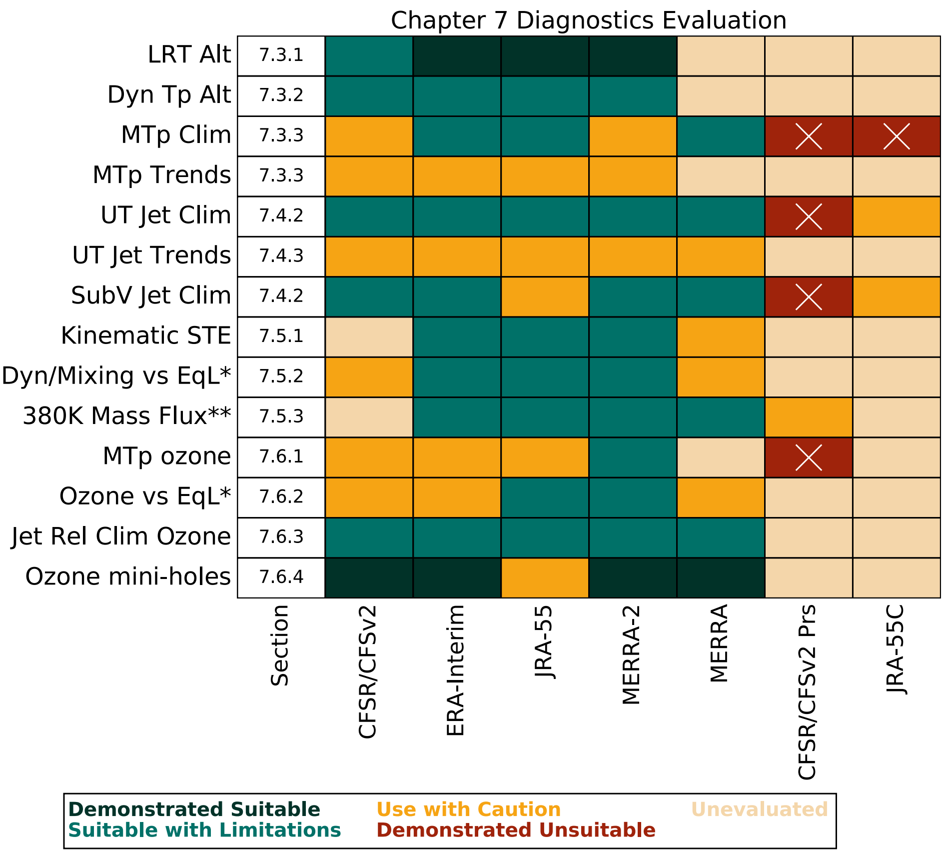
- Errors in transport calculations may also be gleaned from comparison of trajectory calculations driven by the reanalysis winds to long-duration balloon observations when available. However, such observations are infrequent and sometimes assimilated into the reanalyses, which limits their utility for validation studies.

- Given the known errors in trajectory and other transport calculations that arise from coarse temporal resolution of input wind fields (*e.g.*, *Stohl*, 1998; *Bowman et al.*, 2013), more frequent 3D wind field outputs are desired from future reanalyses. Such wind fields, which are already available for ERA5, will allow for improved understanding of transport and STE (*e.g.*, see early work using ERA5 in *Hoffmann et al.*, 2019).

- For studies of reanalysis ozone, several datasets are available for comparisons that have yet to be fully utilized; we recommend further comparisons with data from other satellite instruments (*e.g.*, the Odin OSIRIS and ACE-FTS instruments), ozone sondes, and both campaign and longer term aircraft datasets (*e.g.*, START-08, WISE, IAGOS). Some such studies will be done under the aegis of the SPARC Observed Composition Trends and Variability in the UTLS (OCTAV-UTLS) activity.

- Increased horizontal and vertical grid resolution will also be beneficial for reducing errors in transport calculations and enable analysis of processes at smaller scales.

Figure 7.36 summarizes the results for the main diagnostics evaluated herein. Overall, the latest generation of reanalyses shows good quality for representing UTLS dynamics and transport. Most of the diagnostics discussed herein cannot be verified with observations directly, and, while differences are generally relatively small, the agreement is rarely so good that we can say they are "demonstrated suitable" in cases where direct verification is not possible; hence most of the reanalyses are deemed "suitable with limitations" or "use with caution" for most diagnostics.



* Because the analysis as a function of EqL depends critically on PV (which is used to compute the EqL), those reanalyses where we have concerns about the PV fields are rated "use with caution" even in the absence of obvious "red flags". "CFSR/CFSv2 Prs" indicates CFSR/CFSv2 was used as interpolated to standard pressure levels; otherwise all diagnostics are calculated using model level data for all reanalyses except where specifically noted.

** The 380 K mass flux analysis was done using pressure level data for all reanalyses.

Figure 7.36: *Summary evaluation table for Chapter 7 diagnostics, per "key findings" highlighted above.*

Data availability

Access information for reanalysis datasets is given in *Chapter 2*.

The version 4.2 MLS ozone data are publicly available at: **https://disc.gsfc.nasa.gov/datasets?page=1&source=Aura%20 MLS&processingLevel=2**

JETPAC products at MLS locations used in the ozone analysis are publicly available at **https://mls.jpl.nasa.gov/**

The NWS radiosonde data used were retrieved from the Integrated Global Radiosonde Archive (IGRA; *Durre et al.*, 2016, **https://doi.org/10.7289/V5X63K0Q**).

Additional JETPAC products that are shown herein are not calculated operationally, but diagnostics that have been produced are available upon request.

Acknowledgements

We thank the reanalysis centres for their consistent help throughout the project in providing and helping us understand the reanalysis datasets. We thank the Aura Microwave Limb Sounder (MLS) science team for providing their data, and the MLS team at JPL, especially Brian Knosp and Ryan Fuller, for data management and analysis support. We thank Michelle Santee, Nathaniel Livesey, Kirstin Krüger, Susann Tegtmeier, Daniel Kunkel, and Peter Hoor for helpful discussions and comments during the project, and the reviewers and review editors for their careful reading and helpful comments.

Figures 7.3 and **7.10** are from *Xian and Homeyer* (2019); **Figures 7.4** and **7.5** are from *Millán et al.* (2021); **Figures 7.6, 7.7,** and **7.11 - 7.13** are from *Manney et al.* (2017); **Figures 7.16 - 7.18**, are from or adapted from *Boothe and Homeyer* (2017); **Figures 7.34** and **7.35** are from *Millán and Manney* (2017); all of these reproductions are made under a creative commons attribution 3.0 or 4.0 license **https://creativecommons.org/licenses/by/3.0/** or **https://creativecommons.org/licenses/by/4.0/**, respectively).

References

Abalos, M., B. Legras, B., and E. Shuckburgh, 2016: Interannual variability in effective diffusivity in the upper troposphere/lower stratosphere from reanalysis data. *Q. J. R. Meteorol. Soc.*, **142**, 1847 - 1861, doi: 10.1002/qj.2779.

Allen, D.R. and N. Nakamura, 2001: A seasonal climatology of effective diffusivity in the stratosphere. *J. Geophys. Res.*, **106**, 7917 - 7935, doi: 10.1029/2000JD900717.

Allen, D.R. and N. Nakamura, 2003: Tracer Equivalent Latitude: A Diagnostic tool for isentropic transport studies. *J. Atmos. Sci.*, **60**, 287 - 304, doi: 10.1175/1520-0469(2003)060<0287:TELADT>2.0.CO;2.

Appenzeller, C., J.R. Holton, and K.H. Rosenlof, 1996: Seasonal variation of mass transport across the tropopause. *J. Geophys. Res.*, **101**, 15,071 - 15,078, doi: 10.1029/96JD00821.

Berthet, G., J.G. Esler, and P.H. Haynes, 2007: A Lagrangian perspective of the tropopause and the ventilation of the lowermost stratosphere. *J. Geophys. Res.*, **112**, D18102, doi: 10.1029/2006JD008295.

Bethan, S., G. Vaughan, and S.J. Reid, 1996: A comparison of ozone and thermal tropopause heights and the impact of tropopause definition on quantifying the ozone content of the troposphere. *Q. J. R. Meteorol. Soc.*, **122**, 929 - 944, doi: 10.1002/qj.49712253207.

Birner, T., D. Sankey, and T.G. Shepherd, 2006: The tropopause inversion layer in models and analyses. *Geophys. Res. Lett.*, **33**, L14804, doi: 10.1029/2006GL026549.

Boothe, A.C. and C.R. Homeyer, 2017: Global large-scale stratosphere–troposphere exchange in modern reanalyses. *Atmos. Chem. Phys.*, **17**, 5537 - 5559, doi: 10.5194/acp-17-5537-2017.

Bowman, K.P., *et al.*, 2013: Input Data Requirements for Lagrangian Trajectory Models. *Bull. Amer. Meteorol. Soc.*, **94**, 1051 - 1058, doi: 10.1175/BAMS-D-12-00076.1.

Davis, S.M., *et al.*, 2017: Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP. *Atmos. Chem. Phys.*, **17**, 12743 - 12778, doi: 10.5194/acp-17-12743-2017.

Forster, P.M. and K.P. Shine, 1999: Stratospheric water vapour changes as a possible contributor to observed stratospheric cooling. *Geophys. Res. Lett.*, **26**, 3309 - 3312, doi: 10.1029/1999GL010487.

Fujiwara, M., *et al.*, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp-17-1417-2017.

Gettelman, A., *et al.*, 2011: The extratropical upper troposphere and lower stratosphere. *Rev. Geophys.*, **49**, RG3003, doi: 10.1029/2011RG000355.

Gettelman, A. and T. Wang, 2015: Structural diagnostics of the tropopause inversion layer and its evolution. *J. Geophys. Res.*, **120**, 46 - 62, doi: 10.1002/2014JD021846.

Grise, K.M., *et al.*, 2013: The ozone hole indirect effect: Cloud-radiative anomalies accompanying the poleward shift of the eddy-driven jet in the Southern Hemisphere. *Geophys. Res. Lett.,* **40**, 1 - 5, doi: 10.1002/grl.50675.

Harnik, N., C.I. Garfinkel, and O. Lachmy, 2016: The influence of jet stream regime on extreme weather events. In J. Li, R. Swinbank, R. Grotjahn, & H. Volkert (Eds.), *Dynamics and Predictability of Large-Scale, High-Impact Weather and Climate Events* (Special Publications of the International Union of Geodesy and Geophysics, pp. 79-94). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107775541.007.

Haynes, P. and E. Shuckburgh, 2000a: Effective diffusivity as a diagnostic of atmospheric transport 1. Stratosphere. *J. Geophys. Res.*, **105**, 22,777 - 22,794, doi: 10.1029/2000JD900093.

Haynes, P. and E. Shuckburgh, 2000b: Effective diffusivity as a diagnostic of atmospheric transport 2. Troposphere and lower stratosphere. *J. Geophys. Res.*, **105**, 22,795 - 22,810, doi: 10.1029/2000JD900092.

Hegglin, M.I. and T.G. Shepherd, 2009: Large climate-induced changes in ultraviolet index and stratosphere-to-troposphere ozone flux. *Nature Geosci.*, **2**, 687 - 691, doi: 10.1038/ngeo604.

Hegglin, M.I., *et al.*, 2006: Measurements of NO, NO_y, N_2O, and O_3 during SPURT: implications for transport and chemistry in the lowermost stratosphere. *Atmos. Chem. Phys.*, *6*, 1331 - 1350, doi: 10.5194/acp-6-1331-2006.

Hegglin, M.I., *et al.*, 2016: The GAW/NDACC/SPARC Upper Troposphere and Lower Stratosphere (UTLS) Observation Workshop, *SPARC Newsletter*, **47**, 39 - 43, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Hegglin, M.I. *et al.* 2010: Multi-model assessment of the upper troposphere and lower stratosphere: Extratropics. *J. Geophys. Res.*, **115**, D00M09, doi: 10.1029/2010JD013884.

Highwood, E.J., B.J. Hoskins, and P. Berrisford, 2000: Properties of the Arctic tropopause. *Q. J. R. Meteorol. Soc.*, **126**, 1515 - 1532, doi: 10.1002/qj.49712656515.

Hitchman, M.H. and A.S. Huesmann, 2007: A seasonal climatology of Rossby wave breaking in the 320 - 2000 - K layer. *J. Atmos. Sci.*, **64**, 1922 - 1940, doi: 10.1175/JAS3927.1.

Hoffmann, L., *et al.*, 2019: From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmos. Chem. Phys.*, **19**, 3097 - 3124, doi: 10.5194/acp-19-3097-2019.

Homeyer, C.R., 2014: Formation of the Enhanced-V Infrared Cloud Top Feature from High-Resolution Three-Dimensional Radar Observations. *J. Atmos. Sci.*, **71**, 332 - 348, doi: 10.1175/JAS-D-13-079.1.

Homeyer, C.R. and K.P. Bowman, 2013: Rossby wave breaking and transport between the tropics and extratropics above the subtropical jet. *J. Atmos. Sci.*, **70**, 607 - 626, doi: 10.1175/JAS-D-12-0198.1.

Homeyer, C., K.P. Bowman, and L.L. Pan, 2010: Extratropical tropopause transition layer characteristics from high-resolution sounding data. *J. Geophys. Res.*, **115**, D13108, doi: 10.1029/2009JD013664.

Homeyer, C.R., J.D. McAuliffe, and K.M. Bedka, 2017: On the Development of Above-Anvil Cirrus Plumes in Extratropical Convection. *J. Atmos. Sci.*, **74**, 1617 - 1633, doi: 10.1175/JAS-D-16-0269.1.

Homeyer, C.R., *et al.*, 2011: Dynamical and chemical characteristics of tropospheric intrusions observed during START08. *J. Geophys. Res.*, **116**, D06111, doi: 10.1029/2010JD015098.

Hood, L.L., B.E. Soukharev, M. Fromm, and J.P. McCormack, 2001: Origin of extreme ozone minima at middle to high northern latitudes. *J. Geophys. Res.: Atmos.*, **106**, 20 925 - 20 940, doi: 10.1029/2001JD900093.

Hoor, P., *et al.*, 2004: Seasonality and extent of extratropical TST derived from in-situ CO measurements during SPURT. *Atmos. Chem. Phys.*, **4**, 1427 - 1442.

Hubert, D., et al., 2016: Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records. *Atmos. Meas. Tech.*, **9**, 2497 - 2534, doi: 10.5194/amt-9-2497-2016.

Hudson, R.D., 2012: Measurements of the movement of the jet streams at mid-latitudes, in the Northern and Southern Hemispheres, 1979 to 2010. *Atmos. Chem. Phys.*, **12**, 7797 - 7808, doi: 10.5194/acp-12-7797-2012.

Jaeglé, L., R.Wood, and K. Wargan, 2017: Multiyear Composite View of Ozone Enhancements and Stratosphere-to-Troposphere Transport in Dry Intrusions of Northern Hemisphere Extratropical Cyclones. *J. Geophys. Res.*, **122**, 13,436 - 13,457, doi: 10.1002/2017JD027656.

James, P.M., 1998a: A climatology of ozone mini-holes over the northern hemisphere. *Int. J. Climatol.*, **18**, 1287 - 1303, doi: 10.1002/(SICI)1097-0088(1998100)18:12<1287::AID-JOC315>3.0.CO;2-4.

James, P.M., 1998b: An interhemispheric comparison of ozone mini-hole climatologies. *Geophys Res. Lett.*, **25**, 301 - 304, doi: 10.1029/97GL03643.

Knowland, K.E., L.E. Ott, B.N. Duncan, and K. Wargan, 2017: Stratospheric Intrusion-Influenced Ozone Air Quality Exceedances Investigated in the NASA MERRA-2 Reanalysis. *Geophys. Res. Lett.*, **44**, 10,691 - 10,701, doi: 10.1002/2017GL074532.

Koch, P., H. Wernli, and H.C. Davies, 2006.: An event-based jet-stream climatology and typology. *Int. J. Climatol.*, **26**, 283 - 301, 10.1002/joc.1255.

Koch, G., *et al.*, 2005: A composite study on the structure and formation of ozone miniholes and minihighs over central Europe. *Geophys. Res. Lett.*, **32**, L12810, doi: 10.1029/2004GL022062.

Kunkel, D., *et al.*, 2019: Evidence of small-scale quasi-isentropic mixing in ridges of extratropical baroclinic waves. *Atmos. Chem. Phys.*, **19**, 12 607 - 12 630, doi: 10.5194/acp-19-12607-2019.

Kunz, A., P. Konopka, R. Müller, and L.L. Pan, 2011b: Dynamical tropopause based on isentropic potential vorticity gradients. *J. Geophys. Res.*, **116**, D01110, doi: 10.1029/2010JD014343.

Kunz, A., M. Sprenger, and H. Wernli, 2015: Climatology of potential vorticity streamers and associated isentropic transport pathways across PV gradient barriers. *J. Geophys. Res. Atmos.*, **120**, 3802 - 3821, doi: 10.1002/2014JD022615.

Lacis, A., D.J. Wuebbles, J.A. and Logan, 1990: Radiative forcing of climate by changes in the vertical distribution of ozone. *J. Geophys. Res.*, **95**, 9971 - 9981, doi: 10.1029/JD095iD07p09971.

Lambert, A., *et al.*, 2008: Version 2.2 Level 2 near-real-time data user guide. *Tech. Rep. JPL* **D-48439**, Jet Propulsion Laboratory, 2008.

Lawrence, Z.D., G.L. Manney, and K. Wargan, 2018: Reanalysis intercomparisons of stratospheric polar processing diagnostics. *Atmos. Chem. Phys.*, **18**, 13 547 - 13 579, doi: 10.5194/acp-18-13547-2018.

Levelt, P.F., *et al.*, 2006: Science objectives of the ozone monitoring instrument. *IEEE T. Geosci. Remote* , **44**, 1199 - 1208, doi: 10.1109/TGRS.2006.872336.

Lin, M., *et al.*, 2014: Tropospheric ozone trends at Mauna Loa Observatory tied to decadal climate variability. *Nat. Geosci.*, **7**, 136 - 143, doi: 10.1038/ngeo2066.

Lin, M., *et al.*, 2015: Climate variability modulates western US ozone air quality in spring via deep stratospheric intrusions. *Nature Commun.*, **6**, doi: 10.1038/ncomms8105.

Livesey, N.J. et al., 2020: Version 4.2x Level 2 and 3 data quality and description document. Tech. Rep. JPL D-33509 Rev.Jet Propulsion Laboratory, available at **http://mls.jpl.nasa.gov.**

Long, C.S., *et al.*, 2017: Climatology and Interannual Variability of Dynamic Variables in Multiple Reanalyses Evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14,593 - 14,629, doi: 10.5194/acp-17-14593-2017.

Lorenz, D.J. and E.T. DeWeaver, 2007: Tropopause height and zonal wind response to global warming in the IPCC scenario integrations. *J. Geophys. Res.*, **112**, D10119, doi: 10.1029/2006JD008087.

Mahlman, J.D., 1997: Dynamics of Transport Processes in the Upper Troposphere. *Science*, **276**, 1079 - 1083, doi: 10.1126/science.276.5315.1079.

Mann, M.E., *et al.*, 2017: Influence of Anthropogenic Climate Change on Planetary Wave Resonance and Extreme Weather Events. *Sci. Rep.*, **7**, doi: 10.1038/srep45242.

Manney, G. L., Hegglin, M. I., & Lawrence, Z. D., 2021. Seasonal and Regional Signatures of ENSO in Upper Tropospheric Jet Characteristics from Reanalyses, Journal of Climate, 34(22), 9181-9200.

Manney, G.L. and Z.D. Lawrence, 2016: The major stratospheric final warming in 2016: dispersal of vortex air and termination of Arctic chemical ozone loss. Atmos. Chem. Phys., 16, 15371–15396, doi.org 10.5194/acp-16-15371-2016.

Manney, G.L. and M.I. Hegglin, 2018a: Seasonal and Regional Variations in Long-Term Changes in Upper Tropospheric Jets from Reanalyses. *J. Clim.*, **31**, 423 - 448, doi: 10.1175/JCLI-D-17-0303.1.

Manney, G.L. and M.I. Hegglin, 2018b: Corrigendum for "Seasonal and Regional Variations in Long- Term Changes in Upper Tropospheric Jets from Reanalyses". *J. Clim.*, **31**, 1289 - 1293, doi: 10.1175/JCLI-D-17-0881.1.

Manney, G.L., *et al.*, 2005: EOS Microwave Limb Sounder observations of the Antarctic polar vortex breakup in 2004. *Geophys. Res. Lett.*, **32**, L12811, doi: 10.1029/2005GL022823, 2005.

Manney, G. L. *et al.*, 2009: Satellite observations and modeling of transport in the upper troposphere through the lower mesosphere during the 2006 major stratospheric sudden warming. *Atmos. Chem. Phys.*, **9**, 4775 - 4795, doi: 10.5194/acp-9-4775-2009.

Manney, G.L., *et al.*, 2011: Jet characterization in the upper troposphere/lower stratosphere (UTLS): applications to climatology and transport studies. *Atmos. Chem. Phys.*, **11**, 6115 - 6137, doi: 10.5194/acp-11-6115-2011.

Manney, G.L., *et al.*, 2014: Climatology of Upper Tropospheric/Lower Stratospheric (UTLS) Jets and Tropopauses in MERRA. *J. Clim.*, **27**, 3248 - 3271, doi: 10.1175/JCLI-D-13-00243.1.

Manney, G.L., et al., 2017: Reanalysis comparisons of upper tropospheric/lower stratospheric jets and multiple tropopauses. *Atmos. Chem. Phys.,* **17**, 11541 - 11566, doi: 10.5194/acp-17-11541-2017.

Martineau, P., J.S. Wright, N. Zhu, and M. Fujiwara, 2018: Zonal-mean data set of global atmospheric reanalyses on pressure levels. *Earth Sys. Sci. Data*, **10**, 1925 - 1941, doi: 10.5194/essd-10-1925-2018.

McIntyre, M.E. and T.N. Palmer, 1983: Breaking planetary waves in the stratosphere, *Nature*, **305**, 593 - 600, doi: 10.1038/305593a0.

McLandress, C., *et al.*, 2011: Separating the dynamical effects of climate change and ozone depletion. Part II: Southern Hemisphere troposphere. *J. Clim.*, **24**, 1850 - 1868, doi: 10.1175/2010JCLI3958.1.

Millán, L. and G.L. Manney, 2017: An assessment of Ozone Mini-holes Representation in Reanalyses Over the Northern Hemisphere. *Atmos. Chem. Phys.*, **17**, 9277 - 9289, doi: 10.5194/acp-17-9277-2017.

Millán, L.F., G.L. Manney, and Z.D. Lawrence, 2021: Reanalysis intercomparison of potential vorticity and potential-vorticity-based diagnostics. *Atmos. Chem. Phys.*, **21**, 5355 - 5376, doi: 10.5194/acp-21-5355-2021.

Newman, P.A., L.R. Lait, and M.R. Schoeberl, 1988: The morphology and meteorology of southern hemisphere spring total ozone mini-holes. *Geophys. Res. Lett.*, **15**, 923 - 926, doi: 10.1029/GL015i008p00923.

Olsen, M.A., M.R. Schoeberl, and A.R. Douglass, 2004: Stratosphere-troposphere exchange of mass and ozone. *J. Geophys. Res.*, **109**, D24114, doi: 10.1029/2004JD005186.

Olsen, M.A., A.R. Douglass, and T.B. Kaplan, 2013: Variability of extratropical ozone stratosphere- troposphere exchange using Microwave Limb Sounder observations. *J. Geophys. Res.*, **118**, 1090 - 1099, doi: 10.1029/2012JD018465.

Olsen, M.A., K. Wargan, and S. Pawson, 2016: Tropospheric column ozone response to ENSO in GEOS-5 assimilation of OMI and MLS ozone data. *Atmos. Chem. Phys.*, **16**, 7091 - 7103, doi: 10.5194/acp-16-7091-2016.

Orbe, C., *et al.*, 2018: Large-scale tropospheric transport in the Chemistry–Climate Model Initiative (CCMI) simulations. *Atmos. Chem. Phys.*, **18**, 7217 - 7235, doi: 10.5194/acp-18-7217-2018.

Pan, L.L., *et al.*, 2009: Tropospheric intrusions associated with the secondary tropopause. *J. Geophys. Res.*, **114**, D10 302, doi: 10.1029/2008JD011374.

Pan, L.L. *et al.*, 2010: The Stratosphere-Troposphere Analysis of Regional Transport 2008 Experiment. B*ull. Am. Meteor. Soc.*, **91**, 327 - 342, doi: 10.1175/2009BAMS2865.1.

Pilch Kedzierski, R., L. Neef, and K. Matthes, 2016: Tropopause sharpening by data assimilation. *Geophys. Res. Lett.*, **43**, 8298 - 8305, doi: 10.1002/2016GL069936.

Prather, M.J., *et al.*, 2011: An atmospheric chemist in search of the tropopause. *J. Geophys. Res.*, **116**, D04306, doi: 10.1029/2010JD014939.

Randel, W.J., D.J. Seidel, and L. Pan, 2007: Observational characteristics of double tropopauses. *J. Geophys. Res.*, **112**, D07309, doi: 10.1029/2006JD007904.

Reutter, P., B. Škerlak, M. Sprenger, and H. Wernli, 2015: Stratosphere-troposphere exchange (STE) in the vicinity of North Atlantic cyclones. *Atmos. Chem. Phys.*, **15**, 10,939 - 10,953, doi: 10.5194/acp-15-10939-2015.

Riese, M., *et al.*, 2012: Impact of uncertainties in atmospheric mixing on simulated UTLS composition and related radiative effects. J. *Geophys. Res. Atmos.*, **117**, D16305 , doi: 10.1029/2012JD017751.

Santee, M.L., *et al.*, 2011: Trace gas evolution in the lowermost stratosphere from Aura Microwave Limb Sounder measurements. *J. Geophys. Res.*, **116**, D18306, doi: 10.1029/2011JD015590.

Schoeberl, M.R., 2004: Extratropical stratosphere-troposphere mass exchange. *J. Geophys. Res.*, **109**, D13303, doi: 10.1029/2004JD004525.

Schwartz, M.J., *et al.*, 2015: Climatology and variability of trace gases in extratropical multiple tropopause regions from MLS, HIRDLS and ACE-FTS measurements. *J. Geophys. Res.*, **120**, 843 - 867, doi: 10.1002/2014JD021964.

Seo, K. and K.P. Bowman, 2002: Lagrangian estimate of global stratosphere-troposphere mass exchange. *J. Geophys. Res.*, **107**, 4555, doi: 10.1029/2002JD002441.

Škerlak, B., M. Sprenger, and H. Wernli, 2014: A global climatology of stratosphere-troposphere exchange using the ERA-Interim data set from 1979 to 2011. *Atmos. Chem. Phys.*, **14**, 913 - 937, doi: 10.5194/acp-14-913-2014.

Škerlak, B., et al., 2015: Tropopause folds in ERA-Interim: Global climatology and relation to extreme weather events. *J. Geophys. Res. Atmos.*, **120**, 4860 - 4877, doi: 10.1002/2014JD022787.

Solomon, D.L., K.P. Bowman, and C.R. Homeyer, 2016: Tropopause-Penetrating Convection from Three-Dimensional Gridded NEXRAD Data. *J. Appl. Meteor. Climatol.*, **55**, 465 - 478, doi: 10.1175/JAMC-D-15-0190.1.

Solomon, S., *et al.*, 2010: Contributions of stratospheric water vapor to decadal changes in the rate of global warming. *Science*, **327**, 1219 - 1223, doi: 10.1126/science.1182488.

Song, J., C. Li, J. Pan, and W. Zhou, W., 2011: Climatology of anticyclonic and cyclonic Rossby wave breaking on the dynamical tropopause in the southern hemisphere. *J. Clim.*, **24**, 1239 - 1251, doi: 10.1175/2010JCLI3157.1, 2011.

Sprenger, M. and H. Wernli, H., 2003: A northern hemispheric climatology of cross-tropopause exchange for the ERA15 time period (1979-1993). *J. Geophys. Res.*, **108**, 8521, doi: 10.1029/2002JD002636.

Sprenger, M., M.C. Maspoli, and H. Wernli, 2003: Tropopause folds and cross-tropopause exchange: A global investigation based upon ECMWF analyses for the time period March 2000 to February 2001. *J. Geophys. Res.*, **108**, 8518, doi: 10.1029/2002JD002587.

Sprenger, M., H. Wernli, and M. Bourqui, 2007: Stratosphere-troposphere exchange and its relation to potential vorticity streamers and cutoffs near the extratropical tropopause. *J. Atmos. Sci.*, 64, 1587–1602, doi: 10.1175/JAS3911.1, 2007.

Stohl, A., 1998: Computation, accuracy and applications of trajectories—A review and bibliography. *Atmos. Environ.*, **32,** 947 - 966**,** doi: 10.1016/S1352-2310(97)00457-3.

Stohl, A.:, 2001 A 1-year Lagrangian "climatology" of airstreams in the northern hemisphere troposphere and lowermost stratosphere. *J. Geophys. Res.*, **106**, 7263 - 7279, doi: 10.1029/2000JD900570.

Wang, P.K., *et al.*, 2016a: The origin of the gullwing-shaped cirrus above an Argentinian thunderstorm as seen in CALIPSO images. J. *Geophys. Res. Atmos.*, **121**, 3729 - 3738, doi: 10.1002/2015JD024111.

Wang, R., *et al.*, 2016b: A mechanism to explain the variations of tropopause and tropopause inversion layer in the Arctic region during a sudden stratospheric warming in 2009. *J. Geophys. Res.*, **121**, 11 932 - 11 945, doi: 10.1002/2016JD024958.

Wargan, K. L. and Coy, 2016: Strengthening of the Tropopause Inversion Layer during the 2009 Sudden Stratospheric Warming: A MERRA-2 Study. *J. Atmos. Sci.*, **73**, 1871 - 1887, doi: 10.1175/JAS-D-15-0333.1.

Wargan, K., *et al.*, 2017: Evaluation of the Ozone Fields in NASA's MERRA-2 Reanalysis. *J. Clim.*, **30**, 2961 - 2988, doi: 10.1175/JCLI-D-16-0699.1.

Waters, J.W., et al., 2006: The Earth observing system microwave limb sounder (EOS MLS) on the aura Satellite. *IEEE T Geosci Remote*, **44**, 1075 - 1092, doi: 10.1109/TGRS.2006.873771.

Waugh, D.W., C.I. Garfinkel, and L.M. Polvani, 2015: Drivers of the Recent Tropical Expansion in the Southern Hemisphere: Changing SSTs or Ozone Depletion? *J. Clim.*, **28**, 6581 - 6586, doi: 10.1175/JCLI-D-15-0138.1.

Wernli, H. and M. Bourqui, M., 2002: A Lagrangian "1-year climatology" of (deep) cross-tropopause exchange in the extratropical northern hemisphere. *J. Geophys. Res.*, **107**, 4021, doi: 10.1029/2001JD000812.

WMO, 2011: Scientific assessment of ozone depletion: 2010, *Global Ozone Res. and Monit. Proj. Rep*. **52**, 516 pp., Geneva, Switzerland.

World Meteorological Organization: Meteorology, 1957: A three-dimensional science: Second session of the commission for aerology. *World Meteorol. Organ. Bull.*, **4**, 134 - 138, 1957.

Xian, T. and C.R. Homeyer, 2019: Global Tropopause Altitudes in Radiosondes and Reanalyses. *Atmos. Chem. Phys.*, **19**, 5661 - 5678, doi: 10.5194/acp-19-5661-2019.

Zeng, G., O. Morgenstern,P. Braesicke, and J.A. Pyle, 2010: Impact of stratospheric ozone recovery on tropospheric ozone and its budget. *Geophys. Res. Lett.*, **37**, L09805, doi: 10.1029/2010GL042812, 2010.

Major abbreviations and terms

ACE-FTS	
ANA	Analyzed, referring to MERRA and MERRA-2 products from the analysis step (see *Chapter 2*)
ASM	Assimilated, referring to MERRA and MERRA-2 products from assimilation step (see *Chapter 2*)
ATOVS	Advanced TIROS Operational Vertical Sounder
CFSR	Climate Forecast System Reanalysis of the NCEP
CFSv2	Climate Forecast System version 2
CONUS	CONtiguous United States
CTM	Chemical Transport Model
DJF	December/January/February
DOE	Department of Energy
DT	Double Tropopause
ECMWF	European Centre for Medium-range Weather Forecasts
ENSO	El Niño Southern Oscillation
Eq	Equivalent Latitude
ERA-40	ECMWF 40-year reanalysis
ERA-Interim	ECMWF interim reanalysis
ExUTLS	Extratropical Upper Troposphere and Lower Stratosphere
GMAO	Global Modeling and Assimilation Office
GNSS-RO	Global Navigation Satellite System - Radio Occultation
IAGOS	In-service Aircraft for a Global Observing System
JETPAC	JEt and Tropopause Products for Analysis and Characterization
JJA	June/July/August
JRA-55	Japanese 55-year Reanalysis
K_{eff}	Effective Diffusivity
LRT	Lapse-Rate Tropopause
LS	Lower Stratosphere
MAM	March/April/May
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2

MLS	Microwave Limb Sounder
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NCEP-NCAR DOE R2	Reanalysis 2 of the NCEP and DOE
NCEP-NCAR R1	Reanalysis 1 of the NCEP and NCAR
NASA	National Aeronautics and Space Administration
NH	Northern Hemisphere
NRT	Near Real Time
NWS	National Weather Service
OCTAV-UTLS	Observed Composition Trends And Variability in the Upper Troposphere/Lower Stratosphere
OMI	Ozone Monitoring Instrument
OSIRIS	Optical Spectrograph and InfraRed Imaging System
PJ	Polar Jet
PV	Potential Vorticity
PVU	PV units (defined as 10^{-6} K m^2 (kg s)$^{-1}$)
REM	Reanalysis Ensemble Mean
rms	root mean square
SH	Southern Hemisphere
SON	September/October/November
sPV	scaled Potential Vorticity
START-08	Stratosphere-Troposphere Analyses of Regional Transport 2008
STE	Stratosphere-Troposphere Exchange
STJ	SubTropical Jet
STT	Stratosphere-to-Troposphere Transport
TIROS	The Television Infrared Observation Satellite Program
TOVS-ATOVS	TIROS Operational Vertical Sounde
TST	Troposphere-to-Stratosphere Transport
TTL	Tropical Tropopause Layer
UT	Upper Troposphere
UTLS	Upper Troposphere and Lower Stratosphere
WISE	Wave-driven ISentropic Exchange
WMO	World Meteorological Organization

Chapter 8: Tropical Tropopause Layer

Chapter lead authors

Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Kirstin Krüger	Section for Meteorology and Oceanography, Department of Geosciences, University of Oslo	Norway

Co-authors

Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany
Nicholas A. Davis	Atmospheric Chemistry Observations & Modeling Lab, National Center for Atmospheric Research	USA
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Cameron R. Homeyer	School of Meteorology, University of Oklahoma	USA
Ioana Ivanciu	GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel	Germany
Young-Ha Kim	Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt *previously at: Severe Storm Research Center, Ewha Womans University, South Korea*	Germany
Bernard Legras	Laboratoire de Météorologie Dynamique, CNRS/ENS-PSL	France
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Eriko Nishimoto	NTT DATA INTELLILINK Corporation *previously at: Japan Agency for Marine Earth Science and Technology*	Japan
Matthias Nützel	Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen)	Germany
Robin Pilch Kedzierski	GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel	Germany
James S. Wang	Institute for Advanced Sustainability Studies	Germany
Tao Wang	Jet Propulsion Laboratory, California Institute of Technology	USA
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China

Abstract. This chapter evaluates the tropical transition region between the well-mixed, convective troposphere and the highly stratified stratosphere in the reanalyses. The general tropical tropopause layer structure, as given by the vertical temperature profile, tropopause levels, and the level of zero radiative heating, is analysed. Diagnostics related to clouds and convection in the tropical tropopause layer include cloud fraction, cloud water content, and outgoing longwave radiation. The chapter takes into account the diabatic heat budget as well as dynamical characteristics of the tropical tropopause layer such as Lagrangian cold points, residence times, and wave activity. Finally, the width of the tropical belt based on tropical and extra-tropical diagnostics and the representation of the South Asian Summer Monsoon in the reanalyses are evaluated.

Contents

8.1 Introduction 311
8.2 Temperature and tropopause characteristics 312
8.2.1 Observational data sets 312
8.2.2 Climatology 312
8.2.3 Interannual variability and long-term changes 315
8.2.4 Key findings and recommendations 317
8.3 Clouds and convection 318
8.3.1 Observational data sets 318
8.3.2 Spatial distribution of high clouds 320
8.3.3 Vertical profiles 321
8.3.4 Cloud radiative effects 323
8.3.5 Relationships with other variables 326
8.3.6 Temporal variability 328
8.3.7 Key findings and recommendations 330
8.4 Diabatic heating rates 331
8.4.1 Total diabatic heating 331
8.4.2 Radiative heating 332
8.4.3 Non-radiative heating 337
8.4.4 Key findings and recommendations 338
8.5 Transport 339
8.5.1 Dehydration point distribution 339
8.5.2 TTL residence time 341
8.5.3 TTL tropical upwelling 343
8.5.4 Key findings and recommendations 343
8.6 Wave activity 345
8.6.1 Horseshoe-shaped structure at the 100 hPa temperature 345
8.6.2 Equatorial waves 346
8.6.3 Key findings 348
8.7 Width of the TTL 348
8.7.1 Zonally-resolved subtropical jet diagnostic 349
8.7.2 Zonally-resolved tropopause break diagnostic 351
8.7.3 Zonal mean subtropical jet and tropopause break diagnostics 352
8.7.4 Key findings and recommendations 353
8.8 South Asian summer monsoon 353
8.8.1 Anticyclone: climatology and variability 354
8.8.2 Vertical velocity 359
8.8.3 Diabatic heating 361
8.8.4 Transport 363
8.8.5 Ozone 364
8.8.6 Regional analysis of clouds and radiative effects 366
8.8.7 Key findings and recommendations 371
8.9 Summary, key findings, and recommendations 372
References 376
Appendix A: Supplementary material 382
Major abbreviations and terms 388

8.1 Introduction

The tropical tropopause layer (TTL) is the transition region between the well mixed, convective troposphere and the radiatively controlled stratosphere. The vertical range of the TTL extends from the region of strong convective outflow near 12 - 14 km to highest altitudes influenced by convective overshooting and tropical tropospheric processes up to 18.5 km (**Fig. 8.1**; *Folkins et al.*, 1999; *Highwood and Hoskins*, 1998). Air masses in the TTL show dynamical and chemical properties of both the troposphere and the stratosphere and are controlled by numerous processes on a wide range of length- and time scales (*e.g.*, *Fueglistaler et al.*, 2009a). The complex interactions of circulation, convection, trace gases, clouds and radiation make the TTL a key player in radiative forcing and chemistry-climate coupling (*e.g.*, *Randel and Jensen*, 2013). Most important, the TTL is the main gateway for air entering the stratosphere. Therefore, stratospheric composition and chemistry, in particular of ozone, water vapour and aerosols, is strongly impacted by the composition of air near the tropical tropopause (*e.g.*, *Fueglistaler et al.*, 2011; *Holton and Gettelman*, 2001). The cold point in the inner tropics is of special importance for air masses on their way from the troposphere into the stratosphere, since it sets their stratospheric water vapor content (*e.g.*, *Randel et al.*, 2004; *Mote et al.*, 1996). Together with clouds, such as thin cirrus and convective anvils, water vapor in the TTL has a significant impact on the radiation and tropospheric climate. In general, the chemical and thermal boundary conditions of the TTL are determined by the interplay of rapid tropospheric convection, the stratospheric wave-driven circulation and exchange with mid-latitude air.

Reanalyses provide vertical and horizontal structures for temperature, geopotential height, wind, radiation budgets and cloud properties that are important for studies of atmospheric transport, dynamics and composition in the TTL. Many off-line chemistry-transport models and Lagrangian particle dispersion models are driven by reanalysis data (*e.g.*, *Schoeberl et al.*, 2012; *Krüger et al.*, 2009; *Chipperfield*, 1999). Their representation of the cold point determines how realistically such models simulate dehydration and the entrainment of trace gases or aerosols into the stratosphere. Process studies of TTL dynamics such as equatorial wave variability are also often based on the TTL temperature structure in reanalyses (*e.g.*, *Fujiwara et al.*, 2012). Finally, reanalysis cold point temperature and height have been used in the past for comparison to model results and for investigations of long-term changes (*e.g.*, *Gettelman et al.*, 2010). While many studies have highlighted the characteristics of individual reanalysis products, a comprehensive intercomparison of the TTL among all major atmospheric reanalyses is currently missing.

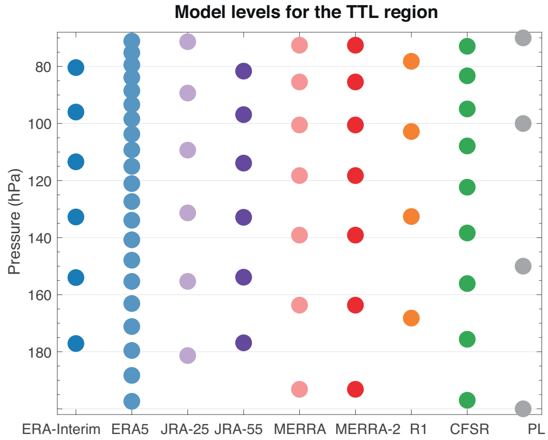


Figure 8.2: *Model-level pressure values for different reanalysis data sets in the TTL using a fixed surface pressure of 1013.25 hPa. Standard pressure levels (PL) in the TTL region are also shown. Adapted from Tegtmeier et al. (2020).*

Given the steep vertical gradient of atmospheric properties in the TTL, the vertical resolution of the reanalysis data is important. Reanalysis models resolve the TTL with different vertical resolutions, as illustrated in **Figure 8.2**. The number of model levels between 200 hPa and 70 hPa varies among the reanalyses from a low of 4 (NCEP-NCAR R1) to a high of 21 (ERA5), corresponding to vertical resolutions between ~1.5 km and ~0.2 km. In addition to the native model levels, all reanalyses provide post-processed data on fixed standard pressure levels with four levels situated between 200 and 70 hPa (**Fig. 8.2**). Detailed descriptions of the reanalysis data and their assimilated observations can be found in *Chapter 2* and *Fujiwara et al.* (2017). If not mentioned otherwise, the MERRA and MERRA-2 ASM products are used.

This chapter investigates whether reanalysis data reproduce the key characteristics of the TTL, including basic processes, such as circulation patterns, radiation and large-scale wave forcing, and their variability in space and time. The general TTL structure as given by the cold point and lapse rate tropopause and the vertical temperature profile is evaluated in *Section 8.2*. Diagnostics on clouds and convection in the TTL include cloud fraction profiles, outgoing longwave radiation, and cloud water content (*Section 8.3*).

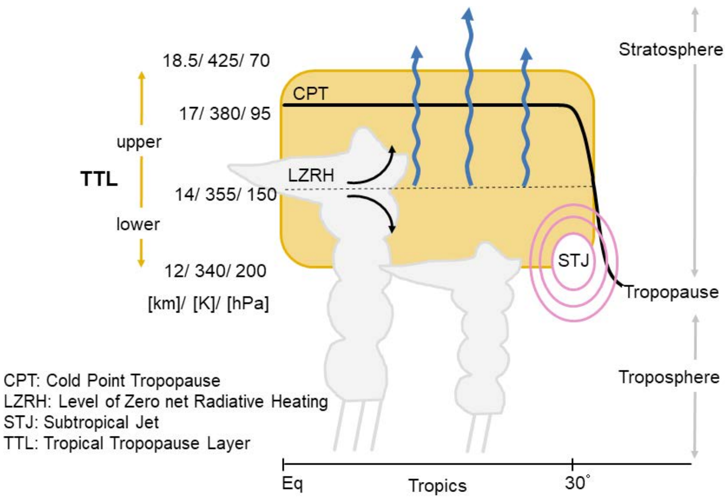


Figure 8.1: *Schematic of the Tropical Tropopause Layer (TTL).*

This chapter also takes into account the diabatic heating rates (*Section 8.4*) as well as dynamical characteristics of the TTL such as transport processes (*Section 8.5*), wave activity (*Section 8.6*), and long-term changes of the width of the TTL (*Section 8.7*). Analysis of the South Asian Summer Monsoon highlights spatial and temporal variations within the TTL (*Section 8.8*). Finally, *Chapter 8* is summarized in *Section 8.9*.

8.2 Temperature and tropopause characteristics

The tropopause is the most important physical boundary within the TTL, serving to separate the turbulent, moist troposphere from the stable, dry stratosphere. The position of the tropopause is diagnosed by the thermal properties of the TTL, as a negative, tropospheric vertical temperature gradient changes into a positive stratospheric temperature gradient. The role of the tropopause as a physical boundary is evident not only from the vertical temperature structure, but also from the distributions of atmospheric trace gases and clouds.

In the tropics, two definitions of the tropopause are widely used: one based on the cold point and one based on the characteristics of the lapse rate. The cold point tropopause is defined as the level at which the vertical temperature profile reaches its minimum (*Highwood and Hoskins*, 1998) and air parcels en route from the troposphere to the stratosphere encounter the lowest temperatures. Final dehydration typically occurs at these lowest temperatures, so that the cold point tropopause effectively controls the overall water vapour content of the lower stratosphere (*Randel et al.*, 2004) and explains its variability (*Fueglistaler et al.*, 2009a). While the cold point tropopause is an important boundary in the tropics where upwelling predominates, this definition of the tropopause is irrelevant for water vapor transport into the stratosphere at higher latitudes where net downwelling occurs. The lapse rate tropopause, on the other hand, offers a globally-applicable definition of the tropopause, defined as the lowest level at which the lapse rate decreases to $2\,\mathrm{K\,km^{-1}}$ or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed $2\,\mathrm{K\,km^{-1}}$ (*World Meteorological Organization*, 1957). The tropical lapse rate tropopause is typically ~0.5 km (~10 hPa) lower and ~1 K warmer than the cold point tropopause (*Seidel et al.*, 2001). In *Section 8.2*, we present a climatology of the tropical tropopause as derived from modern reanalysis data sets and compare it to data from high resolution measurements such as radiosondes or radio occultation. We also investigate temporal variability and long-term changes of TTL and cold point temperatures. All evaluations and further investigations can be found in *Tegtmeier et al.* (2020).

8.2.1 Observational data sets

High-resolution observations of the TTL are available from radiosonde stations in the tropics. However, climate records of radiosonde temperature, height and pressure data often suffer from inhomogeneities or time-varying biases due to changes in instruments or measurement practices (*Seidel and Randel*, 2006). Adjusted radiosonde temperature at 100 hPa, 70 hPa and corresponding trends at the cold point have been created by removing such inhomogeneities (*Wang et al.*, 2012, and references therein). In *Section 8.2*, we use several independently adjusted radiosonde data sets, including RATPAC (*Free et al.*, 2005), RAOBCORE (*Haimberger*, 2007) and HadAT (*Thorne et al.*, 2005) as well as the unadjusted, quality-controlled radiosonde data set IGRA (*Durre et al.*, 2006) covering the S-RIP core time period (1980 - 2010) (see *Chapter 1, Section 1.2*).

Since 2002, high-resolution temperature and pressure data in the TTL are also available from satellite retrievals based on the Global Navigation Satellite System – Radio Occultation (GNSS-RO) technique. Recent studies have demonstrated good agreement between GNSS-RO and radiosonde temperature profiles (*e.g.*, *Ho et al.*, 2017; *Anthes et al.*, 2008). In *Sections 8.2* and *8.8*, we use zonal mean as well as gridded (5°×5°) tropopause data sets constructed from GNSS-RO measurements collected by the Challenging Minisat Payload (CHAMP, *Wickert et al.*, 2001), Gravity Recovery and Climate Experiment (GRACE, *Beyerle et al.*, 2005), Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC, *Anthes et al.*, 2008), Metop-A (*von Engeln et al.*, 2011), Metop-B, Satélite de Aplicaciones Científicas-C/Scientific Application Satellite-C (SAC-C, *Hajj et al.*, 2004), and TerraSAR-X (*Beyerle et al.*, 2011) missions. All data are re-processed or post-processed occultation profiles with moisture information ('wetPrf' product) as provided by the COSMIC Data Analysis and Archive Center (CDAAC, **https://cdaac-www.cosmic.ucar.edu/cdaac/products.html**). Observational temperature records at reanalysis model levels in the TTL region have been determined by interpolating GNSS-RO temperature profiles with the barometric formula, taking into account the lapse rate between levels. For each profile, the cold point and lapse rate tropopause characteristics were identified based on the cold point and WMO criteria (*World Meteorological Organization*, 1957), respectively.

8.2.2 Climatology

Given the strong gradients of temperature and static stability in the TTL, the vertical resolution of the reanalysis data is an important factor in determining the cold point and lapse rate tropopause. For each reanalysis, tropopause heights and temperatures can be derived either from model- or pressure-level data. A comparison of the CFSR cold point tropopause based on model- and pressure-level temperature data is shown here to demonstrate the clear advantage of the finer model-level resolution (**Fig. 8.3**). The cold point tropopause from CFSR model-level data for the time period 2002 - 2010 agrees well with radio occultation results, with differences of less than 1.5 K and 0.2 km at all latitudes. The tropopause derived from CFSR pressure-level data, on the other hand, shows larger differences.

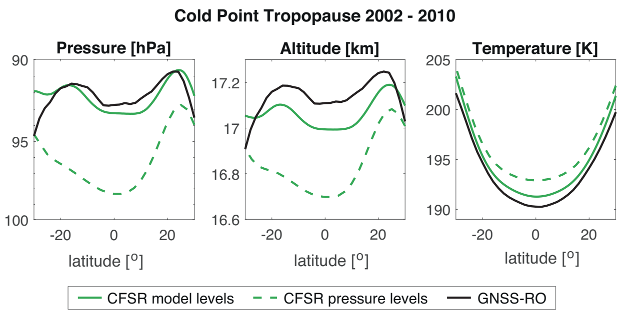


__Figure 8.3:__ Latitudinal distributions of annual mean, zonal mean cold point tropopause pressure (left), altitude (centre) and temperature (right) based on radio occultation data (black) and CFSR model-level (green solid) and pressure-level (green dashed) data during 2002–2010. Adapted from Tegtmeier et al. (2020).

This estimate is up to 0.4 km too low and up to 3 K too warm, illustrating the need to use data with high vertical resolution to identify and describe the tropopause. The following climatological tropopause comparisons are all based on model-level data.

Tropical mean temperatures from reanalyses at two standard pressure levels (100 hPa and 70 hPa) and at the two tropopause levels are compared to radio occultation data for the time period 2002 - 2010 (**Fig. 8.4**). At 100 hPa, reanalysis temperatures agree well with radio occultation observations with differences between - 0.35 K (too cold; ERA-Interim and ERA5) and 0.43 K (too warm; CFSR). At 70 hPa, the agreement is even better, with differences ranging from - 0.29 K (JRA-55) to 0.12 K (JRA-25). However, nearly all reanalyses show warm biases at both tropopause levels, with differences of up to 1.2 K compared to the observations. Most likely, the excess warmth of tropopause estimates based on reanalysis products stems from the limited vertical resolution of the reanalysis models in the TTL region. The best agreement is found for the reanalysis with the highest vertical resolution here (ERA5; 0.05 K too warm at the cold point tropopause). The reanalysis with the lowest vertical resolution (NCEP-NCAR R1) is 2.2 K too warm, outside the range displayed in **Figure 8.4**.

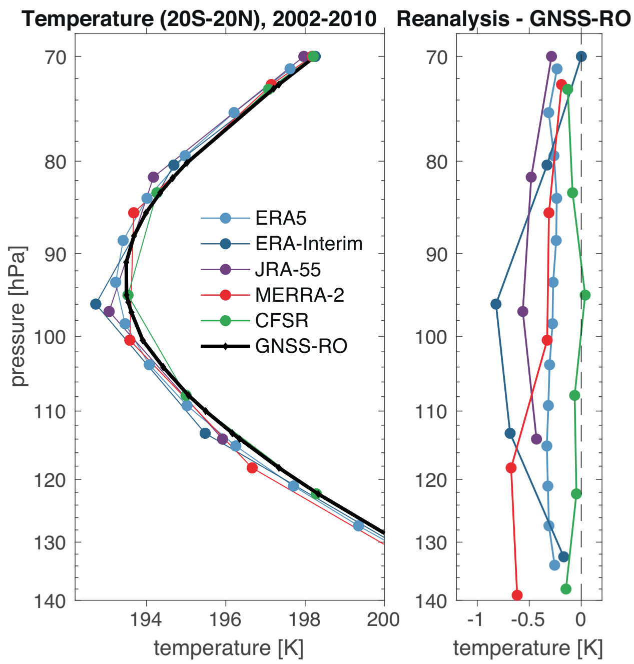


__Figure 8.5:__ Tropical mean (20°S–20°N), annual mean temperature profiles at reanalysis model levels between 140 and 70 hPa (left panel) during 2002–2010 and differences between reanalyses and GNSS-RO temperatures (right panel). Adapted from Tegtmeier et al. (2020).

Temperature profile comparisons between 140 hPa and 70 hPa at the native model level resolution have been conducted for the five most recent reanalyses (ERA5, ERA-Interim, JRA-55, MERRA-2, CFSR). All reanalyses tend to be colder than the observations in the tropical mean (**Fig. 8.5**), but differences are relatively small, and the agreement is good overall. CFSR and ERA5 agree best with the radio occultation data with mean biases of around - 0.06 K and - 0.28 K, respectively, averaged over the whole vertical range. ERA-Interim and MERRA-2 agree very well at upper levels but show relatively large deviations near 100 hPa (ERA-Interim; - 0.82 K) and below 110 hPa (MERRA-2; - 0.67 K), respectively. The evaluation demonstrates that temperature comparisons at standard pressure levels (**Fig. 8.4**) can be biased by up to 0.5 K, with CFSR showing a positive bias (0.45 K) at the 100 hPa standard pressure level but very good agreement (- 0.05 K) at nearby native model levels. Such biases can result from vertical interpolation of temperature data in regions with large lapse rate changes.

Comparing the temperature profiles to the tropopause values (**Figs. 8.4** and **8.5**) reveals that despite the five reanalyses having negative biases at model levels, they mostly have positive biases at the cold point and lapse rate tropopause levels.

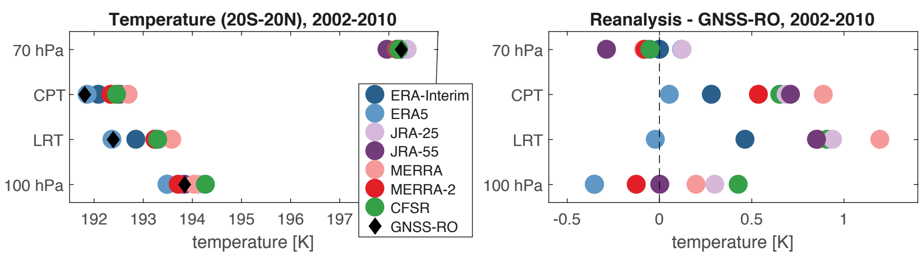


__Figure 8.4:__ Tropical mean (20° S - 20° N), annual mean temperatures at 100 hPa, the lapse rate tropopause (LRT), the cold point tropopause (CPT) and 70 hPa from reanalyses and GNSS-RO observations during 2002 - 2010 (left panel). Differences between the reanalysis and GNSS-RO temperatures are shown in the right panel. At 100 hPa, ERA-Interim is hidden by ERA-5; at the LRT, MERRA-2 is hidden by JRA-55; and at 70 hPa, ERA5 is hidden by JRA-25 and MERRA is hidden by MERRA-2. Adapted from Tegtmeier et al. (2020).

As the discrete values corresponding to reanalysis model levels are unable to reproduce the observed minimum temperature as recorded in a near-continuous profile, this difference is expected for the cold point tropopause. Similarly, the lapse rate tropopause criteria might typically be fulfilled at lower levels for data at coarser resolution, thus resulting in a warm bias at the lapse rate tropopause on average. Overall, our results indicate that the negative temperature bias at model levels is more than cancelled out by the positive bias introduced when calculating the cold point and lapse rate tropopauses. Linking the temperature profile and tropopause comparisons, this 'bias shift' is about 0.3 K for ERA5, 0.6 K for CFSR and 1 K or larger for ERA-Interim, MERRA-2 and JRA-55. In consequence, ERA5, with both a small negative bias at the model levels and a small bias shift provides the most realistic tropopause temperatures compared to GNSS-RO observations. CFSR also has a relatively small bias shift, but the mostly unbiased temperature profile does not permit any error cancelation via this shift, so that cold point and lapse rate tropopause levels based on CFSR are systematically too warm.

Agreement of the reanalysis temperature profiles from ERA5, ERA-Interim, MERRA-2, and CFSR with GNSS-RO data clearly improves for the comparison restricted to the 2007–2010 time period, when the more densely-sampled COSMIC data were assimilated (**Fig. A8.1** in *Appendix A*). Cold biases at model levels are accompanied by warm biases in the tropopause temperatures, which, for ERA-Interim and ERA5, increase after 2007. Here, the advantage of a reduced temperature bias at model levels comes at the expense of an increased temperature bias at the tropopause.

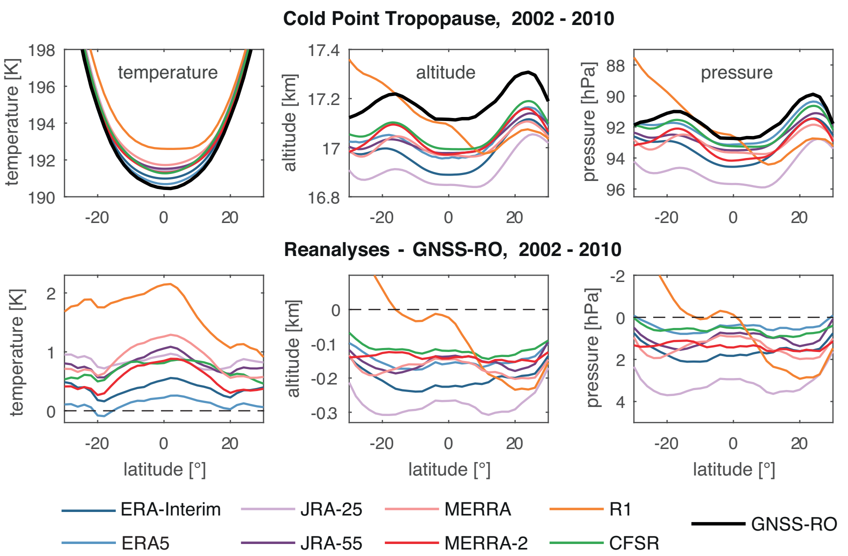


Figure 8.6: *Latitudinal distributions of zonal-mean cold point tropopause temperature (left), altitude (centre) and pressure (right) based on radio occultation data and reanalysis products during 2002 - 2010 (upper row) derived from model level data. Differences between reanalysis and radio occultation estimates are shown in the lower row. Adapted from Tegtmeier et al. (2020).*

Evaluations of the latitudinal structure of the cold point tropopause for 2002 - 2010 are based on comparisons to radio occultation data (**Fig. 8.6**). All reanalysis data produce tropopause levels that are too low and too warm, with the latter related to vertical resolution as explained above. The observations show that average cold point temperatures are lowest right around the equator. The reanalyses fail to reproduce this latitudinal gradient, indicating more constant cold point temperatures across the inner tropics between 10 ° S and 10 ° N with a less pronounced minimum at the equator. As a consequence, the largest differences in cold point tropopause temperatures relative to GNSS-RO data are at the equator and the best agreement is around 20 ° S/20 ° N for all reanalysis data.

The cold point altitude and pressure exhibit little north–south variability, ranging from 16.9 km (94 hPa) to 17.2 km (91.8 hPa). The lowest cold point temperatures are located near the equator, while the highest cold point altitudes are located around 20 ° S/20 ° N due to zonally-variable tropospheric pressure regimes, such as particularly low tropopause pressures over the Tibetan plateau during boreal summer (*Kim and Son*, 2012). The reanalysis data capture most of this latitudinal structure, showing roughly constant differences between about 0.1 km and 0.2 km (0 - 2 hPa, **Fig. 8.6**). The largest differences are found for NCEP-NCAR R1 in the SH, where the cold point tropopause based on *R1* is both higher and warmer than observed. The best agreement with respect to cold point temperatures is found for ERA5 and ERA-Interim, which are around 0.2 K and 0.4 K warmer than the radio occultation data, respectively. All other reanalysis data are in close agreement with each other, with differences from the observations of between 0.5 K and 1 K. The altitude and pressure of the cold point tropopause are captured best by ERA5, CFSR, MERRA, MERRA-2 and JRA-55, which all produce cold point tropopauses that are slightly too low (~0.1 km). ERA-Interim, despite very good agreement in cold point temperature, shows slightly larger biases in cold point altitude (~0.2 km) relative to the GNSS-RO benchmark.

Differences between reanalyses (ERA-Interim, MERRA-2, JRA55, and CFSR) and observations are largest in the inner tropics over central Africa, reaching values 50 % to 100 % greater than the zonal mean differences (**Fig. A8.2** in *Appendix A*). This region is characterized by a local cold point minimum that results from deep convection and its interaction with equatorial waves.

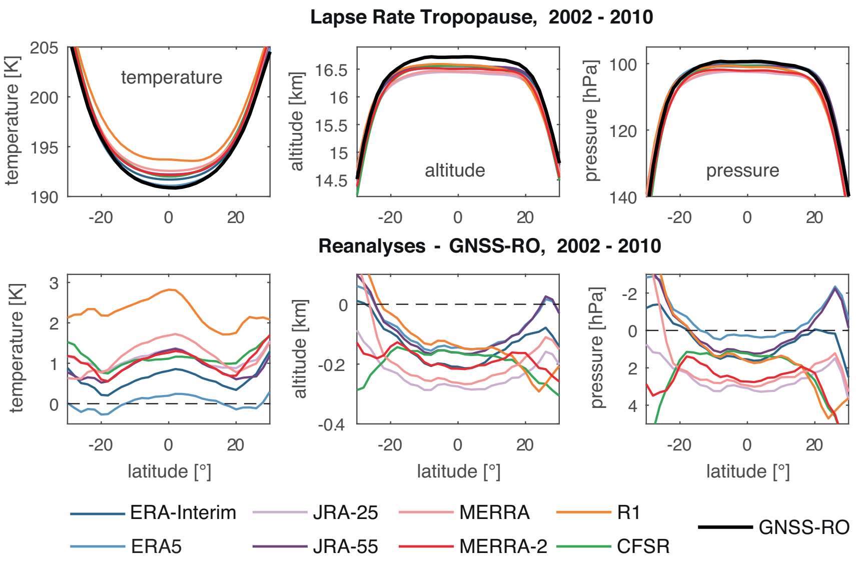


Figure 8.7: *Latitudinal distributions of zonal-mean lapse rate tropopause temperature (left), altitude (centre) and pressure (right) based on radio occultation data and reanalysis products during 2002 - 2010 (upper row) derived from model level data. Differences between reanalyses and radio occultation data estimates are shown in the lower row. Adapted from Tegtmeier et al. (2020).*

One possible explanation for the bias distribution might link the enhanced temperature differences to Kelvin wave activity that maximizes over Central Africa but is weaker over the West Pacific (*Kim et al.*, 2019). For most reanalyses, differences to GNSS-RO over Central Africa are 50 % higher for periods with enhanced wave activity (see CFSR in **Fig. A8.3** of *Appendix A*). *Section 8.8.1* highlights more tropopause analyses for the South Asian Summer Monsoon region and season.

The zonal mean lapse rate tropopause (**Fig. 8.7**) at the equator is found at similar temperatures and heights as the cold point tropopause, being only slightly warmer and lower consistent with *Seidel et al.* (2001). Poleward of 10 ° S/10 ° N, however, the lapse rate tropopause height decreases considerably faster than the cold point height, since here the cold point is more often located at the top of the inversion layer while the lapse rate tropopause is located at the bottom of the inversion layer (*Seidel et al.*, 2001). Lapse rate tropopause temperatures based on reanalysis data are on average about 0.2 K to 1.5 K too warm when compared to radio occultation data (see also **Fig. 8.4** and associated discussion) with best agreement for ERA5 and ERA-Interim. Consistent with this temperature bias, lapse rate tropopause levels based on reanalysis data are about 0.2 km to 0.4 km lower than those based on radio occultation data. The latitudinal structure of lapse rate tropopause temperatures reveals slightly larger biases at the equator and better agreement between 10 ° - 20 ° in each hemisphere, and is generally very similar to the latitudinal distribution of biases in cold point temperatures (**Fig. 8.6**). The altitude of the lapse rate tropopause shows considerable zonal variability, ranging from 14.5 km to 16.7 km. All reanalyses capture the plateau in lapse rate tropopause altitudes between 20 ° S and 20 ° N and the steep gradients in these altitudes on the poleward edges of the tropics.

8.2.3 Interannual variability and long-term changes

The interannual variability of TTL temperatures is strongly affected by both tropospheric (*e.g.*, ENSO) and stratospheric (*e.g.*, QBO, solar, volcanic) variability (*Krüger et al.*, 2008; *Zhou et al.*, 2001; *Randel et al.*, 2000). Time series of 70 hPa temperature anomalies and cold point temperature, pressure and altitude anomalies deseasonalized with respect to the common time period 2002 - 2010 are shown in **Figure 8.8**. The performance of the reanalyses with respect to both the spread among reanalyses and their agreement with observations is much better at the 70 hPa level than at the cold point level. Here, mostly the older reanalyses NCEP-NCAR R1 and JRA-25 show larger deviations when compared to the RAOBCORE radiosonde data. The interannual variability at 70 hPa is dominated by the stratospheric QBO signal, which is reproduced by all reanalyses datasets (see *Chapter 9* for a detailed analysis of the QBO signal). Positive temperature anomalies in response to the eruptions of El Chichón in 1982, and Mount Pinatubo in 1991 can be detected for all reanalysis data consistent with results of *Fujiwara et al.* (2015). In addition to the known signals such as the QBO- and ENSO-driven variations, the time series of tropical zonal mean temperatures shows some inherent variations representing the internal dynamical variability of the troposphere-stratosphere system (*Randel and Wu*, 2015).

The level of agreement among the reanalyses and between reanalyses and observations improves over time, with a step-like improvement around 1998 - 1999 that is likely associated with the TOVS-to-ATOVS transition. The higher vertical resolution of measurements from the ATOVS suite (see, *e.g.*, **Figure 7** in *Fujiwara et al.*, 2017; and **Figure 2.16** of *Chapter 2*) is known to reduce differences among the reanalysis with respect to stratospheric temperature (*Chapter 3*; *Long et al.*, 2017) and polar diagnostics (*Lawrence et al.*, 2018). Within the TTL, temperature biases improve from values of 1 - 2 K to around 0.5 K following the TOVS-to-ATOVS transition. This agreement improves further after 2002, when many of the more recent reanalyses started assimilating AIRS and GNSS-RO data (see, *e.g.*, **Figure 8** in *Fujiwara et al.*, 2017).

At the cold point, NCEP-NCAR R1 is a clear outlier, with much higher temperature anomalies than any other reanalyses during the period prior to 2005 (**Fig. 8.8**). However, differences among the more recent reanalyses are also relatively large, with ERA-Interim (on the lower side) and CFSR (on the upper side) showing differences as large as 2 K in the early years of the comparison.

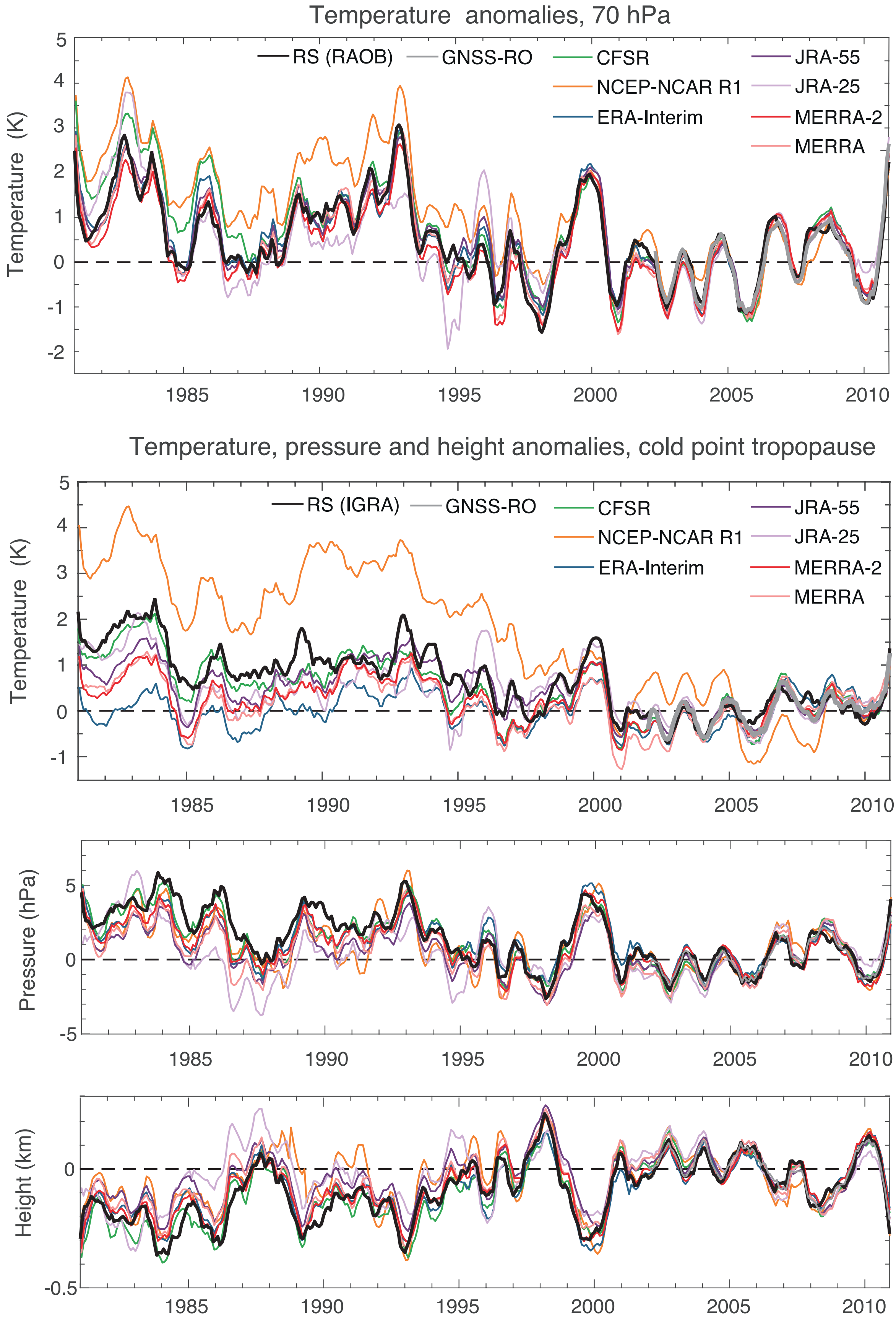


Figure 8.8: *Anomaly time series of 20° S - 20° N, 70 hPa temperature and cold point temperature, pressure and altitude with respect to the reference time period 2002 - 2010 for reanalyses, radiosonde (RAOBCORE and IGRA) and radio occultation data are shown. Adapted from Tegtmeier et al. (2020).*

Given that existing homogenized radiosonde data sets also show deviations of up to 1.5 K at this level (**Figure 2** in *Wang et al.*, 2012), we cannot deduce which reanalysis data set is most realistic. Note that the radiosonde time series from IGRA shown here should not be used for evaluating long-term changes, but only for assessing the representation of interannual variability. Periods of particularly pronounced interannual variability alternate with relatively quiescent ones. The QBO temperature signal at the cold point is weaker than at 70 hPa but still well captured by all of the reanalysis data except for NCEP-NCAR R1 (see *Chapter 9*).

Interannual variability in cold point pressure and altitude (**Fig. 8.8**) shows better agreement among the data sets than that in cold point or 70 hPa temperature. During the first 15 years of the record, the reanalysis cold point tropopause levels are mostly shifted toward higher altitudes and lower pressures, consistent with lower temperatures during this period. Anomalies in cold point temperature are in most cases matched by anomalies in cold point pressure and altitude, with a higher cold-point temperature (*e.g.*, around 1999 - 2000) corresponding to lower tropopause (negative altitude anomaly and positive pressure anomaly) and vice versa. The older reanalyses NCEP-NCAR R1 and JRA-25 again show the largest overall differences. The agreement improves over time, with the most consistent results found for the period after 2002.

Long-term temperature changes are evaluated over the 1979 - 2005 time period due to the availability of adjusted tropopause temperature trends from radiosonde data sets (see *Wang et al.*, 2012 for details). Both radiosonde records suggest significant cooling at the 70 hPa level (**Fig. 8.9**). Temperature trends based on the reanalysis data span almost exactly the same range (- 0.5 to - 1.1 K/decade) as those based on the radiosonde data sets (- 0.5 to - 1 K/decade). All reanalysis- and observationally-based trends are significant at this level, confirming the stratospheric cooling reported by previous studies (*e.g.*, *Randel et al.*, 2009). Satellite data from the Microwave Sounding Unit channel 4 (~13 - 22 km) suggests smaller trends of around - 0.25 K/decade over 1979 - 2005 (*Maycock et al.*, 2018) or - 0.4 K/decade over 1979 - 2009 (*Emanuel et al.*, 2013). However, the much broader altitude range of this MSU channel includes both stratospheric and tropospheric levels, which impedes a direct comparison with trends at 70 hPa.

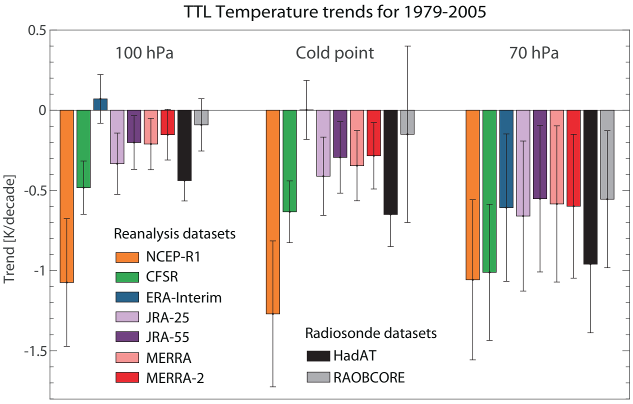


Figure 8.9: *Linear trends of tropical temperature (K/decade) for reanalyses and adjusted radiosonde data at the cold point, 100 hPa and 70 hPa with ± 2σ error bars.*

At the 100 hPa and cold point levels, the situation is completely different. The available adjusted radiosonde data sets show in some cases uncertainties larger than the respective temperature trends at these levels. Only a few of the available data sets indicate a statistically significant cooling based on a methodology that adjusts the cold point trend to account for nearby fixed pressure-level data and day–night differences. Based on five adjusted radiosonde data sets (*Wang et al.*, 2012), we show here the smallest and largest reported trends and consider their range (including the reported error bars) as the observational uncertainty range. Similar to the observations, the reanalysis data suggest a large range in cold point temperature trends, from no trend at all (0 K/decade for ERA-Interim) to a strong cooling of - 1.3 K/decade (NCEP-NCAR R1). The latter is outside of the observational uncertainty range and can thus be considered unrealistic. All other reanalyses suggest small but significant cooling trends of - 0.3 K/decade to - 0.6 K/decade. JRA-25, JRA-55, MERRA, and MERRA-2 agree particularly well and produce trends in the middle of the observational uncertainty range. Overall, due to the large uncertainties in radiosonde-derived cold point temperature trends, all reanalyses except for NCEP NCAR R1 are statistically consistent with at least one of the observational data sets.

8.2.4 Key findings and recommendations

Key findings

- The reanalysis data sets ERA5, ERA-Interim, MERRA-2, JRA-55, and CFSR provide realistic representations of temperature structure within the TTL. There is good agreement between reanalysis tropical mean temperatures and GNSS-RO retrievals, with relatively small cold biases for most data sets (best agreement for CFSR, - 0.06 K). However, the cold point and lapse rate tropopause based on reanalyses show warm biases when compared to observations (best agreement for ERA5, 0.05 K), most likely related to the fact that the discrete values corresponding to reanalysis model levels are unable to reproduce the observed minimum temperature as recorded in a near-continuous profile. (*Section 8.2.2*)

- Interannual variability in reanalysis temperatures is best constrained in the upper TTL (70 hPa), with larger differences at lower levels such as the cold point and 100 hPa. The reanalyses reproduce the temperature responses to major dynamical and radiative signals such as volcanic eruptions and the QBO. Long-term reanalysis trends in temperature at 70 hPa during 1979 - 2005 show good agreement with trends derived from adjusted radiosonde data sets indicating significant stratospheric cooling at this level of around - 0.5 K/decade to - 1 K/decade.

At the cold point, both adjusted radiosonde data sets and reanalyses show large uncertainties in temperature trends with most data sets suggesting small but significant cooling trends. (*Section 8.2.3*)

- Advances in reanalysis and observational systems over recent years have led to a clear improvement in TTL reanalysis products over time. In particular, the reanalyses ERA-Interim, ERA5, MERRA2, CFSR, and JRA-55 show very good agreement after 2002 in terms of the vertical TTL temperature profile, meridional tropopause structure and interannual variability. Step-like improvements also occurred around the TOVS-to-ATOVS transition in 1998 - 1999 and the introduction of COSMIC data in 2006. (*Section 8.2*)

Key recommendations

- In the TTL, temperature on native model levels should be used rather than the standard pressure-surface data sets. Various diagnostics such as the cold point tropopause and the analysis of equatorial waves are demonstrably improved when model-level data are used. The cold point tropopause derived from pressure levels is too warm and too low, while temperature at the 100 hPa pressure level underestimates equatorial wave amplitudes. (*Section 8.2*)

- Cold point and lapse rate tropopause temperature depend on the overall temperature bias and on the vertical resolution of the model level data. For a more realistic representation of the tropical tropopause levels, data sets that combine low temperature biases with high vertical resolution should be used. (*Section 8.2*)

8.3 Clouds and convection

Clouds and convection play important roles in tropical climate and meteorology, including the radiation budget and atmospheric water cycle. Although clouds are primarily model products in reanalyses, many of the variables that influence cloud distributions in the tropics (such as SSTs and atmospheric temperatures, moisture, and winds) are either prescribed as boundary conditions or modified by data assimilation. Differences in cloud fields thus depend on both the physical parameterizations used in the forecast model, and the type and strength of data assimilation constraints on the state of the reanalysis atmosphere. Similarly, the effects of biases in cloud fields may either be pervasive (for variables that are not analyzed, such as radiative heating rates or the top-of-atmosphere energy balance) or mitigated by the data assimilation (for variables that are analyzed, such as temperature and atmospheric humidity). *Chapter 2* of this report provides some information on how cloud fields are generated within the different reanalysis products and how these fields interact with radiation (**Tables 2.4, 2.5**, and **2.6**; see also *Appendix A* of *Wright et al.*, 2020).

In this Section, we examine reanalysis cloud products in the tropics, focusing on the tropical upper troposphere. The variables examined include cloud fraction and cloud water content (CWC) in the upper troposphere, outgoing longwave radiation (OLR), and short-wave and long-wave cloud radiative effects (SWCRE and LWCRE; defined as clear-sky minus all-sky fluxes) at the nominal top-of-atmosphere (TOA). Comparisons are performed on common grids of $2.5° \times 2.5°$ and for overlapping time periods where appropriate. Spatial distributions of cloud cover and cloud radiative effects are evaluated against a reanalysis ensemble mean (REM) that includes ERA-Interim, JRA-55, MERRA-2, and CFSR/CFSv2. ERA5 and MERRA are also included in selected results, but earlier reanalyses (such as ERA-40, JRA-25, NCEP-NCAR R1, and NCEP-DOE R2) and surface-input reanalyses (20CR and ERA-20C) are omitted. Parts of the evaluations and investigations can be found in *Wright et al.* (2020).

8.3.1 Observational data sets

We provide some observational comparisons for context, including observations from the AIRS, CERES, CloudSat, ISCCP, and MODIS satellite missions and TOA radiation products from NASA-GEWEX SRB and NOAA OLR. An important caveat is that satellite observations of clouds and OLR are often not directly comparable to reanalysis products due to biases in observational capabilities, diurnal sampling, and other factors. Observational benchmarks are thus treated more as qualitative than quantitative, especially for cloud fields.

AIRS

We use level 3 data from the Atmospheric Infrared Sounder (AIRS) for observations of the thermodynamic state of the atmosphere, primarily daily means from the AIRS version 6 'TqJoint' collection (*Texeira*, 2013). This collection provides gridded representations of temperature and moisture fields based on consistent sets of initial retrievals in each grid cell, along with quality-controlled representations of cloud properties and many other variables (*Tian et al.*, 2013). As the finest temporal resolutions of other data examined in this intercomparison are daily means, we average data from ascending and descending passes together. Variables used from AIRS TqJoint products include temperature, water vapor mass mixing ratio, and geopotential height, which are used to calculate derived metrics such as relative humidity with respect to liquid water, equivalent potential temperature and moist static energy. AIRS TqJoint products have been acquired from the NASA Goddard Earth Sciences Data and Information Services Center (GESDISC) at **https://daac.gsfc.nasa.gov**.

CERES

We use two TOA radiation flux products from the Clouds and the Earth's Radiant Energy System (CERES) experiment Earth Observing System (EOS) Terra & Aqua collection for the period March 2000 through December 2014. First, we use monthly-mean TOA fluxes calculated from Edition 4.1 of the Energy Balanced and Filled (EBAF) monthly-mean products at $1° \times 1°$ spatial resolution (*Doelling*, 2019). Edition 4.1 of EBAF includes two sets of clear-sky fluxes at TOA (*Loeb et al.*, 2020), one that represents direct observations in 'cloud-free' portions of the grid cell (a traditional approach for observationally-based TOA flux datasets) and one that represents clear-sky fluxes estimated for the entire grid cell. We use the latter, as it is more suitable for comparison with clear-sky fluxes from reanalysis models. Second, we use daily-mean Synoptic Radiative Fluxes and Clouds (SYN1Deg) Edition 4A products at $1° \times 1°$ spatial resolution (*Doelling*, 2017). The SYN1Deg data set provides several estimates of TOA radiative fluxes, including direct measurements, outputs from initial 'untuned' radiative transfer model simulations, and outputs from a second set of radiative transfer simulations in which the model input variables are adjusted to bring the simulated fluxes into better agreement with the observed fluxes. The initial atmospheric state for these radiative computations is taken from the GEOS-5 data assimilation system, which is also used for MERRA-2. Only the 'adjusted' fluxes are used to compute the cloud radiative effects discussed in *Chapter 8*, as these are more analogous to the reanalysis flux products. Results based on the observed fluxes are similar but with some changes in magnitudes. Along with TOA radiative fluxes, the SYN1Deg data set provides estimates of cloud fraction retrieved using measurements collected by MODIS and geostationary satellites. We use these estimates of high cloud fraction in conjunction with the SYN1Deg radiative fluxes when daily data are required. CERES data are provided via the CERES Data Products web interface hosted by the NASA Langley Atmospheric Science Data Center (**https://ceres.larc.nasa.gov**).

CloudSat / CALIPSO

We include several observationally-based cloud products based on measurements from the CloudSat and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite missions. These include two estimates of vertical profiles of cloud fraction, one based on combined information from CloudSat and CALIPSO (*Kay and Gettelman*, 2009) and one based on CALIPSO alone (*Chepfer et al.*, 2010), both provided monthly at $2° \times 2°$ horizontal resolution. The combined CloudSat-CALIPSO product covers the period July 2007 through February 2011, after which CloudSat switched to sunlit-only observations. The CALIPSO-only product is the GCM-Oriented CALIPSO Cloud Product (GOCCP) provided by the Laboratoire de Météorologie Dynamique at the Institut Pierre Simon Laplace (IPSL). We use data from January 2007 through December 2014. CloudSat-CALIPSO combined cloud fractions were provided by Jennifer Kay (*personal communication*, 15 December 2017), and CFMIP-GOCCP products by IPSL (**http://climserv.ipsl.polytechnique.fr/cfmip-obs/goccp_v3.html**; v3.1.2 accessed 21 June 2018). In addition to cloud fraction products, we use ice water content (IWC) measurements from CloudSat, namely version 4 of the 2C-ICE profile product (*Deng et al.*, 2015). This retrieval is based on retrieved ice water path from CloudSat radar reflectivity and the backscatter coefficient from the CALIOP lidar, and uses Rodgers optimal estimation in the retrieval. CloudSat- and CALIPSO-based data sets are provided on a 40-level height grid. We convert these height coordinates to pressure using the barometric equation with a scale height of 7.46 km. This approach introduces uncertainties in the precise vertical location (in pressure) of features observed by CloudSat and CALIPSO, which should be taken into consideration when comparing these features to those produced by the reanalyses.

ISCCP

The International Satellite Cloud Climatology Project (ISCCP) has produced observationally-based descriptions of clouds and their attributes using geostationary and polar-orbiting satellite measurements starting from July 1983 (*Rossow and Schiffer*, 1991, 1999). We use high cloud fractions from the monthly ISCCP HGM product (*Rossow et al.*, 2017), which extend the ISCCP record through June 2017. These data are provided on a $1° \times 1°$ horizontal grid. High clouds are defined as having cloud top pressures less than 440 hPa, and include the cirrus, cirrostratus, and deep convective cloud types. ISCCP HGM products are hosted by NOAA NCEI and are available at **https://www.ncei.noaa.gov/data/international-satellite-cloud-climate-project-isccp-h-series-data/access/isccp-basic/hgm/**.

MODIS

The Moderate Resolution Imaging Spectroradiometer (MODIS) instrument has been flown on the Terra and Aqua satellites starting from early 2000 and mid-2002, respectively. We use high cloud fractions from Collection 6 of the Terra MODIS Level 3 MOD08 Atmosphere Product (*Platnick*, 2015). MODIS gridded cloud products are available from NASA Goddard via the web interface at **https://modis.gsfc.nasa.gov**.

NASA-GEWEX SRB

The NASA Global Energy and Water Cycle Experiment (GEWEX) Surface Radiation Budget (SRB) project has produced radiative fluxes and related variables at both surface and TOA spanning approximately 2.5 decades (*Zhang et al.*, 2013). We use TOA longwave fluxes between January 1984 and December 2007. These products are based on radiative calculations using observed fluxes and ozone together with GEOS-4 analyses of temperature and water vapour.

Pixel-level information from ISCCP is used to derive cloud radiative effects. The NASA GEWEX-SRB data are provided by the NASA Langley Atmospheric Science Data Center.

NOAA OLR

The NOAA Interpolated OLR product (*Liebmann and Smith*, 1996) provides estimates of all-sky OLR at the TOA starting from June 1974. Initial estimates based on radiances observed by polar-orbiting satellites are used to fill gaps via interpolation in time and space. We use monthly-mean estimates of all-sky OLR from this product covering January 1980 through December 2014. The NOAA Interpolated OLR data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at **https://www.esrl.noaa.gov/psd/**.

8.3.2 Spatial distribution of high clouds

Fig. 8.10 shows spatial distributions of high cloud fraction for the REM and ISCCP, as well as differences relative to the REM for ERA-Interim, ERA5, JRA-55, CFSR/CFSv2, MERRA and MERRA-2. The definition of high cloud fraction varies somewhat among these data sets. For example, high clouds are defined as clouds at pressures less than ~500 hPa for JRA-55, as clouds at pressures less than ~400 hPa for CFSR/CFSv2, MERRA, and MERRA-2, and as clouds at pressures less than 0.45 times the surface pressure for ERA-Interim and ERA5 (~450 hPa). High cloud fraction in the ISCCP dataset is defined as clouds with tops at pressures less than 440 hPa (*Rossow and Schiffer*, 1991; 1999). Differences in how cloud fraction is calculated may also play a role. For example, cloud fraction is a prognostic variable in ERA-Interim and ERA5 but is diagnosed as a function of CWC and relative humidity (RH) in CFSR. These details are provided in *Chapter 2* of this report (**Table 2.5**), with additional information and references provided in *Chapter 2E*. We show in *Section 8.3.3* that reanalysis-derived tropical cloud fractions have a minimum between 400 hPa and 500 hPa, so that these differences in the precise definition of high cloud fraction have little impact on the qualitative comparisons presented in **Figure 8.10**.

One of the most striking features of **Figure 8.10** is the systematically larger high cloud fractions produced by MERRA and MERRA-2 relative to the other reanalyses. MERRA and MERRA-2 show tropical mean high cloud fractions greater than 40 %, while all other evaluated reanalyses show tropical mean high cloud fractions less than 35 %. JRA-55 produces the smallest tropical high cloud fractions among the reanalyses, with a tropical mean high cloud fraction of only about 25 %. CFSR/CFSv2 and ERA-Interim also produce tropical mean values slightly less than the REM, but with substantially different spatial distributions. Difference between CFSR/CFSv2 and ERA-Interim are especially pronounced over the Maritime Continent and tropical western Pacific, where CFSR/CFSv2 underestimates the REM, and ERA-Interim exceeds it. These qualitative differences between CFSR/CFSv2 and ERA-Interim are echoed to a lesser extent in other tropical convective regions, such as the Amazon Basin and the Caribbean Sea, and take the opposite sign over mountainous regions such as the Andes and the Tibetan Plateau.

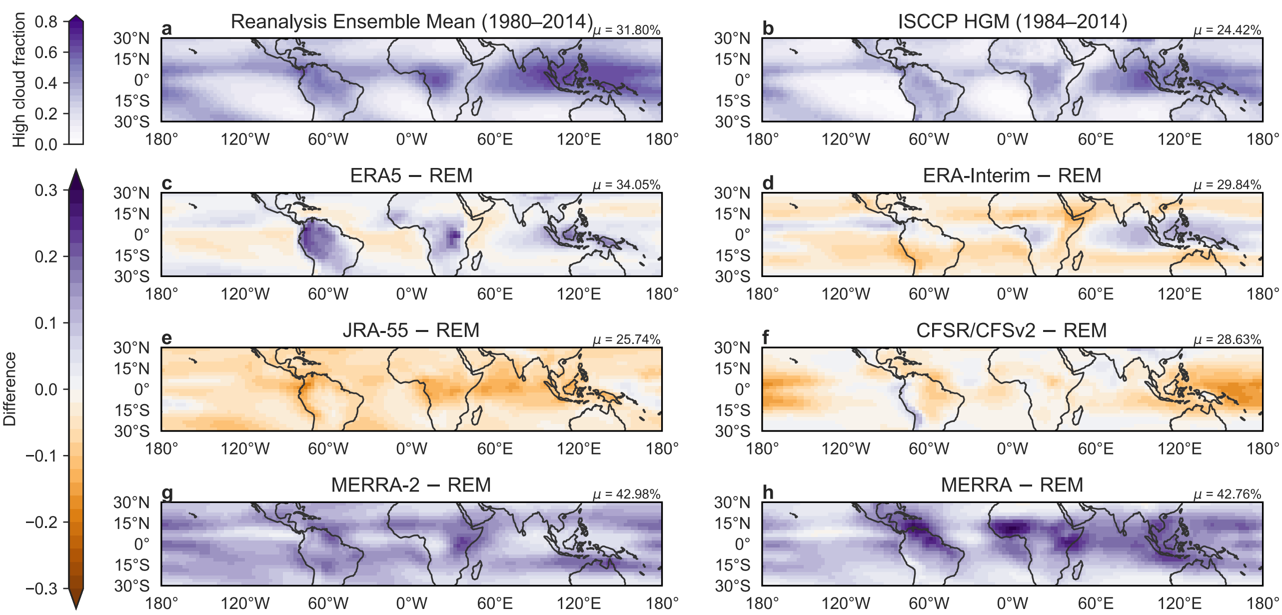


Figure 8.10: *Time-mean spatial distributions of high cloud cover fraction. The upper left panel (a) shows the REM for 1980 - 2014, calculated by averaging the distributions from ERA-Interim, JRA-55, MERRA-2, and CFSR/CFSv2. The upper right panel (b) shows the distribution based on the ISCCP HGM dataset for 1984 - 2014. The remaining panels show differences relative to the REM for (c) ERA-Interim, (d) ERA5, (e) JRA-55, (f) CFSR/CFSv2, (g) MERRA, and (h) MERRA-2. The absolute area-weighted tropical mean (30° S–30° N) (in %) is marked at the upper right corner of each panel. Adapted from Wright et al. (2020).*

Differences between ERA5 and the REM are similar in many ways to those for ERA-Interim, but with further enhancements in the tropical convective regions (especially over land). ERA5 has noticeably larger high cloud fractions than ERA-Interim over tropical South America and Africa, as well as in the South Asian monsoon region, the Pacific portion of the ITCZ, and the SPCZ. Despite these discrepancies, distributions of high cloud cover are nonetheless qualitatively consistent among the reanalyses, with area-weighted pattern correlations against the REM consistently exceeding 0.95. Some possible reasons for the quantitative differences are discussed below in the context of other metrics (see also *Wright et al.*, 2020).

Figure 8.10 also shows the spatial distribution of high cloud fraction based on the ISCCP HGM observationally-based product. ISCCP D2 indicates systematically smaller high cloud fractions than those produced by reanalyses, with a tropical mean of only 24 %. This low bias relative to the REM is consistent among infrared-based observational estimates (see also **Fig. 8.11**) and is not surprising given the expected limitations of these observations. These data products are based on infrared observations near the 11 µm emission band, which are known to underestimate both the top heights of thick high clouds and the occurrence frequency of thin high clouds (*e.g.*, *Pincus et al.*, 2012; *Dessler and Yang*, 2003). MERRA-2 provides an ancillary cloud product based on the COSP satellite simulator (*Bodas-Salcedo et al.*, 2015) that facilitates a more direct comparison. This COSP product emulates what satellites would see if they were observing the reanalysis atmosphere, and includes estimates for MODIS high cloud fraction among other products. **Figure 8.11** shows spatial distributions of tropical high cloud fraction from MERRA-2 and its COSP equivalent, as well as observationally-based distributions from Terra MODIS (*Platnick*, 2015) and the CERES SYN1Deg product (which combines information from Terra MODIS, Aqua MODIS, and geostationary satellites; *Doelling*, 2017). This comparison shows very good agreement between MERRA-2-COSP (25 %) and the satellite-based estimates (24 - 26 %) in the tropical mean. However, it is important to emphasize that this close agreement does not necessarily mean that the larger high cloud fractions in MERRA-2 are more realistic (i.e., that the other three reanalyses substantially underestimate high cloud fraction in the tropics). Rather, it indicates only that MERRA-2 produces a reasonable distribution of the high clouds that can be readily observed by MODIS and similar instruments. A recent study in which a cloud simulator was applied to ERA-Interim outputs also indicated good agreement with observed high cloud fractions in the tropics, but with a slight high bias (~10 %) in the same inner tropical regions where ERA-Interim tends to overestimate the REM (*Stengel et al.*, 2018).

8.3.3 Vertical profiles

The effects of differences in the spatial distribution of cloud fields may be compounded by differences in the vertical distribution of clouds. **Figures 8.12** and **8.13** show zonal-mean vertical distributions of cloud fraction and CWC along with area-mean profiles for the inner tropics (10 ° S - 10 ° N). All reanalyses show maxima in cloud fraction at or just above the base of the TTL (~200 hPa; *Section 8.1*). The peak value in ERA-Interim is centered at 150 hPa, slightly above those in ERA5 (~175 hPa) and MERRA/MERRA-2 (~200 hPa) and slightly below that in JRA-55 (~125 hPa). JRA-55 also shows a secondary, smaller local maximum near 200 hPa. Specific details may be sensitive to our use of data on pressure levels rather than model levels (**Fig. 8.12**), as MERRA and MERRA-2 lack a standard pressure level at 175 hPa. All maxima are most pronounced in the Northern Hemisphere between 5 ° N and 10 ° N, reflecting the preferred position of the ITCZ (*e.g.*, *Schneider et al.*, 2014). CFSR is omitted from **Figure 8.12** because it does not provide a vertically-resolved estimate of cloud fraction.

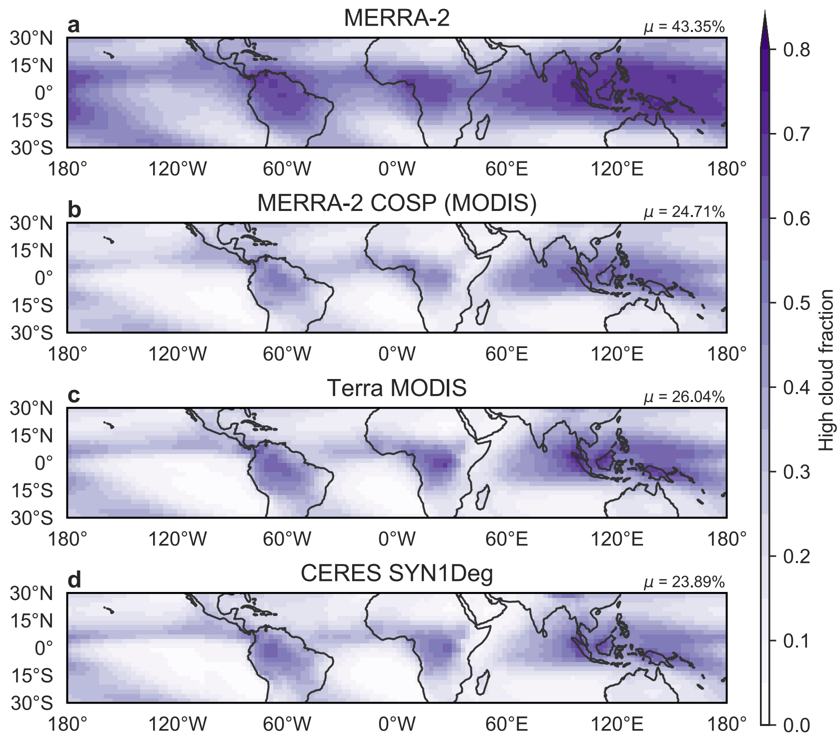


Figure 8.11: *As in* ***Fig. 8.10a****, but for (a) MERRA-2, (b) MERRA-2-COSP, (c) Terra MODIS, and (d) CERES SYN1Deg over the period 2001–2014. Reproduced from Wright et al. (2020).*

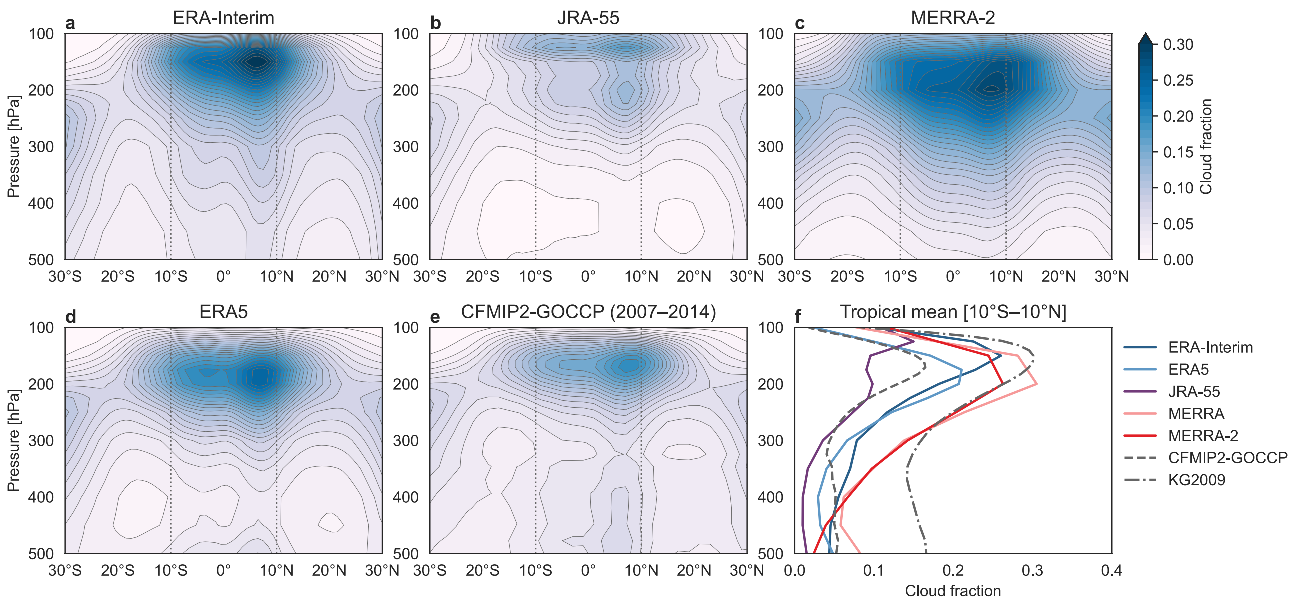


Figure 8.12: *Zonal-mean vertical distributions of time-mean cloud fraction averaged within the tropics (30° S - 30° N) for (a) ERA5, (b) ERA-Interim, (c) JRA-55, and (d) MERRA-2 over 1980 - 2014, along with (e) observational estimates based on the CFMIP2-GOCCP product (2007 - 2014). Profiles shown in panel (f) are averaged over the inner tropics (10° S–10° N), and also include MERRA and a combined CloudSat–CALIPSO product (2007 - 2010; Kay and Gettelman, 2009; KG2009). CFSR is omitted as it does not provide a vertical profile of cloud fraction. Reproduced from Wright et al. (2020).*

Differences among the reanalyses are even more pronounced with respect to time-mean zonal-mean distributions of CWC in the tropical upper troposphere (**Fig. 8.13**). Here, CWC represents the sum of ice and liquid water contents, except for the CloudSat 2C-ICE estimate, which is based on IWC alone. MERRA-2 produces by far the largest CWCs among the reanalyses, with a pronounced peak at 300 hPa. It is worth noting here that although MERRA-2 produces smaller cloud fractions in the tropical upper troposphere than its predecessor MERRA, it produces substantially larger values of CWC. The large values of CWC produced by MERRA-2 have significant impacts on radiative transfer, as outlined in *Section 8.3.3* below (see also *Sect. 8.8.6*), and may also contribute to the more extensive high cloud cover outside the core convective regions relative to MERRA (Fig. 8.10g-h). CFSR/CFSv2 produces a similarly pronounced vertical maximum in cloud water content, but shifted slightly higher in altitude and with a peak magnitude roughly half that produced by MERRA-2 when averaged over 10° S - 10° N. JRA-55 features a qualitatively similar distribution to those of MERRA-2 and CFSR/CFSv2, but with much smaller magnitudes, consistent with other indications that JRA-55 underestimates cloud fields in the tropical upper troposphere (*e.g.*, **Fig. 8.10**). The zonal-mean distribution of CWC in ERA-Interim is remarkably different from that in the other reanalyses, including ERA5, as it shows no distinct maximum in the tropical upper troposphere. Instead, ERA-Interim indicates a monotonic decrease in CWC with increasing altitude above 500 hPa. The difference in vertical profiles of CWC between ERA-Interim and ERA5 may be explained at least in part by changes in the treatment of organized detrainment within the convective scheme. These and other revisions to the cloud and convection schemes (*Bechtold et al.*, 2008; 2014; *Forbes et al.*, 2011) act to enhance detrainment rates in the upper troposphere (200 - 300 hPa) and reduce detrainment closer to the tropopause (100 - 150 hPa) in ERA5 relative to ERA-Interim (see *Wright et al.*, 2020, for details).

Observational context is provided in **Figure 8.12** by vertical profiles of cloud fraction derived from CALIPSO measurements for CFMIP (*Chepfer et al.*, 2010) and derived from combined CloudSat and CALIPSO measurements (*Kay and Gettelman*, 2009). Similar but more limited context is provided in **Figure 8.13** by IWC estimates from the CloudSat-CALIPSO 2C-ICE product (*Deng et al.*, 2015). These data sets are based on active measurements made using radar and lidar profilers, and therefore have different types of biases than cloud fields derived from passive measurements in the 11 µm band (*e.g.*, increased sensitivity to cloud top heights and thin clouds but more limited diurnal sampling). However, although the two observational cloud fraction data sets are based in part on the same underlying observations collected at approximately the same times and locations, the range between these two observational estimates is comparable in magnitude to that among the reanalyses, which complicates evaluation of the reanalysis products. Given also the lack of suitable observation simulators applied to the reanalysis fields, we avoid further quantitative comparison. Qualitatively, the observational estimates are more consistent with the single anvil-type peaks in cloud fraction around 150 - 200 hPa as produced by ERA-Interim, ERA5, MERRA, and MERRA-2 than with the double-peak structure produced by JRA-55.

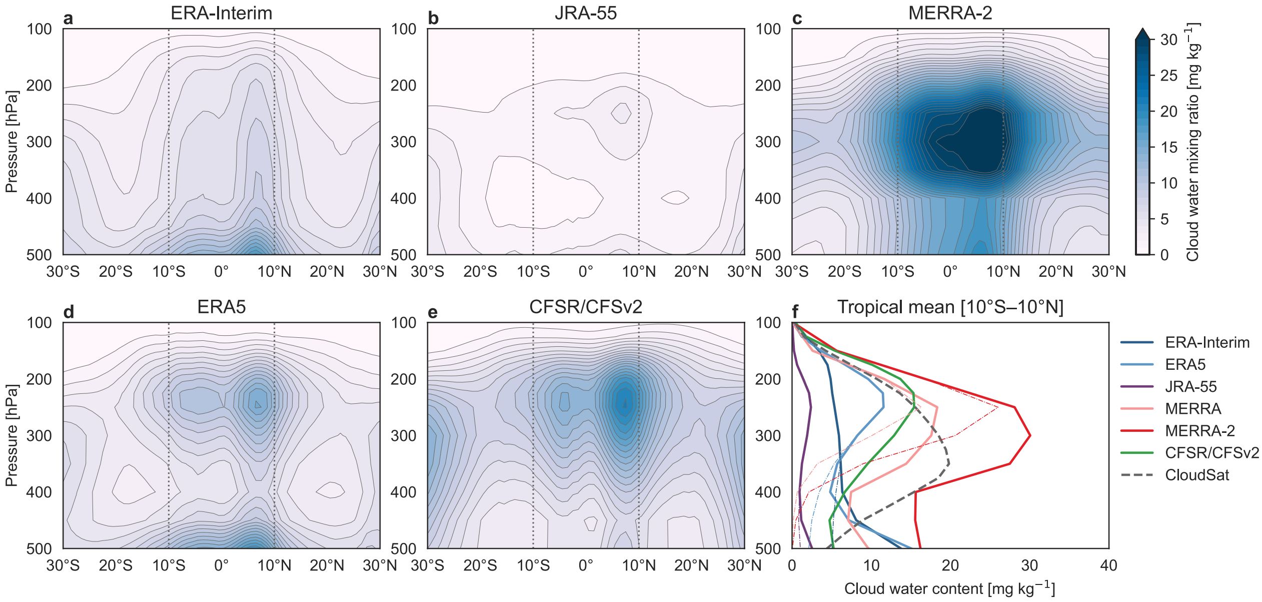


Figure 8.13: *As in* ***Fig. 8.12****, but for time-mean zonal-mean total CWC (LWC + IWC) in mg kg^{-1}. Differences from* ***Fig. 8.12*** *are the inclusion of CFSR/CFSv2 1980 - 2014 mean in panels (e) and (f) and the source of the observational estimate in panel (f). The latter is based on the CloudSat-CALIPSO 2C-ICE total IWC product (cloud ice + snow) for 2007 - 2010. Thin dotted lines in (f) indicate ice-only estimates of CWC for the reanalyses that provide them (all except CFSR/CFSv2). Reproduced from Wright et al. (2020).*

However, none of the reanalyses captures the observed peak in cloud fraction near 500 hPa associated with shallower cumulus congestus clouds. For CWC, the 2C-ICE profile is likewise more consistent with the anvil layers produced by ERA5, MERRA, MERRA-2, and CFSR/CFSv2, although the reanalyses typically show smaller magnitudes and place the peak value at somewhat higher altitudes than observed. These differences are expected, as the 2C-ICE algorithm measures total IWC (including snow) while the reanalyses account only for cloud condensate, again precluding quantitative comparison (*e.g., Li et al.*, 2016; see also *Sect. 8.8.6*). Unlike in cloud fraction, the reanalyses do show larger values of CWC around the cumulus congestus detrainment level (~500 hPa). Although this peak is not present in the observed IWC, this may be explained by the primarily liquid composition of CWC at these levels in the reanalyses (**Fig. 8.13f**).

Differences in the mean vertical profiles of cloud fraction and CWC among reanalyses suggest differences in the preferential location and subsequent evolution of anvil clouds detrained from deep convection. For example, detrainment appears to peak at lower altitudes and higher pressures in MERRA and MERRA-2 than in the other reanalyses. The cloud fraction maximum at 125 hPa in JRA-55 suggests that convective detrainment may be more likely to penetrate across the LZRH (*Section 8.4.2*) in JRA-55 than in other systems, while the peak values of CWC in ERA5 and CFSR/CFSv2 are clearly shifted upward relative to MERRA-2. Such differences reflect the specific treatments of detrainment within the deep convective scheme, but may also indicate systematic differences in the tropical circulation as represented by the reanalysis. The latter may respond to other aspects of the convective scheme (the convective trigger, treatment of mixed-phase condensate, autoconversion, *etc.*), as well as other physical parameterizations (boundary-layer turbulence, interactions between radiation and clouds) and/or the types or treatments of assimilated data (see also *Wright et al.*, 2020).

8.3.4 Cloud radiative effects

Tropical high clouds have substantial climatic impacts, particularly via their influences on the radiation budget (*e.g., Stevens and Schwartz*, 2012). For example, the presence of thick high clouds (such as anvil clouds associated with tropical deep convection) substantially reduces the OLR. This LW effect is offset to some extent by the additional reflection and absorption of solar radiation by thick high clouds. Such compensation does not occur with thin high clouds, which are largely transparent to incoming solar radiation but opaque to outgoing LW radiation. Here, we examine how differences in the distribution of high clouds in reanalyses alter LW and SW fluxes at the nominal TOA. In *Section 8.4.2* we extend this discussion to include the convergence of LW and SW radiation in the tropical UTLS. High clouds are the dominant factor in determining LW cloud impacts, but play a more limited role in SW effects (*e.g., Zelinka et al.*, 2012). We therefore focus primarily on the role of high clouds in altering LW fluxes at the TOA. Additional discussion of SW and net effects has been provided by *Wright et al.* (2020).

Figure 8.14 shows spatial distributions of the OLR and LWCRE based on various reanalysis and observational data sets. The LWCRE is calculated for each data set by subtracting the all-sky OLR from the clear-sky OLR.

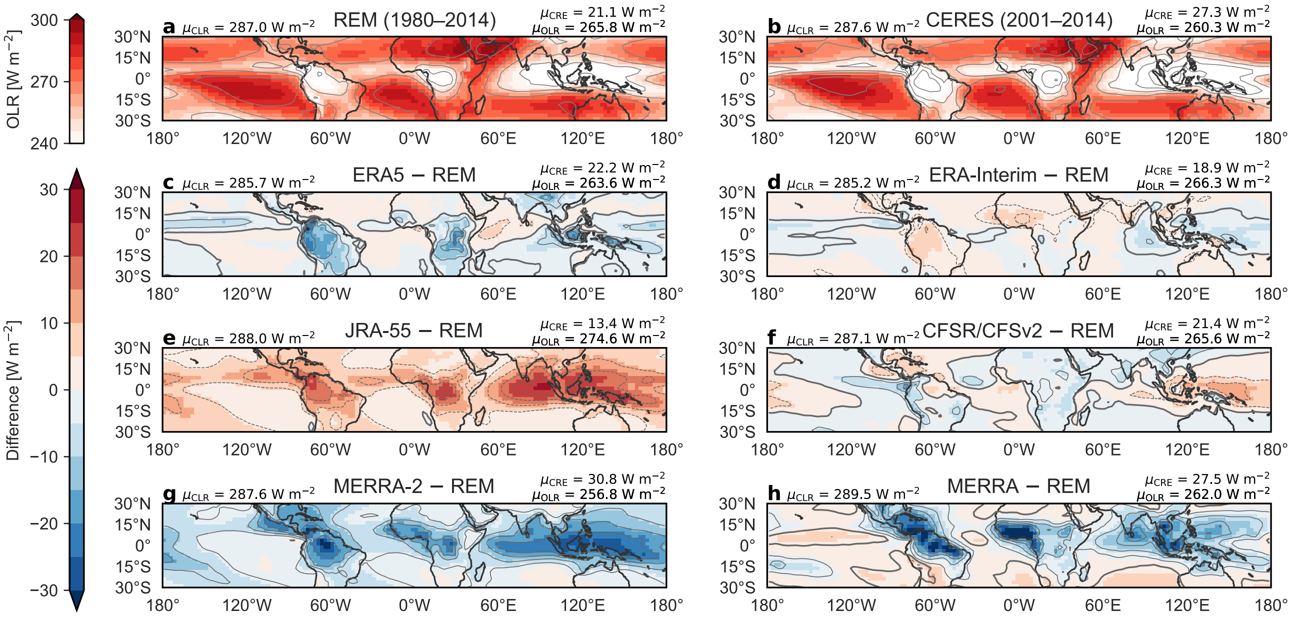


Figure 8.14: *Time- mean spatial distributions of OLR (shading) and LWCRE (contours) [in W m-2]. The upper left panel (a) shows the REM, which is constructed by averaging the climatological means from CFSR/CFSv2, ERA-Interim, JRA-55, and MERRA-2 over 1980 - 2014. The upper right panel (b) shows estimates from CERES EBAF (Edition 4.1) over 2001 - 2014. The remaining panels show differences [in W m-2] relative to the REM for (c) ERA-Interim, (d) ERA5, (e) JRA-55, (f) CFSR/CFSv2, (g) MERRA, and (h) MERRA-2. Note that the REM is biased high relative to CERES EBAF, so that reanalyses with low biases relative to the REM are in better agreement with observations (see text for details). Area-weighted tropical mean (30°S–30°N) values of OLR and LWCRE are shown at the upper right corner of each panel, with corresponding values for clear-sky OLR at upper left. Adapted from Wright et al. (2020).*

These quantities may be derived in slightly different ways for observational and reanalysis data sets. For observational data sets, all-sky fluxes are computed by aggregating all observations. Clear-sky fluxes may be estimated by aggregating observations flagged as cloud-free but may also be derived by running radiative transfer simulations constrained by observed fluxes, with some combination of observed and analysis fields used to specify the atmospheric state. We use the latter type to define 'observational' LWCREs. For reanalyses, all-sky and clear-sky fluxes are computed by running the radiation parameterization for profiles with and without the model-generated cloud fields. As with high clouds, reanalyses generally provide realistic spatial distributions of the time-mean OLR and LWCRE: pattern correlations against the REM are consistently larger than 0.9, and pattern correlations between observational estimates and the REM all exceed 0.97 (including the NOAA OLR and NASA GEWEX-SRB datasets; not shown). Spatial distributions of biases in OLR are qualitatively opposite to spatial distributions of biases in LWCRE (*i.e.*, biases in OLR are positive where biases in LWCRE are negative and vice versa). This situation reflects the preeminent role of clouds in determining the spatial pattern of OLR in the tropics: an underestimate of LWCRE corresponds to an underestimate of cloud impacts on net absorption within the column and thus an overestimate of OLR, while an overestimate of LWCRE has the opposite effect.

The REM indicates a tropical mean OLR of 266 W m^{-2} and a tropical mean LWCRE of 21 W m^{-2} (**Fig. 8.14**). Both CFSR and ERA-Interim produce tropical mean values of OLR and LWCRE that are very close to the REM, but with spatial bias distributions that are qualitatively opposite in many respects. CFSR produces high biases of OLR and low biases of LWCRE relative to the REM over most of the tropical oceans, particularly near the maritime continent, while producing low biases of OLR and high biases of LWCRE over the eastern tropical Pacific and land regions with strong convection, such as equatorial Africa. ERA-Interim, by contrast, produces low biases in OLR and high biases in LWCRE relative to the REM over oceanic deep convective regions, but high biases in OLR and low biases in LWCRE over large parts of the tropical continents. ERA5 produces a slightly smaller tropical-mean OLR and slightly larger LWCRE than ERA-Interim, consistent with its larger high cloud fraction (**Fig. 8.10**). The changes are again most pronounced over tropical land areas with strong convection, especially South America, Africa, and the South Asian monsoon region (**Fig. 8.14**). JRA-55 substantially overestimates OLR and underestimates LWCRE relative to the REM, with biases of nearly 10 W m^{-2} relative to the REM. The biases in JRA-55 are opposite to those in MERRA-2, for which the tropical mean OLR is smaller than the REM by 10 W m^{-2} and the LWCRE is larger than in any other reanalysis. Differences between MERRA-2 and JRA-55 are particularly pronounced in tropical deep convective regions.

Observationally-based estimates of OLR and LWCRE shown in **Figure 8.14** are taken from the CERES EBAF (*Section 8.3.1*).

The tropical-mean OLR based on CERES is smaller than that based on the REM but not quite as small as that based on MERRA-2. We have also examined other observationally-based estimates, such as the NASA GEWEX-SRB product (*Zhang et al.*, 2013) and the NOAA Interpolated OLR product (*Liebmann and Smith*, 1996). The SRB data indicate a tropical mean OLR of 259 W m^{-2} and a tropical mean LWCRE of 28 W m^{-2}, in good agreement with CERES. The NOAA OLR indicates a tropical mean OLR of 252 W m^{-2}, even smaller than that in MERRA-2 (the NOAA OLR does not provide a clear-sky estimate so cannot be used to estimate LWCRE). The REM is thus biased high relative to the NOAA OLR by nearly 15 W m^{-2}. Based on this context, MERRA-2 appears to produce the most realistic values of OLR and LWCRE averaged over the tropics among these reanalyses. Moreover, the magnitudes by which ERA-Interim, ERA5, and CFSR/CFSv2 underestimate the LWCRE are approximately twice as large as the magnitudes by which they underestimate OLR: these reanalyses underestimate clear-sky OLR but overestimate all-sky OLR. However, it is worth noting that these comparisons are not strictly independent, as both the SRB and CERES products use temperature and moisture profiles from the GEOS-4 (SRB) or GEOS-5 (CERES) data assimilation systems during data processing.

Figure 8.15 summarizes joint distributions of daily-mean gridded LWCRE and SW cloud radiative effect (SWCRE) relative to daily-mean gridded high cloud fraction during 2001 - 2010 in the inner tropics (10° S–10° N at 1° × 1° grid spacing). The distributions highlight differences among the LWCREs across different data sets and their relationships with high cloud cover in the tropics. CFSR, ERA-Interim, and JRA-55 underestimate LWCRE relative to CERES, with 75th percentile values between 20 W m^{-2} and 35 W m^{-2} smaller than the CERES SYN1Deg benchmark. This low bias in LWCRE is particularly pronounced in JRA-55, as also indicated by the spatial distributions shown in **Figure 8.14**. MERRA-2 is quantitatively in better agreement with CERES-based estimates, although this reanalysis produces a pronounced modal 'lobe' of strong LWCRE at larger values of high cloud fraction that is not seen in the observations (**Fig. 8.15**). This difference, which is also evident in the relationship between high cloud cover and SWCRE, results from the separate treatments of anvil clouds and in situ clouds by the prognostic cloud scheme in MERRA-2 (*Chapter 2*; **Table 2.5**). As a result, MERRA-2 tends to overestimate the LWCRE in convective regions (**Fig. 8.15**). Like MERRA-2, ERA-Interim shows a bimodal structure in the joint distribution of high cloud cover and LWCRE. However, whereas the large-LWCRE mode in MERRA-2 is centered near high cloud fractions of 60 - 80 %, that in ERA-Interim is associated almost exclusively with high cloud fractions near 100 %. The range of LWCRE produced within the tropics provides another useful metric. CERES indicates that the distribution of LWCRE in the tropics has a long tail at large values (more than 100 W m^{-2}), where the latter is associated with large values of high cloud fraction. Among the reanalyses, only CFSR shows a long tail similar to that found in the CERES estimates. However, CFSR overestimates the occurrence frequency of small values and underestimates the occurrence frequency of large values relative to CERES, as indicated by the sharper curvature of the joint distribution and the smaller 75th percentile value of LWCRE.

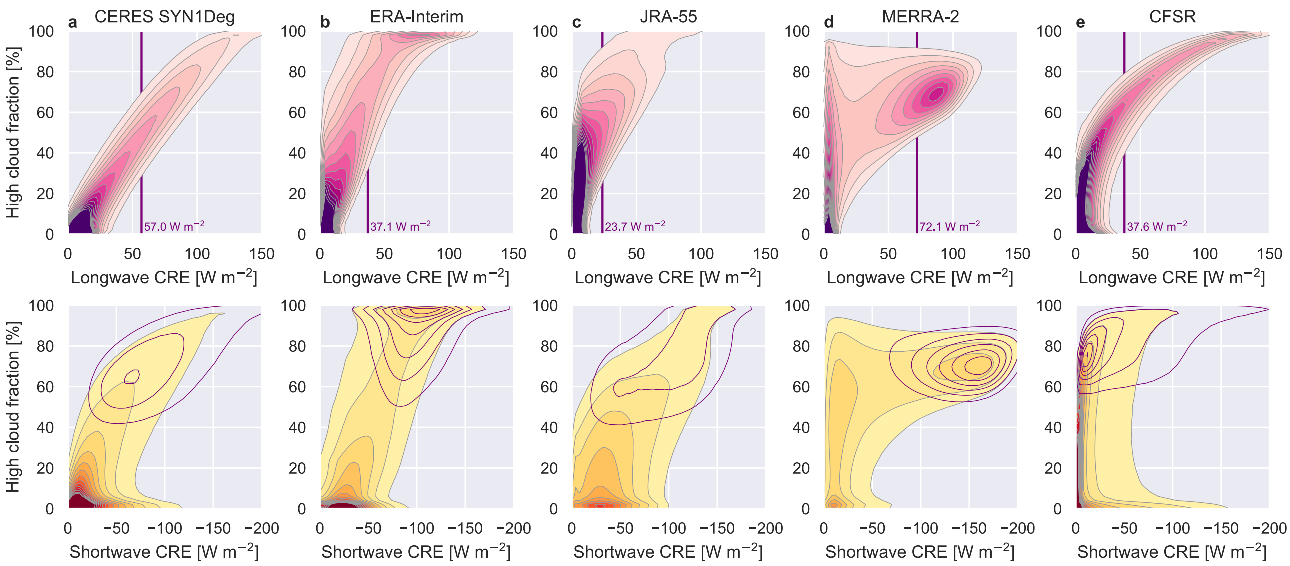


Figure 8.15: *Two-dimensional joint frequency distributions of daily-mean LWCRE (upper) and SWCRE (lower) relative to daily-mean high cloud fraction for 1° × 1° grid cells in the inner tropics (10° S - 10° N) during 2001 - 2010. From right to left, distributions are based on fluxes at the nominal TOA from (a) CERES SYN1Deg, (b) ERA-Interim, (c) JRA-55, (d) MERRA-2, and (e) CFSR. Vertical lines in the upper panels mark the 75th percentile of daily gridded LWCRE. Purple contours in the lower panels show joint distributions of high cloud fraction and SWCRE conditional on the upper quartile of LWCRE (i.e., LWCRE exceeding the 75th percentile). Shading and contours show frequency densities of paired data values (i.e., two-dimensional histograms). The same contour intervals are used for all datasets. Adapted from Wright et al. (2020).*

The joint distributions of high cloud fraction against SWCRE, including the distributions conditioned on large values of LWCRE, also show large differences among the reanalyses. Based on these distributions, it appears that both ERA-Interim and MERRA-2 overestimate SWCREs associated with deep tropical convection, while CFSR underestimates these effects. The prevalence of large values of SWCRE at very small values of high cloud fraction in CFSR suggests that cloud albedo effects are primarily associated with low clouds in this reanalysis, a relationship that also emerges in the CERES-based distribution but is missing or masked by extensive high cloud cover in MERRA-2. The distribution based on JRA-55 is more consistent with the CERES-based distribution, although JRA-55 still overestimates the SWCRE.

8.3.5 Relationships with other variables

The spatial distribution of high cloud cover shown in **Figure 8.10** is controlled to leading order by the spatial distribution of deep convection, which is closely linked to the spatial distribution of SST (*e.g.*, *Fu et al.*, 1996). Other factors include the thermodynamic structure of the atmosphere, large-scale vertical motion, and relative humidity in the mid-troposphere (*e.g.*, *Su et al.*, 2011). **Figure 8.16** shows joint distributions of daily-mean gridded high cloud cover relative to daily-mean gridded SST, potential instability ($\theta_{e,850\mathrm{hPa}}$ - $\theta_{e,500\mathrm{hPa}}$, where θ is equivalent potential temperature), grid-scale vertical velocity at 500 hPa (ω), and grid-scale RH at 500 hPa during 2001 - 2010. These relationships (and results for ERA5, omitted here) have been discussed in more detail by *Wright et al.* (2020); here, we briefly touch on some key points.

The reanalyses generally capture the relationship between SST and high cloud cover, in which tropical convection (associated with large values of high cloud fraction) tends to cluster over the largest SSTs. However, apart from CFSR, this relationship is usually stronger in the reanalyses than observed. Relationships with potential instability in the lower troposphere show larger discrepancies among the reanalyses.

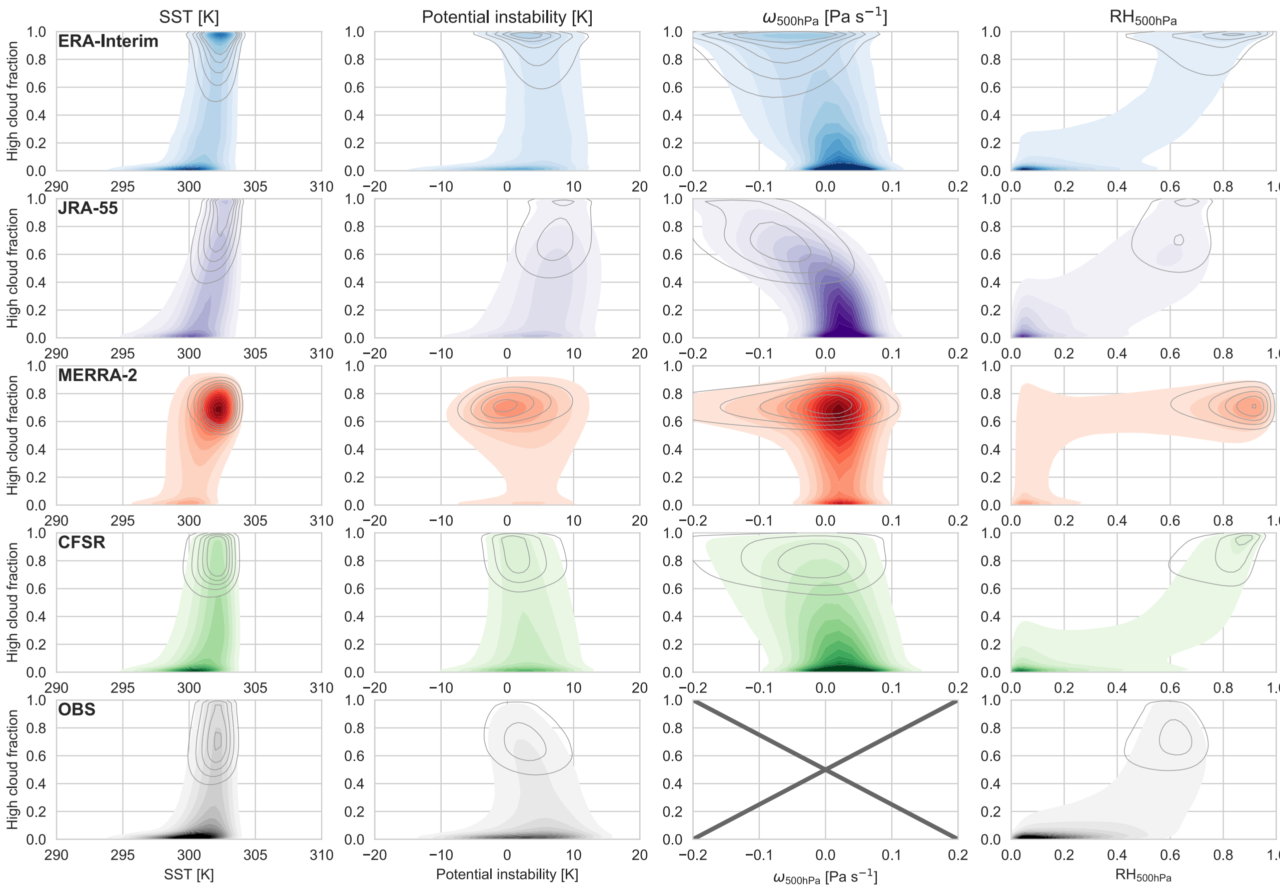


Figure 8.16: *As in **Fig. 8.15**, but for daily-mean gridded high cloud cover against SST (far left), potential instability in the lower troposphere (centre left), mid-tropospheric vertical velocity (centre right), and mid-tropospheric RH (far right) in the inner tropics (10 °S - 10 °N) during 2001 - 2010. Distributions are shown for ERA-Interim (first row; blue), JRA-55 (second row; purple), MERRA-2 (third row; red), CFSR (fourth row; green), and observational benchmarks (bottom row; grey). Daily-mean observational estimates are from CERES SYN1Deg (high cloud cover and LWCRE), Optimum Interpolation Sea Surface Temperature (OISST) v2 (SST), and AIRS TqJoint (potential instability and RH500; limited to 2003 - 2010). Grey contour lines in each panel mark distributions corresponding to the upper quartile of LWCRE as noted in **Fig. 8.15**.*

Whereas MERRA-2 and CFSR agree well with the distribution based on CERES and AIRS, both ERA-Interim and JRA-55 show larger values of potential instability associated with larger values of LWCRE. For ERA-Interim, this difference may be explained by the convective closure, which consumes convective instability more slowly at the relatively coarse horizontal resolution used for that reanalysis (*Bechtold et al.*, 2008; their **Fig. 1**). In this case, substantial instability may remain in the column even after convection has produced extensive high cloud cover. For JRA-55, this difference is related to the convective trigger function, which requires that convective cloud base be sited at the level closest to ~900 hPa. As a result, moist entropy that builds up at 850 hPa may not be released until instability develops at 900 hPa as well. This leads to a clear 'kink' in the profile of moist static energy in the JRA-55 lower troposphere that does not appear in other reanalyses or in AIRS observations (**Fig. 8.17**; see also expanded version in *Wright et al.*, 2020), and illustrates the extent to which details of the convection schemes can imprint on analyzed variables. Relationships with mid-tropospheric vertical velocity are strongest in JRA-55 and ERA-Interim and weakest in MERRA-2 (**Fig. 8.16**). This difference also relates to differences in the convective trigger functions; namely that the trigger functions in JRA-55 and ERA-Interim explicitly use grid-scale vertical velocities to represent large-scale controls on convective activity, while that in MERRA-2 does not. Relationships with mid-tropospheric RH are qualitatively consistent among the reanalyses, except for the distinct lobe of high cloud cover at large RH in MERRA-2. Similar lobes are evident in other joint distributions based on MERRA-2 (including those in **Fig. 8.15**), and result from different treatments of anvil condensate and in situ condensate in the prognostic large-scale cloud scheme in MERRA-2. The other striking feature of the RH distributions in **Figure 8.16** is the difference in mid-tropospheric RH associated with the strongest deep convection. The values of RH used to construct this figure are all calculated relative to saturation with respect to liquid water. The smaller values of RH in JRA-55 relative to MERRA-2, for example, can thus be understood in terms of different treatments of the liquid–ice transition (see, *e.g.*, **Fig. 2.3** in the extended digital version of *Chapter 2*). Whereas JRA-55 assumes that condensate is entirely liquid at 0 °C and entirely ice at -15 °C, with a linear partitioning between these two endpoints, ERA-Interim partitions condensate using a quadratic function of temperature between 0 °C and -23 °C. The effects of these different treatments are also evident in the ice fraction of CWC in these reanalyses (**Fig. 8.13f**). Comparison against the CERES/AIRS distribution shown at the lower right of **Figure 8.16** suggests that the mid-tropospheric RH distribution based on JRA-55 is more realistic than the others at this spatial resolution (daily means at 1°). CFSR provides the closest match with observations for distributions of high cloud cover against SST and potential instability.

Figure 8.17 shows that MERRA-2 has larger values of upper tropospheric moist static energy (MSE) in convective regions relative to the other reanalyses. This difference reflects both a systematic moist bias, perhaps due to greater detrainment of cloud water and subsequent condensate evaporation (**Fig. 8.13**; see also **Fig. 8.22**), and a systematic warm bias, possibly linked to more intense cloud radiative heating at anvil level (see **Fig. 8.23**; *Sect. 8.4.2*). For example, at 300 hPa, the greater MSE associated with the upper quartile of LWCRE in MERRA-2 relative to ERA-Interim is on average 62 % due to differences in the dry enthalpy component (cpT) and 35 % due to differences in the latent energy component (Lvq), with the residual discrepancy (3 %) arising from differences in geopotential. This difference in upper tropospheric MSE is systematic throughout the tropics (see, *e.g.*, profiles corresponding to the lower quartiles of LWCRE in **Fig. 8.17**), but with temperature biases a proportionally greater contributor outside of the main deep convective regions. Greater upper tropospheric MSE in MERRA-2 implies larger gross moist stability and especially a stabilization of the upper troposphere that may suppress the average depth of convection. Indeed, the lower, more extensive anvil deck in MERRA-2 appears to be a primary factor in the relatively strong above-cloud radiative cooling in this reanalysis, as well as the inability of convective heating to compensate for this effect (*Sect. 8.4*).

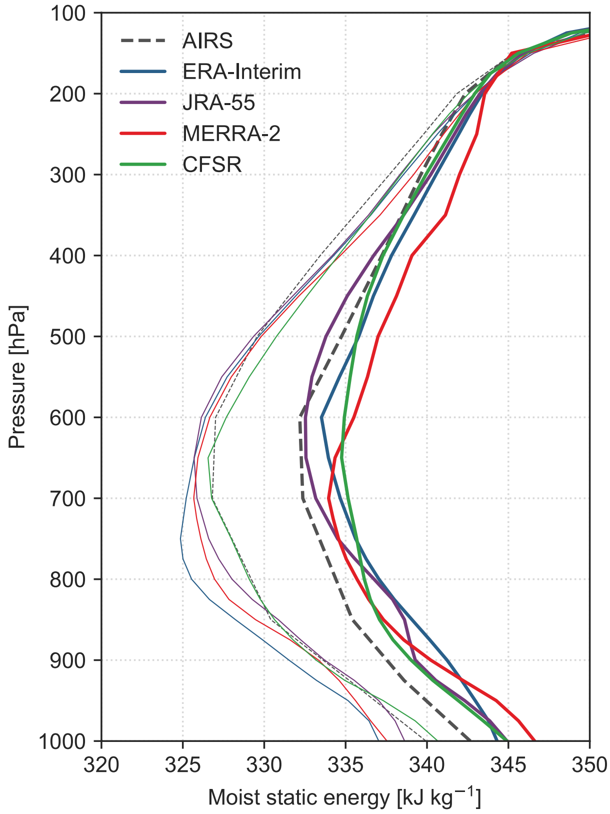


Figure 8.17: *Composite vertical profiles of moist static energy for ERA-Interim (blue), JRA-55 (purple), MERRA-2 (red), and CFSR (green) averaged for the upper (thick lines) and lower (thin lines) quartiles of daily-mean gridded LWCRE in the inner tropics (10 °S–10 °N) during 2001 - 2010. Profiles calculated from AIRS observations (September 2002 - December 2010) are shown as grey dashed line for context. Adapted from Wright et al. (2020).*

8.3.6 Temporal variability

Mean annual cycles of high cloud cover and OLR averaged over the inner tropics, NH subtropics, and SH subtropics (**Fig. 8.18**) show that ERA-Interim, ERA5, JRA-55, MERRA-2, and CFSR/CFSv2 all capture the main characteristics from observations. For the tropics, high cloud cover reaches a maximum in April and a minimum in August, although ERA-Interim and ERA5 show extended minima that span August and September and the minimum in MERRA is one month earlier than observed. The annual cycle in CFSR/CFSv2 has a smaller amplitude than indicated by observations, primarily due to a weaker minimum during boreal summer. The inner tropical latitude band omits many monsoon regions, so that the annual cycle of cloud fraction in the inner tropics depends largely on the migration of the ITCZ and the extent to which it passes out of the 10°S-10°N band during solstice seasons. The less pronounced annual minimum in CFSR may thus indicate weaker migration of the ITCZ rainband in this reanalysis, or artificial damping of the seasonal signal owing to positive biases in cloud cover outside convective regions and negative biases within convective regions (**Fig. 8.10**). JRA-55 and MERRA-2 produce larger amplitudes, but are otherwise qualitatively consistent with the observed annual cycle (**Fig. 8.18**), while ERA5 shows good agreement in both timing and amplitude despite its larger annual mean value. Annual cycles of cloud cover in the subtropics of both hemispheres are more consistent. Although the reanalyses tend to overestimate their amplitudes in these domains, it is unclear whether this results from issues in the reanalyses or shortcomings in the observational analyses, such as sampling biases or limited sensitivity to optically thin clouds. MERRA-2 also shows a more rapid increase of high cloud cover in the NH subtropics during boreal spring that is neither observed nor indicated by any other reanalysis. For OLR, the annual cycles are again broadly consistent with variations qualitatively opposite to those in cloud cover. MERRA and MERRA-2 consistently overestimate the observed amplitude and JRA-55 consistently underestimates the observed amplitude. ERA-Interim also produces weaker annual cycles than observed in the inner tropics and the SH subtropics, while the annual cycle based on CFSR/CFSv2 is again much weaker than observed in the inner tropics. Both the character and magnitude of monthly deviations from the annual mean are captured well by ERA5.

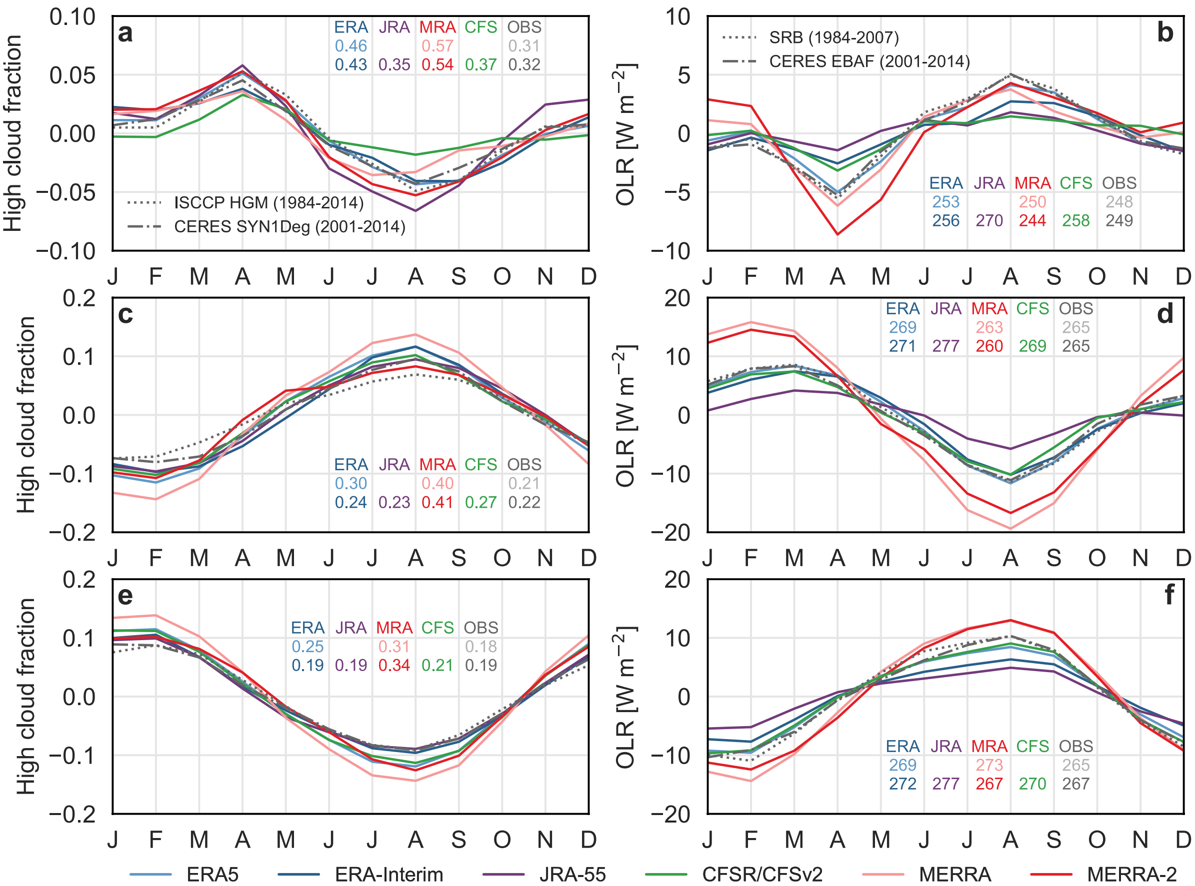


Figure 8.18: *Mean annual cycles of (a, c, e) high cloud fraction and (b, d, f) OLR as anomalies from the annual mean averaged over the (a-b) inner tropics (10°S–10°N), (c-d) NH subtropics (10°N-30°N) and (e-f) SH subtropics (30°S-10°S). Data are shown from ERA5, ERA-Interim, JRA-55, CFSR/CFSv2, MERRA-2, MERRA during 1980-2014, and observational estimates as indicated in the legend. Annual-mean reference values for each data product are listed in the corresponding panel, with lighter text for ERA5, MERRA, and ISCCP/SRB.*

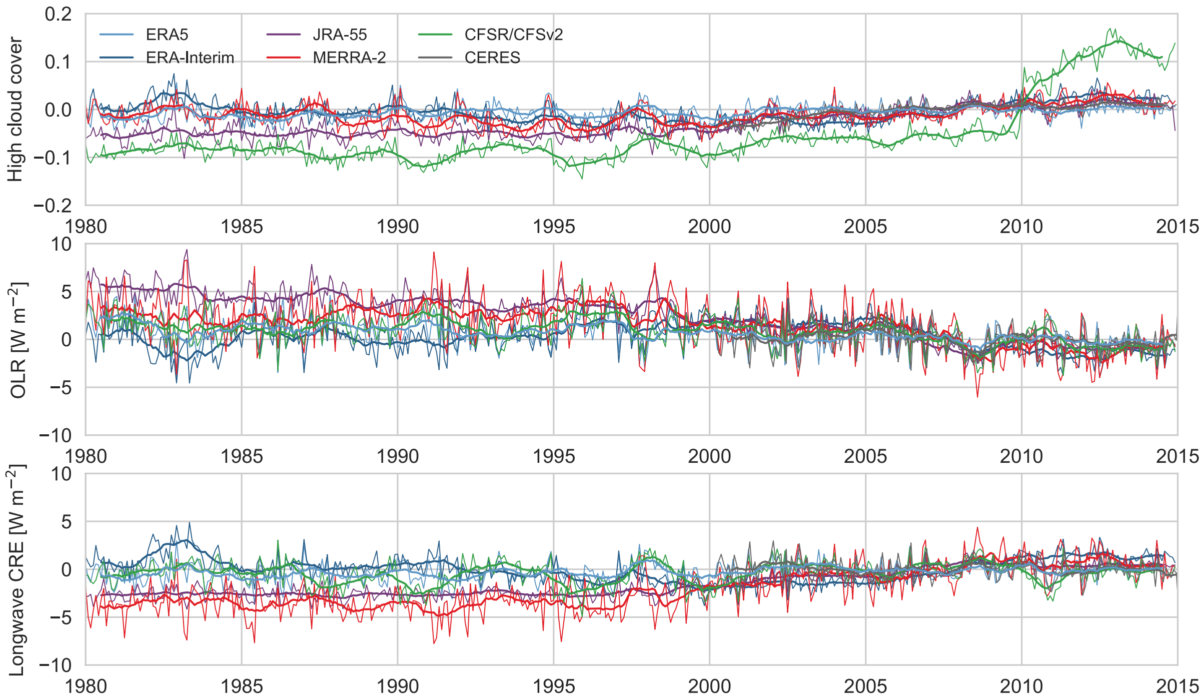


Figure 8.19: *Time series of monthly-mean anomalies of (a) high cloud fraction, (b) OLR, and (c) LWCRE averaged over the inner tropics (10 °S - 10 °N) relative to the mean annual cycle during 2001 - 2014. Data are shown for ERA5, ERA-Interim, JRA-55, MERRA-2, CFSR/CFSv2, and various CERES-based estimates (March 2000 - December 2014). CERES data are from the SYN1Deg product for high cloud fraction and from the EBAF product for OLR and LWCRE.*

Annual mean values listed in **Figure 8.18** confirm that maximum high cloud fractions and minimum OLR occur in the inner tropical band, with somewhat larger values of high cloud fraction and smaller values of OLR in the NH subtropics relative to the SH subtropics. This hemispheric asymmetry is stronger in cloud fraction than in OLR and is much more pronounced in MERRA and MERRA-2 than in the other data sets.

Figure 8.19 illustrates the long-term variability of monthly anomalies in high cloud fraction, OLR, and LWCRE in reanalyses and CERES-based observational products averaged over the inner tropics (10°S - 10°N). Anomalies are calculated relative to the mean annual cycle over 2001 - 2014. Variability in tropical mean high cloud fraction is primarily seasonal, with few robust signals at interannual time scales. One exception is transient increases in high cloud fraction and LWCRE coupled with decreases in OLR following the very strong El Niño events in 1982 - 83 and 1997 - 98, which are relatively robust among the reanalyses. However, the most pronounced variations in this figure appear to be artificial rather than physical. Most notably, the tropical-mean high cloud fraction in CFSR jumped suddenly by more than 10 percentage points between the end of 2009, when CFSR was initially planned to end, and the beginning of 2010. Tropical-mean high cloud fraction then increased again at the beginning of 2011 with the transition to CFSv2, approaching a value (0.54) close to that produced by MERRA-2 (0.56 for 2011 - 2014). The bridge year 2010 is not well documented, but has also been found to show discontinuities in other variables (*e.g.*, stratospheric water vapour; *Davis et al.*, 2017).

Discontinuities in the CFSR time series are not limited solely to the CFSR/CFSv2 transition, with transient reductions in tropical-mean high cloud fraction after every production stream transition in the initial 1979 - 2009 run (1 January 1987, 1990, and 1995; 1 April 1999 and 2005; see also *Chapter 2*, **Table 2.24**). However, although these latter stream-related discontinuities are reflected in OLR and LWCRE (as is the transition at the beginning of 2010), neither OLR nor LWCRE shows large changes following the transition to CFSv2 in January 2011. This peculiar feature is discussed in detail in *Appendix B* of *Wright et al.* (2020), along with possible reasons for the jump in high cloud fraction. Although it has been suggested that CFSv2 can serve as an extension of CFSR, researchers should use this reanalysis with caution in any study that spans the 2010 bridge year or the 2011 transition to CFSv2. JRA-55 shows a gradual increase in high cloud fraction from 1980 to the early 2000s, along with corresponding changes in OLR (towards smaller values) and LWCRE (towards larger values). Although the signs of these trends are reproduced across most of the reanalyses (*Wright et al.*, 2020, their **Fig. 14**), the relatively strong changes in JRA-55 bring values of all three variables closer to those in other reanalyses by the later part of the record. Despite this improvement, biases in tropical-mean OLR from JRA-55 relative to ERA-Interim and CFSR/CFSv2 remain large (~10 W m^{-2}, reduced from ~15 W m^{-2} in the 1980s). MERRA-2 shows qualitatively similar drifts in OLR (toward smaller values) and LWCRE (towards larger values). However, the lack of any corresponding change in high cloud fraction suggests that other factors (such as changes in cloud top height or cloud water path) must be responsible for these drifts.

With the notable exception of the initial five years in ERA-Interim, both the ERA-Interim and ERA5 time series are relatively stable, with no major long-term drifts.

Wright et al. (2020) also summarize relative variance and cross-correlations in deseasonalized anomalies among the reanalyses and observationally based data sets examined in this section (their **Fig. 15**). Noting that ERA5 and MERRA-2 are most consistent with observations in terms of tropical-mean OLR and LWCRE (**Fig. 8.14**), we focus on these two products here. Whereas ERA5 produces the strongest correlations against observationally based data sets among the five examined reanalyses, correlations based on MERRA-2 are consistently among the weakest. Likewise, while ERA5 captures the magnitude of observed variance in OLR and LWCRE well, MERRA-2 overestimates variance in both. Taking all factors into account (no major drifts or jumps, consistently high correlations, and standard deviations and seasonal cycles close to observationally based benchmarks), ERA5 provides the best representation of temporal variability in tropical OLR and LWCRE among recent reanalyses.

8.3.7 Key findings and recommendations

Key findings

- Tropical high cloud fields are substantially different among reanalyses, with tropical-mean high cloud fractions ranging from 25 % (JRA-55) to 43 % (MERRA-2) and cloud water contents (CWCs) in the upper troposphere spanning more than a factor of 10. However, simulated cloud satellite products based on ERA-Interim and MERRA-2 indicate that both reanalyses reproduce high cloud fractions as observed by passive satellite instruments well despite differences of nearly 15 % in tropical-mean high cloud fraction. (*Section 8.3.2*)

- Observed vertical profiles of cloud fields in the tropical UT are in better agreement with models that produce pronounced convective anvils in both cloud fraction and CWC (ERA5, MERRA-2, CFSR/CFSv2) than with those that do not (ERA-Interim, JRA-55), although issues with the altitude and extent of deep convective detrainment remain to be resolved. (*Section 8.3.3*)

- Differences in high cloud fields project directly onto differences in OLR. MERRA-2 produces the largest tropical-mean longwave cloud radiative effect (LWCRE) and smallest tropical-mean OLR among the evaluated reanalyses, while JRA-55 produces the smallest tropical-mean LWCRE and a tropical-mean OLR approximately 10 - 15 $W\,m^{-2}$ larger than any other evaluated reanalysis. Comparison with observations suggests that the larger time-mean LWCRE and smaller time-mean OLR produced by MERRA-2 are more realistic. ERA-Interim, ERA5, and CFSR/CFSv2 underestimate clear-sky OLR relative to CERES observations (suggesting a high bias in GHG absorption) even as they overestimate all-sky OLR (a low bias in LWCRE). (*Section 8.3.4*)

- Details of model physical parameterizations (*e.g.*, clouds and convection) can have systematic impacts not only on forecast variables (*e.g.*, diabatic heating), but also on analysed variables (*e.g.*, temperature and specific humidity) and derived variables that rely on them (*e.g.*, moist static energy). These effects are illustrated by biases in moist static energy in the JRA-55 lower troposphere (related to restrictions in convective cloud base) and in the MERRA-2 tropical upper troposphere (related to the representation of convective anvil clouds) relative to other reanalyses. (*Section 8.3.5*)

- Despite some differences in amplitude and timing, reanalyses generally reproduce annual cycles of high cloud fraction and OLR averaged over the tropics and subtropics. However, interannual variations in these variables show drifts and discontinuities that appear to arise mainly from changes in assimilated observations and/or production streams rather than physical factors. Among reanalysis estimates of tropical high cloud cover and OLR, ERA5 shows greater stability in time (1980 - 2014), as well as stronger correlations and smaller standard deviations relative to observations (2001 - 2014). This stability may be surprising in light of other key findings in this report regarding temporal variability in ERA5, such as evident discontinuities of global averaged temperature in the middle and upper stratosphere (*Chapter 3*). (*Section 8.3.6*)

Key recommendations

- Despite suggestions that CFSv2 can serve as an extension of CFSR, discontinuities in clouds and other products mean that researchers should use this reanalysis with caution in any study that spans the 2010 bridge year or the 2011 transition to CFSv2 (see also *Chapter 2*, *Section 2.5*). (*Section 8.3.6*)

- Long-term drifts in high cloud fraction, OLR, and LWCRE are present in almost all reanalyses, and show little agreement in terms of sign, timing, or magnitude. These products should generally not be used for trend or time series analysis without independent verification. Among the reanalyses, ERA5 shows greater stability in time and stronger correlations with observed variability for these cloud and radiation metrics, and may therefore offer a more reliable characterization of long-term variations in these metrics relative to earlier reanalyses. (*Section 8.3.6*)

- Evaluation of co-variability between high cloud fraction and other variables shows that the separate treatment of anvil and in situ large-scale clouds in GEOS-5 (as applied in MERRA-2) produces some unrealistic behaviours, particularly with respect to radiative transfer. A revised prognostic treatment of cloud condensate may be necessary to resolve these issues. (*Sections 8.3.4 - 8.3.5*)

- Differences in the parameterizations of clouds and convection imprint not only on the distributions of clouds and other forecast fields, but also on metrics that are directly affected by the data assimilation, such as the vertical profile of moist static energy in the tropical atmosphere. These differences can often be traced back to assumptions made in the parameterization and may thus present viable targets for model improvement. Data users should be alert to potential impacts of these issues on the generation and interpretation of reanalysis-based diagnostics. (*Section 8.3.5*)

8.4 Diabatic heating rates

Diabatic heating rates or temperature tendencies are useful diagnostics of reanalysis behavior and performance. These heating rates are virtually impossible to measure directly (although some components can be inferred from observations, as discussed in *Sect. 8.8.6*) and are therefore poorly constrained. In reanalyses these terms are influenced to some extent by the impacts of observational data assimilation on temperature, moisture, winds, and other variables, but they still differ substantially across reanalyses (*Wright and Fueglistaler*, 2013). The magnitude and distribution of diabatic heating within the TTL provide insight into the circulation of this region, and can help to diagnose the sources and characteristics of differences in this circulation amongst reanalyses.

Diabatic heating is a fundamental component of the temperature budget, as expressed via the thermodynamic energy equation:

$$\frac{\partial T}{\partial t} + v \cdot \nabla \mathrm{T} - \omega\left(\frac{\kappa \mathrm{T}}{\mathrm{p}} - \frac{\partial T}{\partial p}\right) = \frac{Q}{c_p} \qquad (8.1).$$

We use the temperature form of the thermodynamic equation here for consistency with reanalysis diabatic heating diagnostics, which are reported as tendencies in temperature (T) rather than potential temperature (θ). The three terms on the left-hand side are the time rate of change, the horizontal advection (v = <u,v> the horizontal wind vector), and the vertical advection including adiabatic effects (p pressure, ω pressure vertical velocity, and $\kappa = R/c_p$, with R the gas constant and c_p the specific heat of air at constant pressure). These terms are balanced on the right-hand side by diabatic heating (Q/c_p). Diabatic heating is often separated into contributions from different physical processes as follows:

$$\frac{Q}{c_p} = \frac{Q_{rad}}{c_p} + \left(\frac{Q_{mst}}{c_p} + \frac{Q_{mix}}{c_p}\right) \qquad (8.2).$$

Here Q_{rad}/c_p represents diabatic heating due to radiative transfer, Q_{mst}/c_p represents heating due to moist physics, and Q_{mix}/c_p represents heating due to shear–flow (turbulent) mixing. The latter two terms are not always provided separately for reanalyses, and are therefore often combined into a single "residual" term, which represents heating due to non-radiative physics. The turbulent mixing term is a non-negligible component of the residual near the tropopause, but is orders of magnitude smaller than the moist physics term in most of the tropical upper troposphere (*Wright and Fueglistaler*, 2013).

Diabatic temperature tendencies in reanalyses are computed by tracking the evolution of temperature before and after physical parameterizations are applied. For example, the radiative heating rate over a forecast represents the cumulative changes calculated by the radiation parametrization over all radiation time steps included in that forecast, while the heating due to moist physics includes not only latent heating and cooling associated with the phase changes of water, but also heat transport that occurs within parameterized convection. Heating due to moist physics can be decomposed into terms due to convection and large-scale condensation, while convective heating can be further decomposed into terms due to deep and shallow convection. It is important to emphasize that the diabatic heat budget is not closed in reanalyses: energy is not conserved. This lack of closure occurs because the data assimilation step can cause changes in temperature that add or remove heat from the system. We can think of this assimilation increment as a separate "diabatic" term in the thermodynamic energy equation (*e.g.*, Q_{assim}/c_p). The assimilation increment may be useful for identifying biases in the atmospheric model, but its interpretation is complicated. Biases that are corrected by the assimilation may originate in one or more of the diabatic terms (*e.g.*, radiation or convection), but they may also originate from errors in the temperature advection terms or unknown biases in the observations. The role of the assimilation increment (and the lack of closure that it implies) is important to keep in mind, but we do not examine it here. Please also see the footnote on diabatic heating rates in reanalyses in *Section 12.1.3*.

This Section extends the intercomparison presented by Wright and *Fueglistaler* (2013) in both temporal coverage and reanalyses examined. Specifically, we add results for ERA5, JRA-55 and MERRA-2. We also add some new metrics, particularly with respect to variability in the LZRH, and incorporate some new methodological approaches following *Zhang et al.* (2017).

8.4.1 Total diabatic heating

Figure 8.20 shows zonal-mean estimates of total diabatic heating based on eight reanalyses: two from ECMWF (ERA-Interim and ERA5), two from JMA (JRA-55 and JRA-25), two from NASA GMAO (MERRA and MERRA-2), and two from NCEP (CFSR and NCEP-NCAR R1). Diabatic terms were not archived for the CFSv2 (*i.e.*, post-2011), so all comparisons are conducted for the period 1980 - 2010. Among the newer estimates aligned along the upper row, ERA-Interim and JRA-55 have strong similarities, as do MERRA-2 and CFSR. ERA5, in the lower row, shows stronger similarities with MERRA-2 and CFSR than with ERA-Interim and JRA-55.

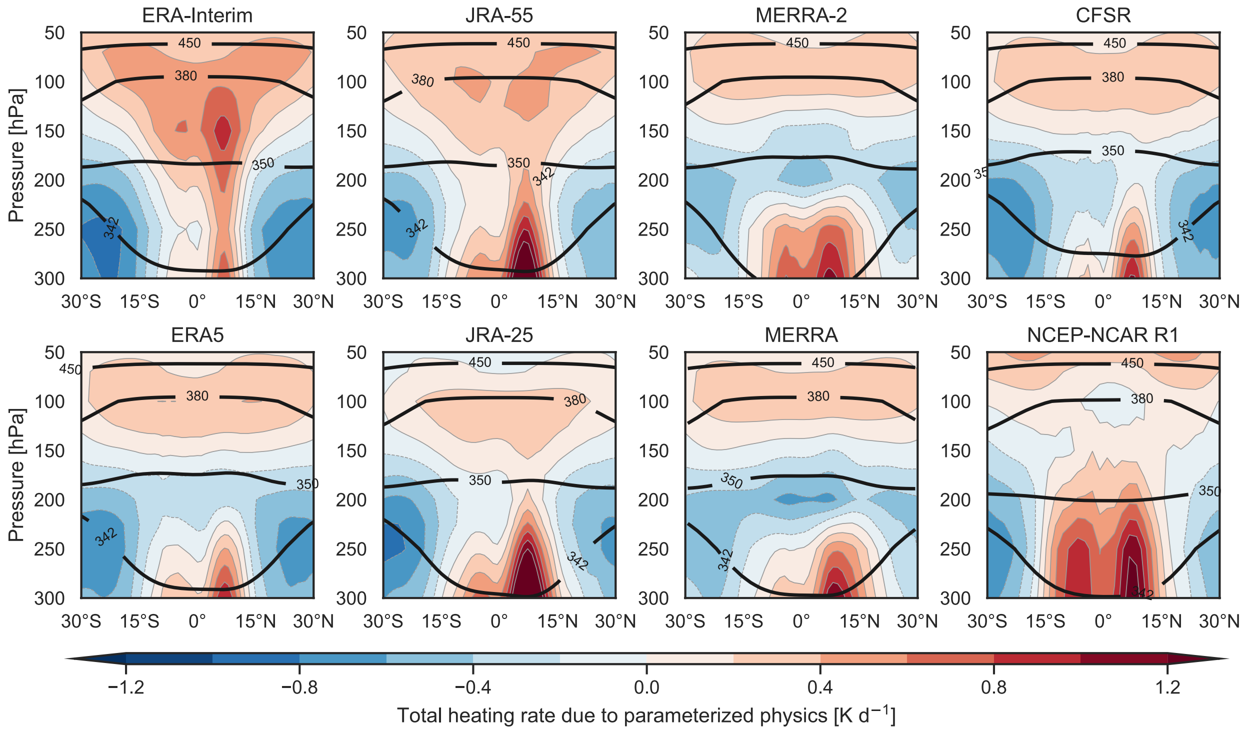


***Figure 8.20:** Zonal mean total diabatic temperature tendencies (Q/c_p in K day^{-1}; shading and gray contours) and potential temperature (θ in K; black contours) averaged over 1980 - 2010 for two generations of reanalyses from ECMWF (far left), JMA (center left), NASA GMAO (center right), and NOAA NCEP (far right). Updated from Wright and Fueglistaler (2013).*

All five systems show relatively strong positive heating rates in the inner tropics near 300 hPa. The largest time-mean values at this level are located near 5 - 10 °N and are associated with latent heating in the ITCZ, especially during NH summer (see also **Fig. A8.4**). Secondary maxima in the SH indicate the effects of seasonal migrations in the ITCZ averaged across longitudes (**Fig. 8.20**), particularly its zonal-mean position during SH summer (see also **Fig. A8.4**). The most pronounced difference among the reanalyses is the diabatic 'chimney' that extends upward across the 350 K isentropic surface (~190 hPa) in the zonal-mean distributions based on ERA-Interim and JRA-55 (**Fig. 8.20**). This feature is missing from the zonal-mean distributions based on MERRA-2, CFSR, and ERA5. The time-mean cooling at this level in the latter two systems is physically unreasonable in the sense that it implies a net downward mass flux across the 350 K isentropic surface that lacks a compensating return flow (diabatic heating rates at 350 K outside the subtropics are also negative in the time mean; not shown), and is also inconsistent with diabatic heating rates diagnosed from the thermodynamic equation (*e.g.*, **Fig. 8.33**). CFSR does include seasonal chimneys of diabatic ascent across this layer (**Fig. A8.4**), as does ERA5 (not shown); however, MERRA-2 does not. Although both ERA-Interim and JRA-55 contain diabatic 'chimney' features within the tropical UT, the mechanisms behind this feature differ between the two reanalyses. Whereas the chimney in JRA-55 is primarily convective in origin (as discussed in the context of **Fig. 8.26** below), it is aided considerably by radiative effects (especially cloud radiative effects) in ERA-Interim (see discussion of **Figs. 8.23** and **8.24** below). Other important differences include the magnitude of heating within the TTL, which is much larger in ERA-Interim than in any other reanalysis, and the latitudinal structure of heating in the LS, which shows a pronounced 'V'-shaped structure in ERA-Interim and JRA-55 that is much weaker in ERA5, MERRA-2, and CFSR.

There are evident improvements in the diabatic heating distributions between the earlier reanalyses JRA-25, MERRA, and NCEP-NCAR R1 and their more recent counterparts in the upper row of **Figure 8.20**. For example, a layer of spurious negative heating rates in the LS of JRA-25/JCDAS has been eliminated in JRA-55, the negative heating rates centered at 200 hPa in MERRA are still present but less intense in MERRA-2, and several problematic features in NCEP-NCAR R1 have been eliminated in CFSR (see also discussion of diabatic heating in ERA-40 relative to ERA-Interim by *Fueglistaler et al.*, 2009b). We focus mainly on the more recent reanalyses included in **Figure 8.20** (ERA-Interim, ERA5, JRA-55, MERRA-2, and CFSR) in the following discussion.

8.4.2 Radiative heating

Radiative heating rates represent net convergence of energy in the form of radiation. These are often decomposed into separate terms due to LW and SW radiative transfer, as these parts of the spectrum are treated separately in the model physics (*Chapter 2*; **Table 2.4**; see also **Fig. 2.2** and additional discussion in *Chapter 2E*).

Some systems provide a further decomposition into all-sky and clear-sky radiative heating rates, which allows a deeper look at the influence of clouds in the diabatic heat budget.

One useful paradigm for understanding differences (and fluctuations) in the distribution of radiative heating is the Newtonian cooling approximation, which approximates radiative heating or cooling (Q_{rad}/c_p) as a constant relaxation rate α times the difference between the actual temperature T and a radiative equilibrium temperature T_{eq}:

$$\frac{Q_{rad}}{c_p} \approx -\alpha \cdot \left(T - T_{eq}\right) \quad (8.3).$$

This equation indicates that, all else remaining equal, an increase in temperature results in enhanced radiative cooling (or reduced radiative heating), while a decrease in temperature results in enhanced radiative heating (or reduced radiative cooling). It also indicates that radiative heating rates can change due to changes in the equilibrium temperature. This equilibrium temperature depends on the composition and thermodynamic structure of the atmosphere throughout the vertical column. For example, the radiative equilibrium temperature at the cold point tropopause may vary due to the presence or absence of clouds in the column below it, or due to differences in the types and/or characteristics of clouds when they are present. The radiative equilibrium temperature might also change due to variations in ozone or other radiatively active constituents. We use the Newtonian cooling approximation to examine potential causes of some key differences in radiative heating among the reanalyses.

Zonal mean radiative heating

Figure 8.21 shows zonal-mean time-mean distributions of total radiative heating and its LW and SW components based on ERA-Interim, JRA-55, MERRA-2, and CFSR. All four distributions show radiative heating in the lower stratosphere overlying radiative cooling in the upper troposphere, but with important differences in both the spatial distributions and the magnitudes of certain features.

Starting from lower altitudes and moving upward, we find that LW cooling in the upper troposphere is stronger in ERA-Interim than in MERRA-2, with ERA5, JRA-55 and CFSR falling between these two. This difference in LW cooling between ERA-Interim and MERRA-2 is exacerbated by differences in SW heating.

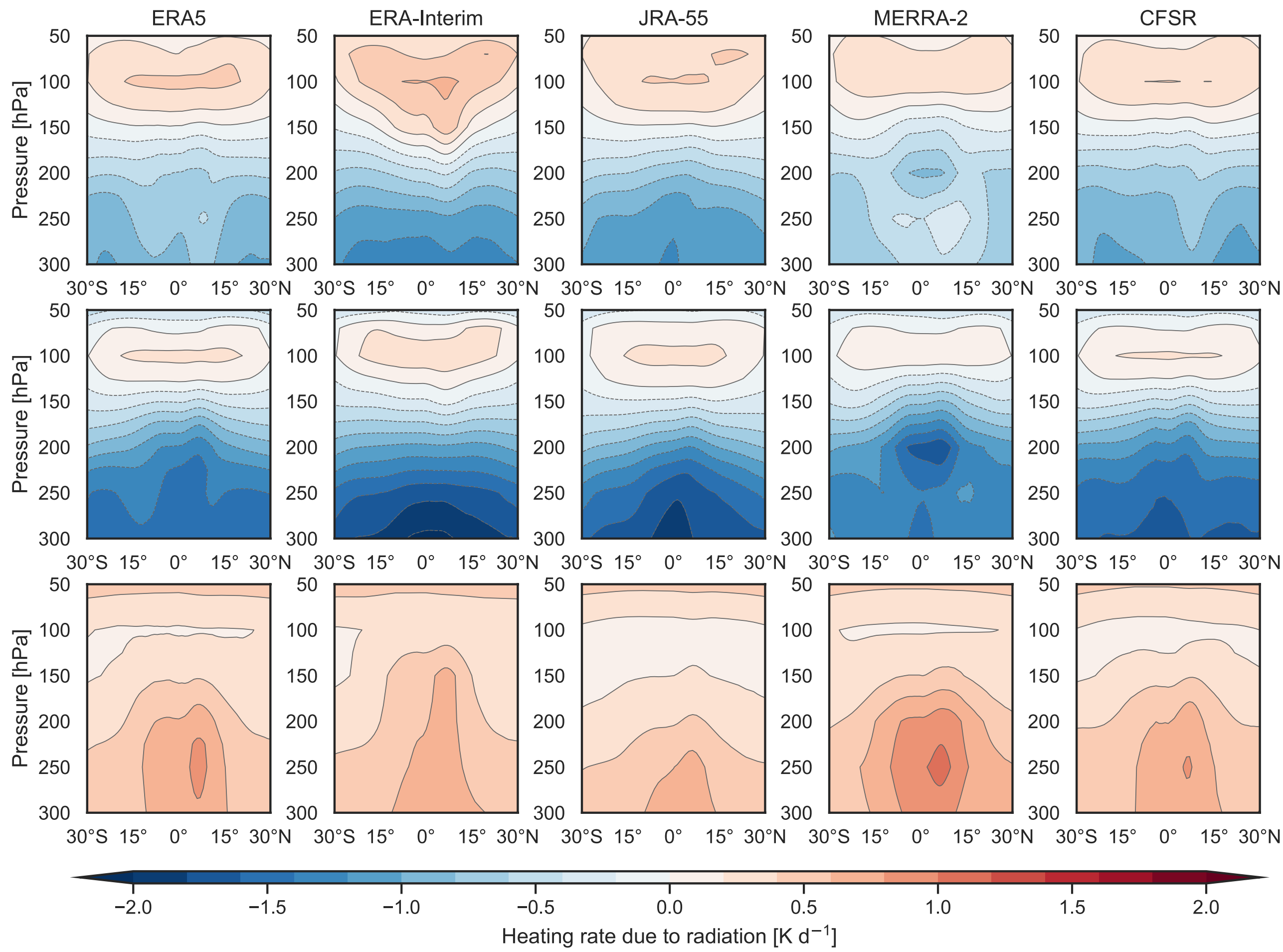


***Figure 8.21:** Zonal mean total radiative heating (Q_{rad}/c_p in K day^{-1}; top) and its LW (center) and SW (bottom) components in ERA5, ERA-Interim, JRA-55, MERRA-2, and CFSR for 1980 - 2010. Updated from Wright nd Fueglistaler (2013).*

SW heating partially offsets LW cooling at these altitudes in all four reanalyses but is evidently stronger in MERRA-2 than in the others, particularly in the inner tropics below 200 hPa. These differences can be explained to some extent by differences in composition: in particular, MERRA-2 has larger concentrations of water vapour in the tropical UT than ERA-Interim (**Fig. 8.22**). Larger concentrations of water vapour will tend to enhance SW absorption, and may also enhance LW absorption relative to emission (depending on conditions in the overlying column). However, the main reason behind these discrepancies is differences in cloud fields. MERRA-2 includes thicker and more extensive anvil clouds in the UT (**Figs. 8.10**, **8.12**, and **8.13**), which enhances SW absorption in the cloud layer, as well as LW absorption below the anvil layer and LW emission above. The latter (enhanced LW emission near the anvil layer top) produces the inner tropical maximum in LW cooling at 200 hPa in MERRA-2, which is not seen in ERA-Interim and is much more pronounced than in JRA-55 or CFSR. This strong LW cooling centered at 200 hPa is associated with enhanced LW emission from the tops of convective anvil clouds in the tropical UT. We discuss these cloud radiative effects in more detail in the following subsection.

Moving upward through the lower TTL we reach the LZRH, which separates net radiative cooling in the tropical UT from net radiative heating in the tropical LS. Differences in the LZRH are treated in more detail later in this Section. For now, we note only that the zero contour is distended downward toward the equator in ERA-Interim but upward in MERRA-2, while it is approximately isobaric in ERA5, JRA-55 and CFSR (**Fig. 8.21**). These differences again involve both LW and SW components. LW cooling in the tropical UT transitions more rapidly to LW heating with increasing height in ERA-Interim, in contrast to the strong LW cooling above anvil cloud tops in MERRA-2. Meanwhile, the tropical maximum in SW heating extends slightly higher in altitude in ERA-Interim than in MERRA-2, despite the larger SW heating rates below 200 hPa in the latter.

Differences among the reanalyses remain substantial in the LS. The strongest radiative heating is found in ERA-Interim, followed in decreasing order by ERA5, JRA-55 and CFSR, while the weakest is found in MERRA-2. Both the magnitudes of diabatic heating and the vertical location of maximum heating rates within the LS have pronounced seasonal cycles that also differ among the reanalyses (**Fig. A8.5**). Several studies have reported that these differences have significant impacts on transport statistics and the rate of ascent in the tropical LS inferred from Lagrangian trajectory simulations (*e.g.*, *Tao et al.*, 2019; ; *Abalos et al.*, 2015; *Schoeberl et al.*, 2012; see also *Sect. 8.5*). These differences are contributed primarily by the LW component in the layer centered around the CPT. LW heating rates within this layer are strongest in ERA-Interim and weakest in MERRA-2, with JRA-55 and CFSR again falling in the middle. The origins of these differences appear to be more varied than those in the UT, although differences in cloud fields again play a role. Thicker and more extensive high cloud cover in MERRA-2 reduces the upwelling flux of LW radiation from the troposphere, which in turn lowers the radiative equilibrium temperature. Following the Newtonian cooling approximation outlined above, this reduced upwelling flux will reduce LW heating rates around the CPT, thus providing a plausible explanation for the relatively weak LW heating in MERRA-2 (which has the strongest LWCRE; **Fig. 8.14**) and the much stronger LW heating in JRA-55 (which has the weakest LWCRE). We return to this idea in the following Section. Considering again the Newtonian cooling approximation, differences in the local temperature may be influential. Among these four reanalyses, the ERA-Interim CPT is coldest by around 0.2 ~ 0.4 K on average (**Fig. 8.5**), consistent with stronger LW heating rates assuming similar radiative equilibrium temperatures (**Fig. 8.21**). Differences in composition, both at the level in question and elsewhere in the column, may also play a role in determining the radiative equilibrium temperature. For example, JRA-55 produces larger values of ozone mixing ratio within the tropical LS (**Fig. 8.22**; see also *Chapter 4*, **Fig. 4.16**) than do the other three reanalyses examined in this Section.

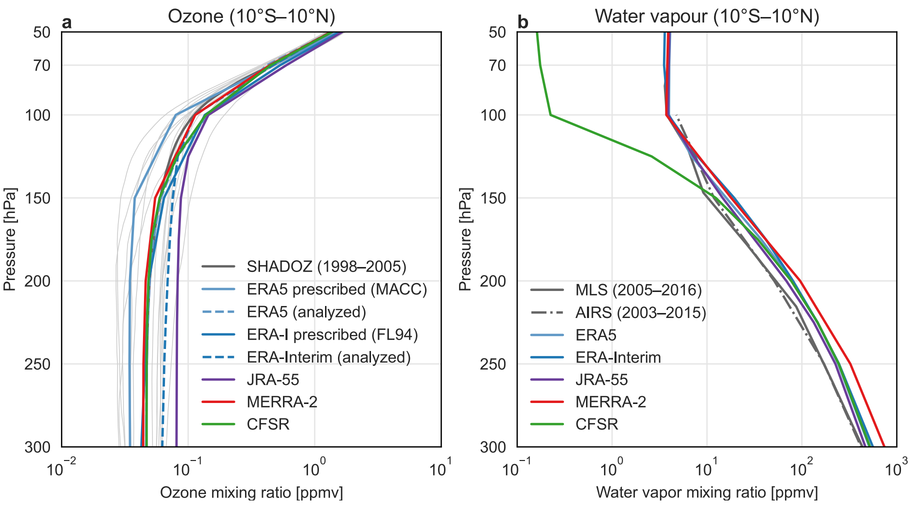


Figure 8.22: *Mean vertical profiles of (a) ozone and (b) water vapour averaged within 10°S-10°N. In addition to analyzed ozone and water vapour from ERA5, ERA-Interim, JRA-55, MERRA-2, and CFSR averaged over 1980-2010, the prescribed ozone climatologies used in ERA5 (based on the MACC reanalysis climatology for 2003-2011), and ERA-Interim (Fortuin and Langematz, 1994) are included in (a) along with observational estimates of ozone volume mixing ratios from SHADOZ; and water vapour volume mixing ratios from Aura MLS and AIRS are included in (b). The thick grey line in (a) indicates the climatological mean ozone profile averaged from observations at 13 SHADOZ sites in the tropics between 1998 and 2005 (Fueglistaler and Fu, 2006; Thompson et al., 2003); thin grey lines indicate climatological mean profiles at individual SHADOZ sites.*

This difference in local ozone loading would tend to increase the radiative equilibrium temperature and thus intensify LW heating assuming similar local temperatures.

Cloud effects on radiative heating

Among the reanalyses considered in this study, only the ECMWF and NASA GMAO systems provide vertically-resolved estimates of radiative heating under clear-sky conditions. Clear-sky radiative heating rates and cloud radiative effects based on ERA5, ERA-Interim and MERRA-2 are shown in **Figure 8.23** (distributions for the earlier ERA-40 and MERRA reanalyses are similar to ERA-Interim and MERRA-2, respectively). Although the clear-sky radiative heating rates are more consistent between these reanalyses than the all-sky radiative heating rates shown in **Figure 8.21**, there remain some important differences. For example, the clear-sky LZRH is shifted upward in ERA-Interim relative to MERRA-2 (**Fig. 8.23**). Clear-sky radiative cooling in the upper troposphere and clear-sky radiative heating in the stratosphere are also enhanced in ERA-Interim relative to MERRA-2, with ERA5 intermediate between these two. In ERA-Interim, clouds cause radiative heating throughout the upper troposphere, with a maximum impact around 150 hPa, where cloud fraction is also at a maximum (**Fig. 8.12**). The distribution in ERA5 is qualitatively similar but shifted downward toward higher pressures, with the zero-line near 125 - 150 hPa. In MERRA-2, by contrast, clouds cause radiative heating in the lower part of the upper troposphere (200 - 300 hPa) but radiative cooling in the upper part (100 - 200 hPa) (**Fig. 8.23**). This dipole is centered on the anvil layer (**Fig. 8.12**), and indicates that clouds enhance absorption in the lower part of the anvil, where CWC is large, and enhance emission in the upper part of the anvil, where cloud fraction remains large but CWC declines sharply (**Fig. 8.13**). Clouds act to reduce radiative heating in the lower stratosphere (50 - 100 hPa) in all three reanalyses (**Fig. 8.23**). This can be understood as clouds reducing the upwelling LW flux from the troposphere, which in turn reduces the convergence of LW radiation in the lower stratosphere. As mentioned above, this effect is most pronounced in MERRA-2 (see also *Tao et al.*, 2019).

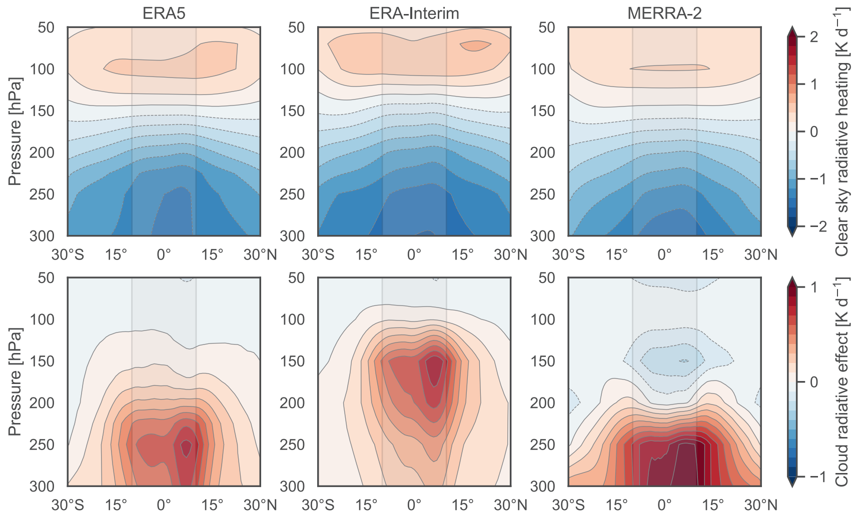


Figure 8.23: *Comparison of zonal mean clear-sky radiative heating rates (top; contour interval 0.2 K day^{-1}) and cloud radiative effects (bottom; contour interval: 0.1 K day^{-1}) in the tropical UTLS based on ERA5, ERA-Interim, and MERRA-2 during 1980 - 2010.*

To extend this analysis to include JRA-55 and CFSR, we construct composite mean profiles of radiative heating rates conditional on the four quartiles of LWCRE in each reanalysis. This is an adaptation of an approach employed previously by *Zhang et al.* (2017), who composited heating rates on quantiles of OLR rather than LWCRE (results are similar for both approaches; *Wright et al.*, 2020). **Figure 8.24** shows these composite profiles for the period 2001 - 2010, separated into total, LW, and SW radiative heating.

Among these five reanalyses, cloud effects on radiative heating rates are weakest in ERA-Interim through most of the tropical UTLS (except for the 100 - 200 hPa layer) and strongest in MERRA-2. The results for these two reanalyses are basically consistent with those based on **Figure 8.23**, with cloud impacts on radiative heating rates in MERRA-2 qualitatively opposite to those in ERA-Interim through much of the upper troposphere. The response in ERA-Interim is concentrated between 100 hPa and 200 hPa, where radiative heating rates are evidently enhanced by the presence of high clouds (**Fig. 8.24**). At lower altitudes in the upper troposphere (200 - 400 hPa), cloud-induced increases in SW heating are effectively balanced by cloud-induced increases in LW cooling in this reanalysis. ERA5, JRA-55 and CFSR show only weak cloud impacts on total radiative heating at pressures less than 175 hPa. This insensitivity of total radiative heating rates reflects a near-complete compensation between enhanced LW cooling and enhanced SW heating at these altitudes. Cloud-related perturbations in the LW and SW components extend upward to around 100 hPa in CFSR, but to only around 150 hPa in JRA-55. MERRA-2 produces the largest cloud impacts on radiative heating rates. Indeed, direct comparison of cloud radiative effects between MERRA and MERRA-2 (not shown) indicates that cloud radiative impacts in MERRA are further amplified in MERRA-2, consistent with an increase in tropical mean CWC in the upper troposphere between MERRA and MERRA-2 (**Fig. 8.13f**). The effects of high clouds in MERRA-2 are to reduce radiative heating rates in the 100 - 200 hPa layer (largely due to enhanced LW cooling, partially offset by enhanced SW heating), and increase radiative heating rates at pressures larger than 200 hPa (**Fig. 8.24**). The latter is the result of enhanced SW heating near the top of the anvil layer (200 - 250 hPa) and enhanced LW heating near the base of the anvil layer (300 - 350 hPa), again taking the MERRA-2 profile of tropical-mean CWC (**Fig. 8.13 f**) as a guide.

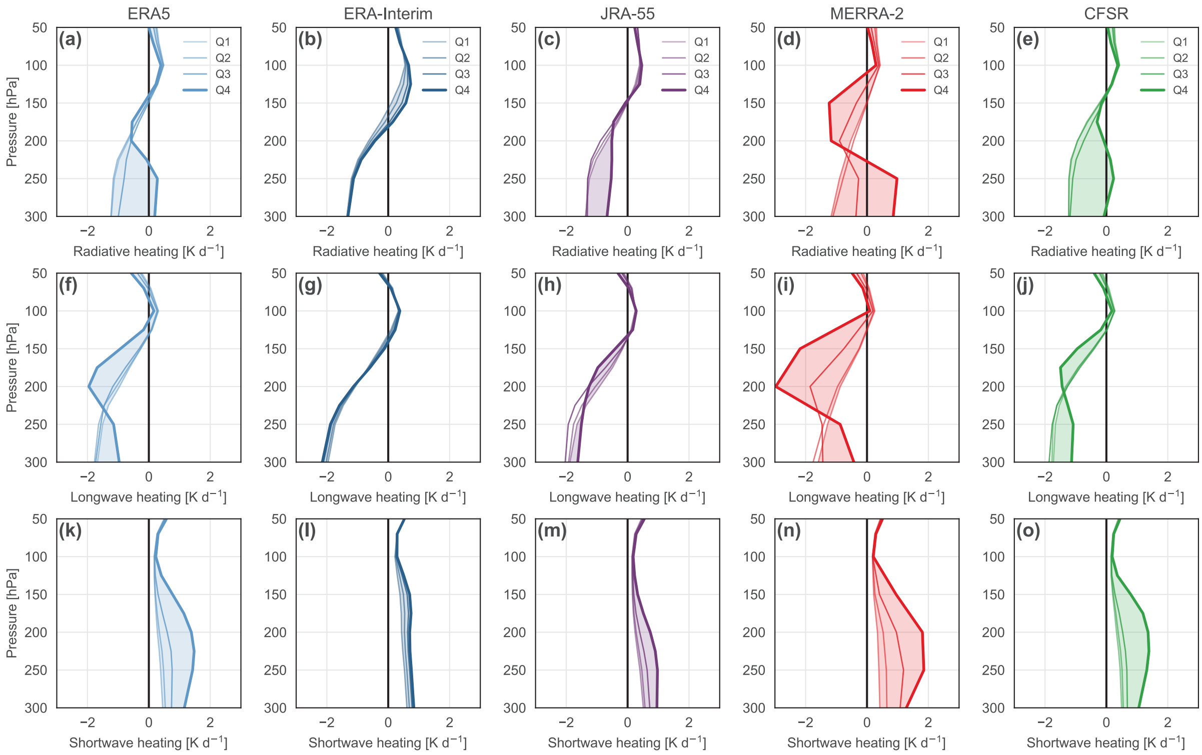


***Figure 8.24:** Composite mean profiles for total (top), LW (center), and SW (bottom) radiative heating on the four quartiles of LWCRE for the ERA5 (far left; light blue), ERA-Interim (center left; dark blue), JRA-55 (center; purple), MERRA-2 (center right; red), and CFSR (far right; green) reanalyses within the inner tropics (10 °S - 10 °N) during 2001 - 2010. Q1 represents daily gridded heating rates for which the LWCRE is in the lowest 25 % of all daily gridded values (i.e., predominantly clear sky). Q2 and Q3 represent the lower middle and upper middle quartiles, respectively, while Q4 represents heating rates for which the associated LWCRE exceeds the 75th percentile (corresponding to extensive high cloud cover; **Fig. 8.15**). Adapted from Wright et al. (2020).*

Level of zero net radiative heating (LZRH)

Differences in the radiative impacts of clouds in the tropical upper troposphere can in turn translate into differences in transport through the tropical tropopause layer. One commonly-used metric in this regard is the LZRH, which marks the boundary between negative radiative heating rates (corresponding to net descent across isentropic surfaces) in the tropical troposphere and positive radiative heating rates (corresponding to net ascent) in the tropopause layer and lower stratosphere (**Fig. 8.1**; *Gettelman et al.*, 2002; *Folkins et al.*, 1999). We identify this level by using linear interpolation of daily-mean gridded radiative heating rates in ln(p) to determine the zero crossing. We further require that radiative heating rates remain positive above the identified LZRH to at least the 70 hPa isobaric level. **Figure 8.25** shows distributions of the LZRH based on each reanalysis.

Differences in the LZRH distributions are largest between ERA-Interim and MERRA-2. Neglecting the influence of clouds, the primary mode of the ERA-Interim LZRH distribution (~140 hPa) is shifted to slightly higher altitudes than that in MERRA-2 (~150 hPa). The altitudes of these primary modes reflect the vertical locations of the clear-sky LZRH in each system (**Fig. 8.23**). The more striking distinction between ERA-Interim and MERRA-2 concerns the impacts of clouds on the LZRH altitude (blue and red distributions in **Fig. 8.25**). Whereas clouds tend to lower the LZRH in ERA-Interim (to around 170 ~ 180 hPa on average), clouds significantly raise the LZRH in MERRA-2 (to around 110 hPa). This difference has important implications for the efficiency of mass and constituent transport from the convective detrainment level (200 ~ 300 hPa) into the tropical lower stratosphere ($p < 100$ hPa). In MERRA-2, the cloudy and clear-sky modes of the distribution are almost completely distinct, suggesting that transport regimes in the tropical upper troposphere are approximately binary in this model. By contrast, the breadth of the LZRH distribution based on ERA-Interim (and especially the breadth of the distribution associated with the largest values of LWCRE) indicates that ERA-Interim produces a broad spectrum of cloudy states (**Fig. 8.25**). This diagnostic thus helps to clarify the environmental conditions associated with the two very different tropical mean cloud water content profiles in **Figure 8.13f**, with the pronounced anvil layer in MERRA-2 in sharp contrast to the gradual decrease of cloud water content with height in ERA-Interim. Distributions of the LZRH location based on ERA5, JRA-55, and CFSR are more consistent with each other (**Fig. 8.25**). Each distribution has one major mode, although the LZRH tends to be shifted to a slightly higher altitude in CFSR (~135 hPa) than in ERA5 (~140 hPa) or JRA-55 (~150 hPa).

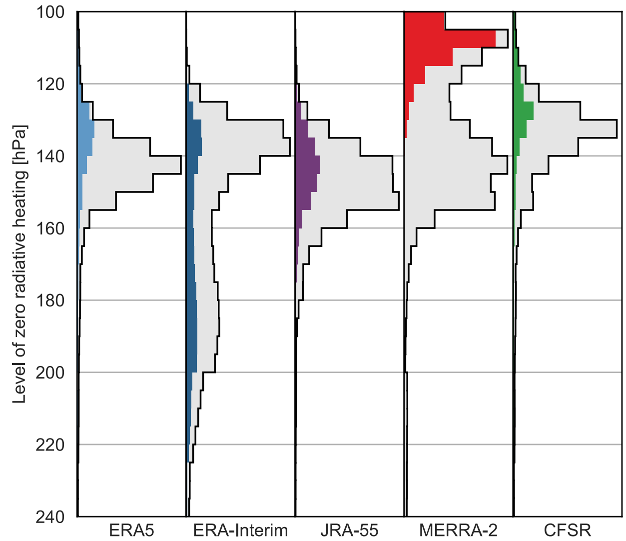


***Figure 8.25:** Histograms of daily-mean gridded LZRH locations in the pressure vertical coordinate within the inner tropics (10°S-10°N) during 2001-2010. Light grey shading indicates distributions for all daily-mean gridded samples. Colored shading in each column indicates distributions for the upper quartile of LWCRE (corresponding to Q4 in **Fig. 8.24**) based on the corresponding reanalysis dataset (see **Fig. 8.15**). Adapted from Wright et al. (2020).*

ERA5, JRA-55, and CFSR all indicate a slight upward shift toward lower pressures (by ~5 hPa) in the location of the LZRH for the largest values of LWCRE, much less than that indicated by MERRA-2 but still opposite in sign to that indicated by ERA-Interim.

Although cloud effects raise the LZRH in most of the reanalyses, results based on applying radiative transfer models to observed cloud distributions suggest that cloud effects should lower the LZRH (*e.g., Yang et al.*, 2010; *Fueglistaler and Fu*, 2006; *Corti et al.*, 2005). This disagreement appears to arise from a combination of the reanalyses locating the peak positive SW effect at lower altitudes and overestimating the negative LW effect relative to the observationally-based estimates (**Fig. 8.24**; *cf., Yang et al.*, 2010, their **Fig. 10**). The lower vertical location of cloud-induced SW heating could indicate that the reanalyses underestimate the depth of convective anvil clouds. This is a known problem in MERRA-2 (*A. Molod, personal communication*), although it is not immediately evident from **Figures 8.12** and **8.13** whether similar issues affect the other reanalyses and to what extent. An overestimated LW effect could result from systematic underrepresentation of thin cirrus and their radiative effects within the TTL (*e.g., Corti et al.*, 2005), especially as we represent cloud effects here in terms of the relative magnitude of the LWCRE.

8.4.3 Non-radiative heating

Through most of the troposphere, non-radiative heating is dominated by latent heating associated with precipitation. These effects remain important within the lower part of the UTLS, but approach zero around and above the tropopause. **Figure 8.26** shows the residual (non-radiative) component of the total temperature tendencies from ERA-Interim, JRA-55, MERRA-2, and CFSR. The two peaks in tropical heating corresponding to the zonal mean locations of the ITCZ during solstice seasons (near 5°S and between 5°N and 10°N) are again readily identifiable. The major discrepancies concern the depth of the heating, which are broadly consistent with the vertical distributions of cloud fields in these reanalyses (**Figs. 8.12-8.13**). The depth of strong residual heating is shallowest in MERRA-2 (**Fig. 8.26**), consistent with the extensive anvil layer at relatively low altitudes in this reanalysis (**Figs. 8.12-8.13**). Heating is also relatively shallow in ERA5, for which the anvil layer is only slightly deeper than that in MERRA-2, and extends progressively deeper in CFSR, ERA-Interim, and JRA-55, consistent with the greater heights associated with convective anvils in these systems.

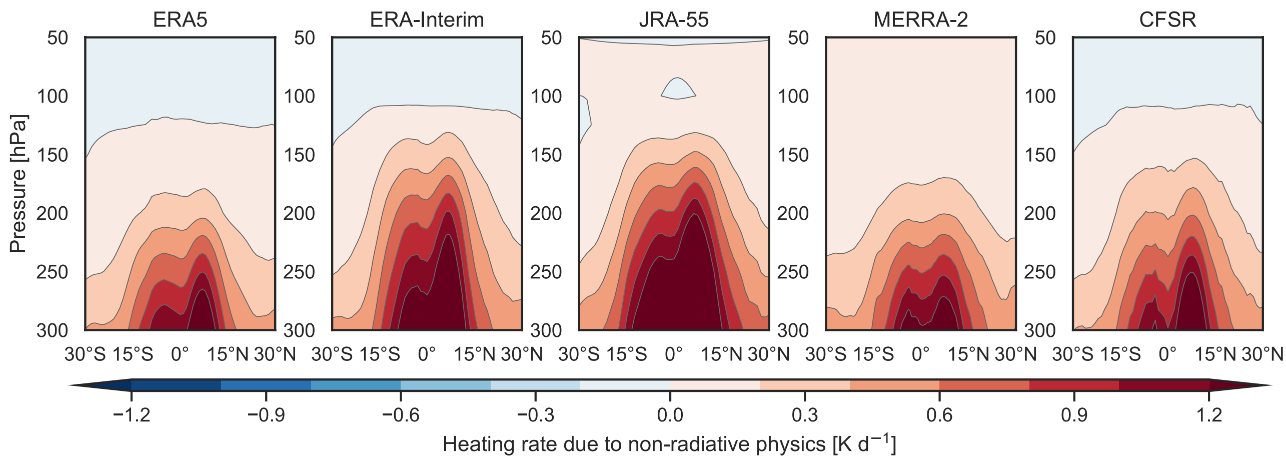


***Figure 8.26:** Zonal-mean time-mean residual (non-radiative) temperature tendencies [in K day^{-1}] from ERA5, ERA-Interim, JRA-55, MERRA-2, and CFSR over 1980-2010. The residual terms are calculated as total heating rates (**Fig. 8.20**) minus radiative heating rates (**Fig. 8.21**), and include moist physics, parameterized turbulence, and any other physics that are implemented in ways that can directly affect the heat budget (e.g., gravity wave drag). Updated from Wright and Fueglistaler (2013).*

Together with cloud radiative effects, differences in these terms are an important contributor to differences in total diabatic heating in the lower TTL (**Fig. 8.20**): shallower latent heating coupled with enhanced cloud-top LW cooling creates the diabatic 'transport barrier' that emerges in MERRA-2 and, to a lesser extent, ERA5 and CFSR (see also discussion of MERRA by *Wright and Fueglistaler*, 2013).

Figure 8.27 shows zonal-mean time-mean temperature tendencies from parameterized mixing in ERA-Interim, JRA-55, MERRA-2, and CFSR (*Chapter 2*, **Table 2.8**; see also **Fig. 2.4** and further discussion in *Chapter 2E*). Although ERA5 and ERA-Interim do not directly provide separate moist physics and vertical mixing terms, it is possible to infer turbulent mixing due to shear-flow instability from offline calculations (*Flannaghan and Fueglistaler*, 2011). This inferred heating due to turbulent mixing in ERA-Interim is larger than the heating due to parameterized mixing in the other reanalyses, with stronger cooling between 10 hPa and 50 hPa and stronger warming between 200 hPa and 100 hPa. However, it remains about an order of magnitude smaller than the radiative terms through most of this vertical range, and the residual term (**Fig. 8.26**) is evidently dominated by contributions from moist physics rather than parameterized mixing in the UT. We have not performed this calculation for ERA5.

The dipole patterns seen in ERA-Interim, CFSR, and (to a lesser extent) JRA-55 imply mixing of the upper troposphere with the lower stratosphere in the inner tropics. This mixing has pronounced zonal asymmetries, and often shows a maximum above the tropical Indian Ocean (see **Fig. 8.60**; and also **Fig. A8.10**). The pattern in CFSR is similar to that in ERA-Interim, but weaker in amplitude. The qualitative similarity between these two reanalyses likely arises because both models use modified versions of the LTG (*Louis et al.*, 1982; *Louis*, 1979) vertical diffusion scheme in the free atmosphere. The difference in magnitude likely relates to how the mixing coefficient is specified in the upper troposphere. This coefficient has been reported to be unrealistically large above the boundary layer in ERA-Interim (*Bechtold et al.*, 2008), while that in CFSR was reformulated specifically to mitigate extremely strong turbulent mixing at upper levels in NCEP-NCAR R1 (*Wright and Fueglistaler*, 2013; *Saha et al.*, 2010) The pattern in JRA-55 is substantially different, with cooling at the tropopause and weak warming above, coupled with warming in the subtropical LS in both hemispheres. Heating rates due to parameterized mixing in MERRA-2 are several orders of magnitude smaller than those in the other three reanalyses, as diffusion coefficients in MERRA-2 are very small above the atmospheric boundary layer (*Chapter 2*, **Table 2.8**).

8.4.4 Key findings and recommendations

Key findings

- There are large differences among reanalysis diabatic heating products within the TTL, which are known to influence transport statistics and rates of ascent in trajectory simulations of cross-tropopause transport in this region. Differences among reanalysis diabatic heating rates in the tropical UTLS are not limited to any one component: longwave, shortwave, and non-radiative components all show substantial discrepancies. (*Section 8.4*)

- Differences in radiative heating rates primarily trace back to the differences in cloud fields, but there are important discrepancies in clear-sky radiative heating as well. In many cases, these discrepancies can be explained by systematic differences in composition and temperature structure. (*Section 8.4.2*)

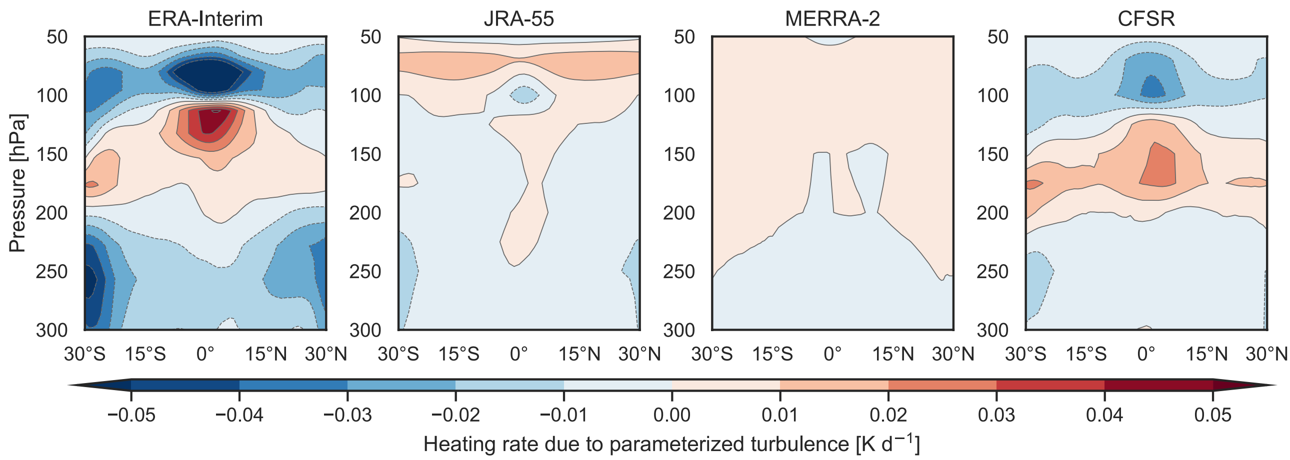


Figure 8.27: *Zonal-mean time-mean temperature tendencies due to parameterized turbulence, Qmix [in K day^{-1}], in ERA-Interim, JRA-55, MERRA-2, and CFSR. The tendency terms for JRA-55, MERRA-2, and CFSR are averaged over 1980 - 2010 from archived data. (There was no turbulence term available for ERA5.) The tendency term for ERA-Interim is not archived by ECMWF and has been estimated from an offline calculation (Flannaghan and Fueglistaler, 2011) of 6-hourly analysis temperatures and winds using the revised Louis scheme as detailed in the ECMWF IFS documentation, part IV. The ERA-Interim result is averaged over 2001 - 2010; results are not sensitive to the chosen period and can be taken as representative.*

- Discrepancies in heating due to parameterized turbulent mixing are very poorly constrained. These terms may be influential near the tropopause, especially when radiative heating rates are small, though they are typically several orders of magnitude smaller than other terms in the diabatic heat budget. (*Section 8.4.3*)

Key recommendations

- Given large differences in reanalysis diabatic heating products and related metrics within the tropical UTLS, researchers using these fields to drive or nudge model simulations of this region should use multiple reanalyses whenever possible. (*Section 8.4*)

8.5 Transport

Transport through the TTL controls the entrainment of tropospheric air into the stratosphere (*Highwood and Hoskins*, 1998). As discussed in *Section 8.4*, the TTL encompasses the level of zero radiative heating, which marks the transition from negative to positive heating rates and creates a barrier for the large-scale transport into the stratosphere (*Folkins et al.*, 1999). Above the LZRH, vertical motion balances the radiative heating according to thermal balance and air is slowly ascending (*Section 8.5.3*). The lower TTL is penetrated by deep convection which becomes increasingly rare with altitude (*Liu and Zipser*, 2005), while the vertical motion outside of convective towers is weak. Quantifying transport paths across the TTL for a better understanding of the chemical composition of air entering the stratosphere is often done based on CTMs and Lagrangian models driven by meteorological reanalyses. Studies have focused on the stratospheric dehydration point (*Bonazzola and Haynes*, 2004; *Fueglistaler et al.*, 2004) and the residence time of air through the TTL (*Krüger et al.*, 2009). *Sections 8.5.1* and *8.5.2* analyze how these two quantities are represented in the different reanalysis data sets.

8.5.1 Dehydration point distribution

It has been known since Brewer's seminal work on stratospheric circulation that tropical tropopause temperature is the key driver of stratospheric water vapor (H_2O) concentration (*Brewer*, 1949). As parcels approach and pass through the cold point tropopause, condensation occurs, thereby regulating the parcel's H_2O concentration to local saturation levels (*e.g.*, *Holton and Gettelman*, 2001; *Fueglistaler et al.*, 2009a). The dehydration process thus primarily depends on the air parcel temperature history.

The details of the transport and dehydration process can be understood by performing Lagrangian trajectory simulations, which track the temperature history of a large number of individual air parcels. The approach applied here is based on a forward trajectory model, following the details described in *Schoeberl and Dessler* (2011), with trajectories calculated using the Bowman trajectory code (*Bowman et al.*, 2013; *Bowman*, 1993). In the forward trajectory mode, the number of trajectories that contribute to dehydration events in a particular geographic region depend on the circulation and temperature structures of the respective reanalysis. We conduct diabatic Lagrangian runs in isentropic coordinates. The parcel initiation level is chosen to be the 370 K isentrope, which is generally above the level of zero radiative heating in the tropics but below the tropical tropopause (~375 - 380 K; see **Fig. 8.1**). In the TTL, water vapor is conserved along the trajectories except when saturation occurs. Water vapor excess is instantaneously removed from the parcel to keep the relative humidity with respect to ice from exceeding 100 %. Along each trajectory, we define the point with the lowest temperature and minimum saturation mixing ratio as the final dehydration point (FDP). The FDP determines the final H_2O mixing ratio of each trajectory as it enters the stratosphere and is equivalent to the Lagrangian cold point. Details of the trajectory model, the setup of the simulations and the FDP calculations are given in *Wang et al.* (2015) and *Schoeberl et al.* (2013). The trajectory model is driven by meteorological reanalyses on model levels, except for CFSR where, due to availability at the time, the model was driven by data on pressure levels.

The distribution of FDP temperatures and frequencies derived from trajectory simulations driven by modern reanalyses for 2007 - 2010 are shown in **Figure 8.28**. The Lagrangian cold point temperatures (black and white contour lines) show strong deviations, with ERA-Interim having the lowest and MERRA the highest dehydration temperatures among the model level data sets. Trajectories driven by CFSR data on pressure levels show unrealistically warm cold points consistent with the Eulerian cold point comparison (*Section 8.2*). Despite different background temperatures for the different reanalysis data, the dehydration patterns given by their frequency distribution agree quite well. The strongest dehydration occurs over the tropical western Pacific, South America, and Africa, where frequent deep convection leads to a cooling above, which results in a higher and colder tropopause. The Asian monsoon during summertime is another important dehydration center (see *Sections 8.8* and *8.8.6*).

Figure 8.29 shows the evolution of the FDP distribution as a function of latitude. The largest occurrence frequencies migrate northward from boreal winter to summer, with the most intense dehydration occurring during the NH winter season when the tropopause is coldest. However, different reanalyses demonstrate different seasonal changes in FDP frequencies. These differences are caused by the combined effects of differences in the circulation and differences in the background temperatures. For example, the run driven by ERA-Interim shows less dehydration events between May and August than JRA-55.

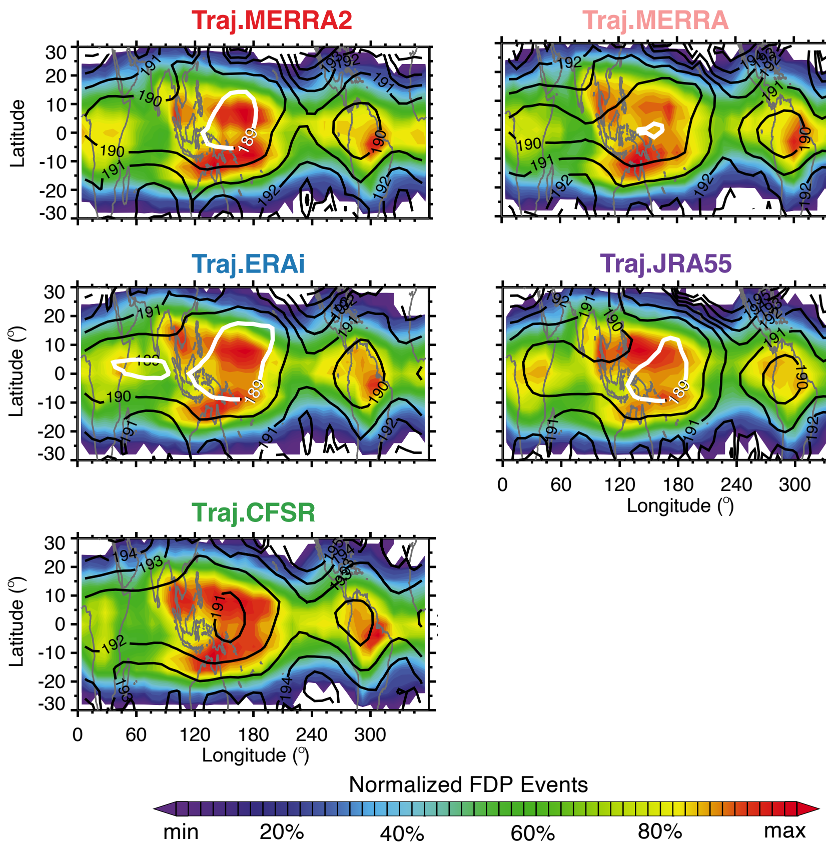


Figure 8.28: *Distribution of annual mean final dehydration points (FDPs) derived from trajectory model simulations driven by different reanalyses for 2007-2010. For convenience of intercomparing the different reanalyses with largely varying total FDP events, we show the percentiles of the FDP event distribution. Temperatures associated with final dehydration are shown as black contours at 1 K intervals, with the 189 K isotherm highlighted in white.*

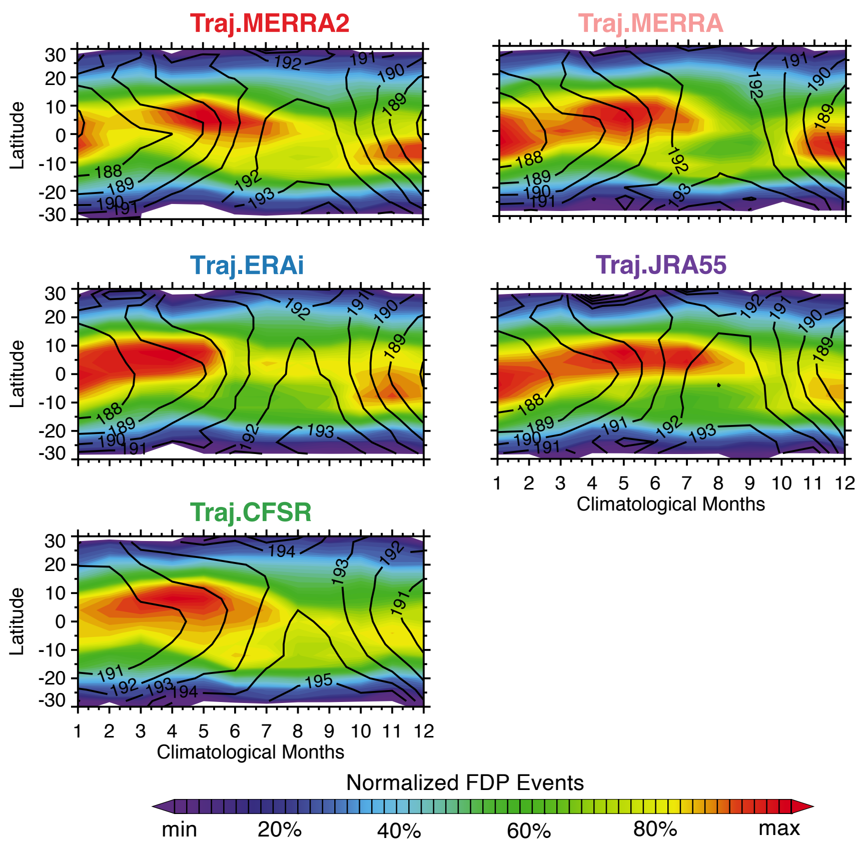


Figure 8.29: *Like Figure 8.28 but for the seasonal cycle of final dehydration points (FDPs).*

8.5.2 TTL residence time

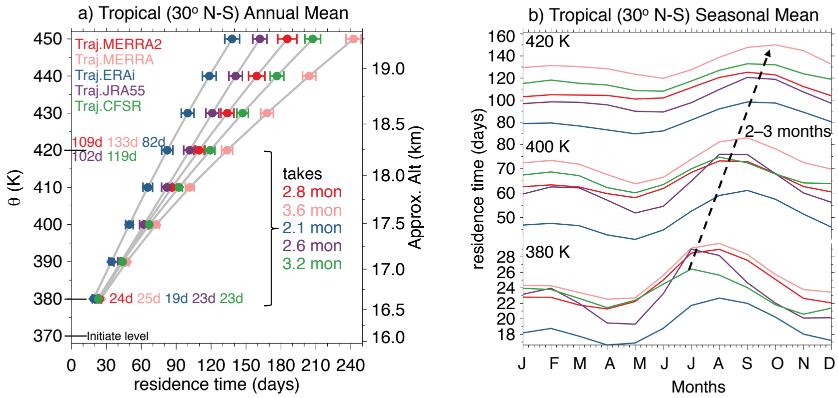


Figure 8.30: *Tropical (30°N-30°S) a) annual mean and b) seasonal mean residence times derived from trajectories driven by modern reanalyses in the upper TTL. All residence times are for transport from the 370K initiation level to the specified isentropic surface.*

One of the advantages of trajectory modeling is that it retains the history of each individual parcel. For an upward moving parcel released at a fixed isentropic surface, we can examine the time the parcel takes to ascend to a specified higher isentropic level for the first time. We refer to this as the residence time (τ) of that parcel in the layer between the two isentropic levels. Based on the trajectory runs described in *Section 8.5.1*, we calculate the residence times for air mass transport between 370 K and 450 K according to the five reanalyses. The mean profile of residence time quantifies the speed of the upwelling branch of the Brewer-Dobson Circulation (*Chapter 5*). Note, however, that atmospheric mixing is not taken into account in the trajectory calculations presented here, which can impact the total transport velocity and the water vapour tape recorder upwelling.

Figure 8.30a shows the residence time averaged over the tropics (30 °N - S) starting from the 370 K level for 2005 - 2010. Only 23 - 25 days are required for newly-initiated parcels to ascend across the tropopause (~ 380 K). One exception is ERA-Interim, which has larger heating rates in the TTL (*Wright and Fueglistaler*, 2013; *Wang et al.*, 2014; *Section 8.4*) and thus relatively rapid parcel ascent (only 19 days) across the tropopause. Within the TTL, parcels stay for ~ 3 months or longer when using the MERRA, MERRA-2, or CFSR circulations, whereas parcels only stay for ~ 2 months when using the ERA-Interim or JRA-55 circulations. Below the 370 K potential temperature level, MERRA-2 diabatic heating rates are often negative and cannot be used to drive tropical upwelling simulations (see also *Section 8.4*). Overall, the vertical range and seasonality for residence time based on different reanalyses is in qualitative agreement with previous studies on residence time (*Ploeger et al.*, 2010; *Krüger et al.*, 2009) and trace gas seasonality in the TTL (*Ploeger et al.*, 2012). The residence time shows a seasonal dependence (**Fig. 8.30b**), with parcels ascending faster (slower) during boreal winter-spring (summer-fall), thus resulting in shorter (longer) residence times. Deviations of the seasonal cycle are most pronounced around the cold point tropopause, where the amplitude of the seasonal cycle based on JRA-55 is more than twice as large as that based on CFSR.

Figure 8.31 shows horizontal distributions of the 370 K - 380 K residence time from different trajectory runs during the boreal winter and summer seasons averaged over 2005 - 2010. All trajectory runs show the tropical Western Pacific and the Asian monsoon as two distinct centers of strong ascent during boreal winter and summer, respectively. Differences in residence time depend on the magnitudes of total diabatic heating rates among the different reanalyses, with broad agreement that heating rates from ERA-Interim are the largest overall among these five reanalyses (see *Section 8.4*). Apart from the overall differences, the spatial distribution of the residence times also varies among the reanalyses.

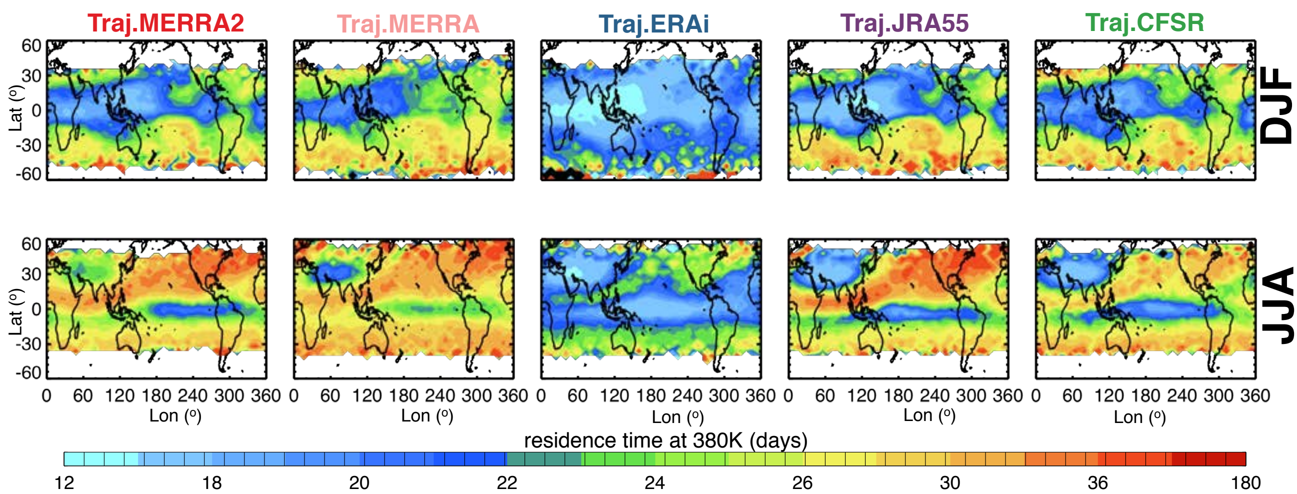


Figure 8.31: *Regional differences of residence time at 380 K (started from 370 K), driven by modern reanalyses during DJF (first row) and JJA (second row) of 2005 - 2010.*

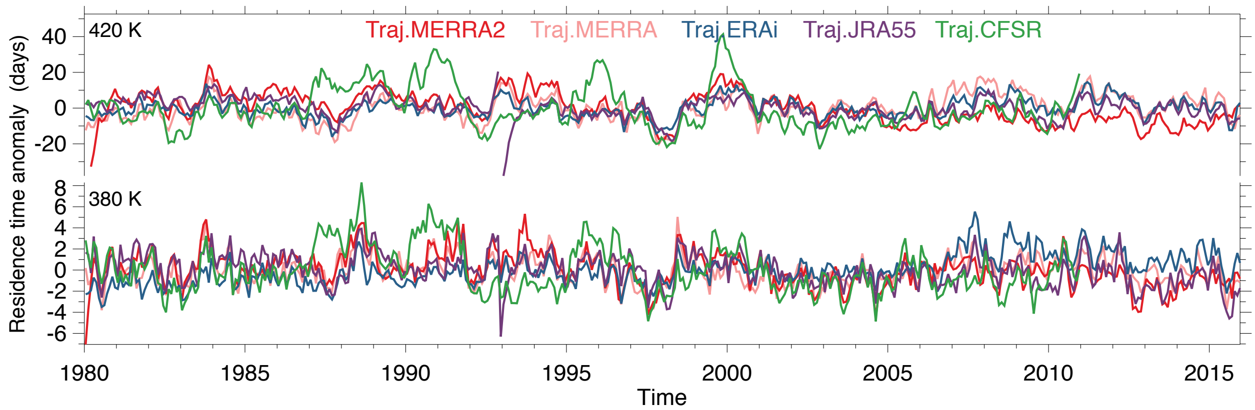


Figure 8.32: *Tropical (30°N-30°S) residence time anomalies (by removing annual cycle) at 380K and 420K (both started from 370 K), for 1980-2015 derived from trajectories driven by modern reanalyses. CFSR only extends to 2010.*

One of the most apparent differences is that ERA-Interim shows a much weaker equator-to-subtropics gradient in residence time during JJA than any other reanalysis considered here.

Generally, all runs produce clear annual cycles of residence time as shown in **Figure 8.30b**, although the details differ. **Figure 8.32** compares interanual anomalies derived by substracting the annual cycle of tropical residence times at 380K and 420K during 1980-2015. Examined in anomaly space, all runs yield similar interannual variability of residence time, mostly characterized by short-term fluctuations with no apparent long-term changes. This is consistent with the study by *Krüger et al.* (2009), who found a significant anticorrelation of TTL residence time with planetary wave driving in the extratropical lower stratosphere. Larger fluctuations are evident in the run driven by CFSR, which has distinct maxima in some years. These maxima probably result from artefacts of the stream transitions which started on 1 January in 1987, 1995, and 2010 and on 1 April in 1999 and 2005. Sudden drops in the CFSR heating rates in the lower stratosphere (~83hPa) occur in 1987, 1990, 1995, 1999, 2005, and 2010, consistent with the signal in the residence time shown here. While cold point temperature anomalies show step like improvements in inter-reanalysis agreement around 1998-1999 and 2006, the same is not true for the residence time. This result demonstrates that vertical transport driven by diabatic heating rates is less impacted by changes in the assimilated observational data sets. Note, however that this conclusion is limited to the TTL (substantial discontinuities in heating rates appear at higher altitudes around the TOVS-ATOVS transition; see, *e.g.*, *Abalos et al.*, 2015) and does not mean that heating rates should be considered reliable in this region. Indeed, heating rates in the TTL based on different atmospheric reanalyses show substantial disagreements in both climatology (*Section 8.4*; see also *Wright and Fueglistaler*, 2013) and trends (*e.g.*, *Linz et al.*, 2019).

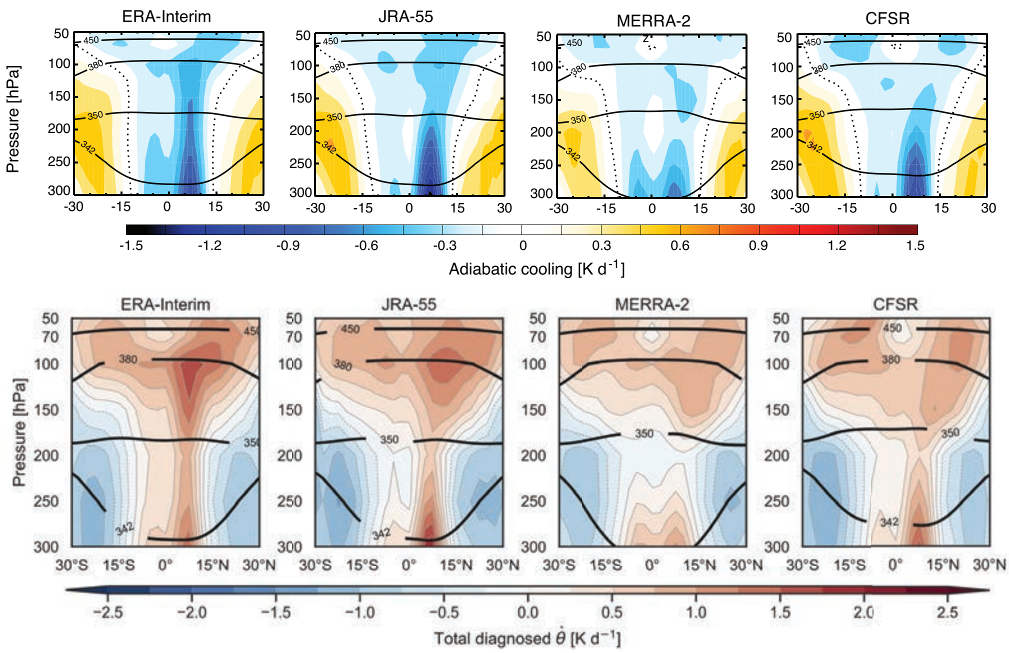


Figure 8.33: *Upper panels: annual mean adiabatic temperature tendencies due to vertical advection by w*. Lower panels: annual mean total diabatic temperature tendencies diagnosed from analyzed winds and temperatures as described by Martineau et al. (2018).*

8.5.3 TTL tropical upwelling

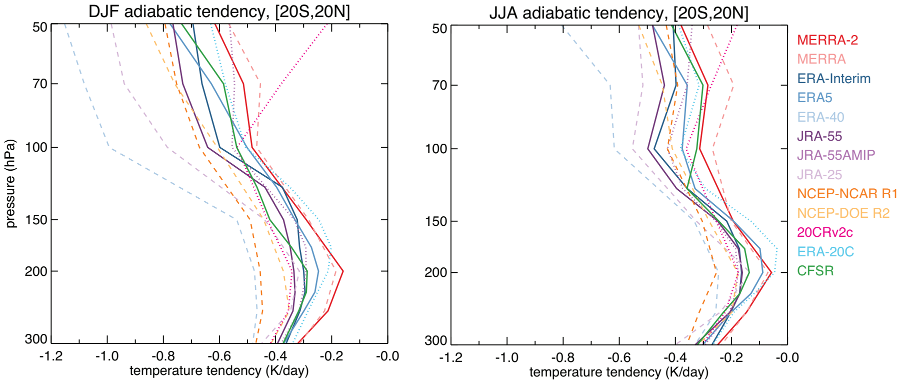


Figure 8.34: *Tropical (20 °S – 20 °N) mean adiabatic temperature tendencies due to vertical advection by w^* for DJF (left) and JJA (right). MERRA-2, ERA-Interim, ERA5, JRA-55 and CFSR are shown as solid lines, MERRA, ERA-40, JRA-25, NCEP-NCAR R1, and NCEP-DOE R2 as dashed lines and JRA-55 AMIP, 20CRv2c and ERA-20C as dotted lines.*

The residual mean upwelling in the Brewer-Dobson Circulation leads to adiabatic cooling in the tropics (see *Chapter 5*). Adiabatic cooling then leads in turn to radiatively-driven diabatic heating that pulls temperatures back toward radiative equilibrium. In principle, we expect in the TTL an approximate balance between diabatic heating and adiabatic cooling in the climatological mean. Near the bottom of the TTL, latent heating may contribute to diabatic heating, while clouds may impact the radiation budget up to the cold point (see also *Sect. 8.4*). **Figure 8.33** compares diabatic and adiabatic contributions to the temperature budget in the TTL for the climatological zonal mean in order to investigate if they balance each other out as expected. Here, the diabatic component represents the total heating rate diagnosed using the zonal-mean thermodynamic equation as described by Martineau et al. (2018). The term represents the sum of the time rate of change, meridional advection, vertical advection, meridional eddy, and vertical eddy terms. The adiabatic cooling represents the vertical advection tendency due to the residual mean vertical velocity (w^*) and is obtained as minus w^* times the vertical temperature gradient.

The overall structure and magnitudes of heating and cooling patterns between the deep tropics and the subtropics confirm the approximate balance mentioned above. The strongest adiabatic cooling / diabatic heating arises near the top of deep convection embedded in the ITCZ (between 5 - 10 °N and 300 - 250 hPa). There are also signatures of the double peak in tropical upwelling in the lowermost stratosphere with maxima in adiabatic cooling near 15 °N/S (*Ming et al.*, 2016). ERA-Interim and JRA-55 show similar double peak structures in the diabatic heating fields, which are weaker and shifted to slightly higher altitudes in MERRA-2 and CFSR. Moreover, MERRA-2 shows diabatic cooling at 200 hPa. This feature is less pronounced than in diabatic heating rates based on parameterized physics (**Fig. 8.20**), but it is neither balanced by adiabatic warming in this region nor present in the other reanalyses. This indicates that other contributions to the heat budget are important in this region in MERRA-2. Data assimilation plays a key role in the difference between the physical temperature tendencies shown in **Figure 8.20** and the diagnosed heating rates shown in **Figure 8.33**. However, for MERRA-2 in the tropics, the influence of data assimilation is on average negative below 200 hPa and positive above 200 hPa (**Fig. A8.6**, *Appendix A*), leaving temperature tendencies at the 200 hPa level largely unaffected.

Overall, both adiabatic and diabatic tendencies are more consistent in the lower stratosphere than in the upper troposphere. **Figure 8.34** further shows that this consistency is not present among older reanalyses (*cf.*, the dashed lines showing a large range of adiabatic cooling values in the lower stratosphere). Amongst the more recent products, MERRA-2 consistently shows the smallest tendencies. JRA-55's AMIP version roughly agrees with JRA-55, although it shows somewhat smaller values throughout the profile. ERA-20C is also consistently biased towards smaller values compared to ERA-Interim.

8.5.4 Key findings and recommendations

Key findings

- Lagrangian transport studies demonstrate large differences in reanalysis temperatures at the dehydration point, however, the data sets agree on the spatial distribution of dehydration locations. Given warm biases at the Eulerian cold point tropopause, Lagrangian dehydration points can be expected to be up to 1 K too warm. (*Section 8.5.1*)

- Diabatic vertical ascent appears to be faster in ERA-Interim, which produces a TTL residence time (between 370 K and 400 K) of ~2 months, in contrast to residence times of ~3 months or longer based on MERRA, MERRA-2, or CFSR. Despite the large differences in absolute values, all reanalysis data sets produce roughly similar distributions, seasonal cycles, and interannual variations of TTL residence time. (*Section 8.5.2*)

Key recommendations

- Lagrangian studies above 370 K (120 hPa), based on diabatic trajectories show more realistic tropical ascent rates when based on MERRA-2 or CFSR. Below 370 K (120 hPa), however, diabatic heating rates in these two data sets imply time-mean descent and therefore require careful treatment of convective detrainment source terms. (*Section 8.5*; see also *8.4* and *8.8*)

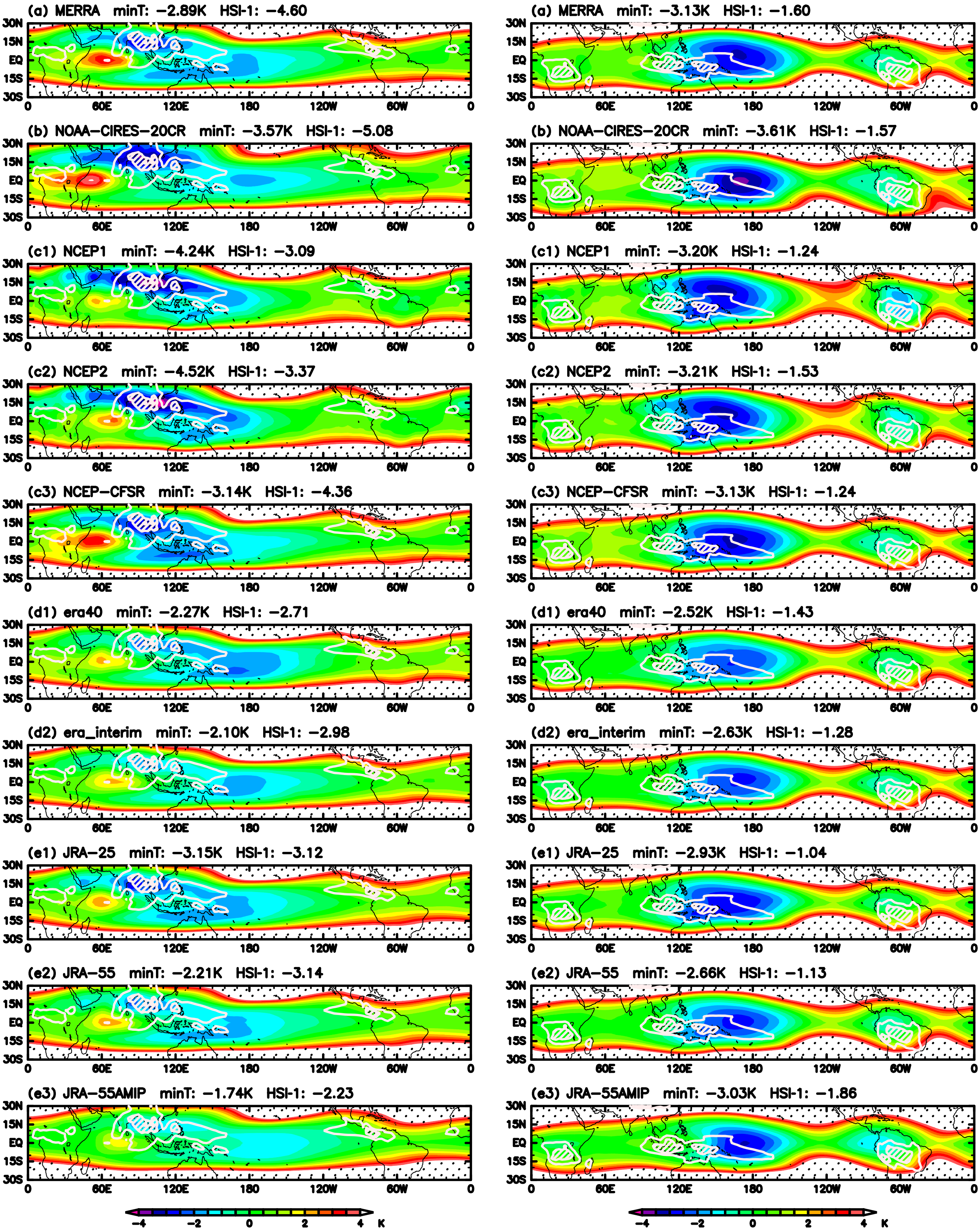


Figure 8.35: *Longitude-latitude sections of the climatological temperature anomaly at 100 hPa during (left) June - August and (right) December - February 1979 - 2005, derived from (top to bottom) MERRA, 20CR v2, NCEP-NCAR R1, NCEP-DOE R2, CFSR, ERA-40, ERA-Interim, JRA-25, JRA-55, and JRA-55AMIP reanalysis datasets. The anomaly is calculated from the tropical mean value (10 ° S - 10 ° N) in each season. Values of the minimum temperature anomaly and the minimum HSI-1 are shown in the legend of each panel. Observed climatological OLR is also shown as white contours (for 180, 200, and 220 W m*$^{-2}$*).*

8.6 Wave activity

Tropical convective activity has unique horizontal patterns in different seasons and sub-seasonal variability, resulting in variabilities in tropical tropopause temperature at these spatio-temporal scales through equatorial wave dynamics. These variabilities strongly influence transport and dehydration in the TTL. In this section, we discuss tropical 100 hPa wave activity at seasonal and sub-seasonal time scales in temperature and winds in multiple reanalyses. (See *Chapter 9* for equatorial wave activities at higher altitudes.)

8.6.1 Horseshoe-shaped structure at the 100 hPa temperature

Low temperatures at 100 hPa generally occur over the equator in the eastern hemisphere and extend northwestward and southwestward to form a horseshoe-shaped structure (*e.g., Highwood and Hoskins*, 1998). This structure resembles a theoretical stationary wave response known as the Matsuno-Gill pattern (*Gill* 1980; *Matsuno,* 1966), which is a superposition of the Rossby and Kelvin wave responses to tropical convective heating. The magnitude of the 100 hPa temperature anomalies is different among reanalyses, although the climatological anomaly patterns have common features, including the horseshoe-shaped structure (*e.g., Fujiwara et al.*, 2012).

Figure 8.35 shows the horizontal distributions of the temperature anomalies at 100 hPa in JJA and in DJF from 10 reanalysis datasets. The values for each season are climatological averages over 27 years (1979 - 2005). Active convective regions based on NOAA OLR data are also shown. In JJA, off-equatorial strong heating in the Asian monsoon region in combination with equatorial heating around the maritime continent results in a horseshoe-shaped structure in the 100 hPa temperature, which is equatorially asymmetric. In DJF, equatorial heating around the maritime continent and western Pacific results in a dominant Kelvin wave response. Note that in individual months and years the Rossby wave response can be observed as well causing the horseshoe-shaped signal during these time periods (*e.g.*, **Fig. 1** of *Nishimoto and Shiotani*, 2012). Negative temperature anomalies show larger magnitude in MERRA, 20CR v2, NCEP-NCAR R1, NCEP-DOE R2, and CFSR than in ERA-40, ERA-Interim, JRA-25, JRA-55, and JRA-55AMIP, in both seasons. In addition, positive temperature anomalies located around 60 °E in the northern summer, which are surrounded by the negative anomalies, have larger amplitudes in MERRA, 20CR v2, and CFSR.

In order to investigate the longitudinal and seasonal variations of the horseshoe shaped temperature structure in a more quantitative way, *Nishimoto and Shiotani* (2012) defined the index (HSI-1) from two preliminary indices, which represent the Rossby and Kelvin wave responses.

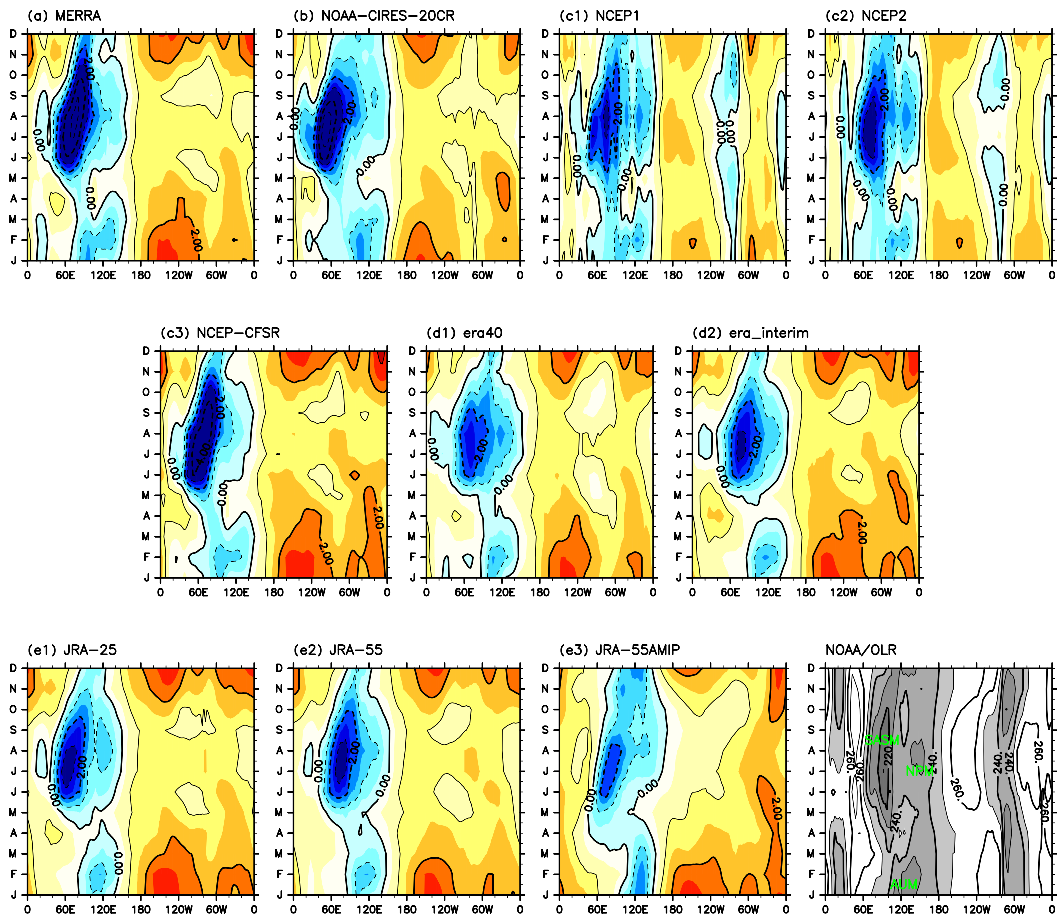


Figure 8.36: *Longitude-time sections of climatological HSI-1 derived from MERRA, 20CR v2, NCEP-NCAR R1, NCEP-DOE R2, CFSR, ERA-40, ERA-Interim, JRA-25, JRA-55, and JRA-55AMIP reanalysis datasets and climatological NOAA/OLR averaged over 15 °S - 20 °N.*

As representation of the Rossby response, the index HSI-R is defined as the meridional curvature of the 100 hPa temperature at the equator as a function of longitude x and time t:

$$\text{HSI-R}(x,t) = \frac{T_N(x,t) + T_S(x,t)}{2} - T_{Eq}(x,t) \quad (8.4),$$

where T_{Eq} denotes the temperature at the equator, and T_N and T_S are the temperatures averaged over 10 °N - 15 °N and 10 °S - 15 °S, respectively. If low temperatures occur in the 10° - 15° latitude bins as the Rossby response, the HSI-R index becomes **negative.As** a representation of the Kelvin response, the index HSI-K is defined as the zonal gradient of the 100 hPa temperature along the equator:

$$\text{HSI-K}(x,t) = T_{Eq}(x + \Delta x/2, t) - T_{Eq}(x - \Delta x/2, t) \quad (8.5),$$

where a differentiation length Δx is set at 20° longitude. When the temperature structure represents the Kelvin response, the HSI-K index becomes negative.

As the HSI-R and HSI-K values change accordingly with a positive correlation in response to heating generated by convection, the index HSI-1 is defined by using the first component of the empirical orthogonal function analysis of HSI-R and HSI-K values:

$$\text{HSI-1}(x,t) = a \times \text{HSI-K}(x + \alpha, t) + b \times \text{HSI-R}(x,t) \quad (8.6).$$

In the horseshoe-shaped structure, negative values of HSI-K are located slightly to the east of the negative values of HSI-R in agreement with the Matsuno-Gill pattern, so that we set the longitudinal phase lag of HSI-K relative to HSI-R at $\alpha = +15$°. The coefficients $a = 0.504$ and $b = 0.864$ are derived from the monthly mean composite data of MERRA, ERA-Interim, JRA-55, and CFSR. In this section, we apply the HSI-1 index to 9 reanalyses and compare the results quantitatively.

Longitude-time sections of climatological monthly HSI-1 values and climatological monthly NOAA/OLR (as a proxy for convective activity) values averaged over 15 °S - 20 °N are provided in **Figure 8.36**. The seasonal variation is almost identical among the reanalyses, and negative HSI-1 values exist in the eastern hemisphere with peaks in the northern and southern summer seasons. During NH summer, the negative HSI-1 values extend from 40 °E - 150 °E in every reanalysis and the peak is located between 60 °E - 100 °E in July or August. The amplitude of the negative HSI-1 values is very large in MERRA, 20CR v2, and CFSR corresponding to the large positive temperature anomalies located around 60 °E as shown in **Figure 8.35**. During SH summer, the negative values extend over 60 °E - 150 °E in MERRA, 20CR v2, NCEP-NCAR R1, NCEP-DOE R2, and CFSR, whereas those in ERA-40, ERA-Interim, JRA-25, JRA-55, and JRA-55AMIP extend narrower in longitude (80 °E - 150 °E).

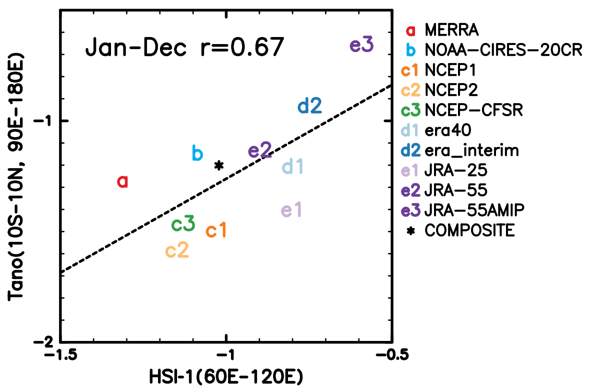


Figure 8.37: *Scatter plot of the climatological HSI-1 averaged over 60 °E - 120 °E vs. the climatological temperature anomaly averaged over 10 °S - 10 °N and 90 °E - 180 °E. The composite (*) is made from MERRA, ERA-Interim, JRA-55, and CFSR.*

As *Nishimoto and Shiotani* (2012) showed based on monthly mean ERA40 data, the seasonal cycle of negative HSI-1 values is significantly related to that observed in convective activities over three monsoon regions: the South Asian Summer Monsoon and the North Pacific monsoon areas during the northern summer and the Australian monsoon area during the southern summer. This relationship is expected theoretically because the Matsuno-Gill pattern is the response from tropical heating including off-equatorial heating within the tropics (*Gill*, 1980). The correlation coefficient between the climatological monthly HSI-1 values averaged over 40° E - 150° E and the OLR values averaged over 60° E - 180° E is larger than 0.8 for every reanalysis dataset and statistically significant.

Figure 8.37 shows a scatter plot of the climatological HSI-1 value averaged over 60° E - 120° E and the climatological temperature anomaly averaged over 10° S - 10° N and 90° E - 180° E among various datasets. Among the reanalysis datasets, a positive relationship is found between the climatological HSI-1 value and the climatological temperature anomaly. The HSI-1 value ranges from -1.3 to -0.6 while the temperature anomaly ranges from -1.6 K to -0.7 K. The datasets can be divided into two groups depending on whether the HSI-1 value is smaller or larger than - 1.0. The former group with the smaller HSI-1 values includes MERRA (a), 20CR v2 (b), and the NCEP series (c1-c3) of reanalyses, while the latter group includes the ECMWF (d1-d2) and JRA series (e1-e2) of reanalyses. These results suggest that the strength of the horseshoe-shaped structure, which controls the magnitude of cold temperature anomaly, is dependent on the inherent dynamical model or assimilation system used in reanalysis.

8.6.2 Equatorial waves

Significant sub-seasonal variability is found in temperature, horizontal winds, and other parameters in the TTL (*Fueglistaler et al.*, 2009a). This is due to various types of equatorial waves, intraseasonal oscillations/the Madden-Julian Oscillation (MJO) (*Madden and Julian*, 1994), and other disturbances that are primarily generated by tropical organized convection (*e.g.*, *Kiladis et al.*, 2009). Previous case studies have investigated the roles of equatorial Kelvin waves in the TTL for large temperature changes, ozone transport, dehydration, turbulence generation, and cirrus variations (see, *e.g.*, *Fujiwara et al.*, 2012 and the references therein).

In this Section, we discuss the wave activity at the 100 hPa level using various reanalysis data sets. The data used are sub-daily (3-hourly for MERRA and MERRA-2; 6-hourly for the other reanalyses). Characteristics of equatorial waves for different stratospheric levels can also be found in *Section 9.3* and *Kim et al.* (2019) using the same method as described here.

The method of obtaining the wave activity is based on the zonal wavenumber–frequency (k - ω) spectral analysis with equatorially symmetric and antisymmetric decomposition with a background spectrum estimation. The power spectral densities (PSDs) of the symmetric and antisymmetric components of variables (*e.g.*, for temperature, $T_S = [T(\lambda,\varphi,t) + T(\lambda,-\varphi,t)]/2$ and $T_a = [T(\lambda,\varphi,t) - T(\lambda,-\varphi,t)]/2$, respectively) are calculated as a function of k and ω for each month, after applying a 90-day window centered on the month (see *Kim et al.*, 2019, for further details). The PSDs are then averaged over 15°N - 15°S. The background spectra of the symmetric and antisymmetric components each are obtained following *Fujiwara et al.* (2012), by iterating 1-2-1 running average 23 times along the zonal wavenumbers and 7 times along the frequencies. The PSDs are presented in the variance-preserving form with log-scale axes, *i.e.*, $PSD(k,\omega) = |F(k,\omega)|^2 k\omega(\ln 10)^2/(\Delta k \Delta\omega)$, where F is the Fourier coefficient of the symmetric or antisymmetric component windowed, and the spectral intervals Δk and $\Delta\omega$ are 1 and 1/90 cyc day^{-1}, respectively.

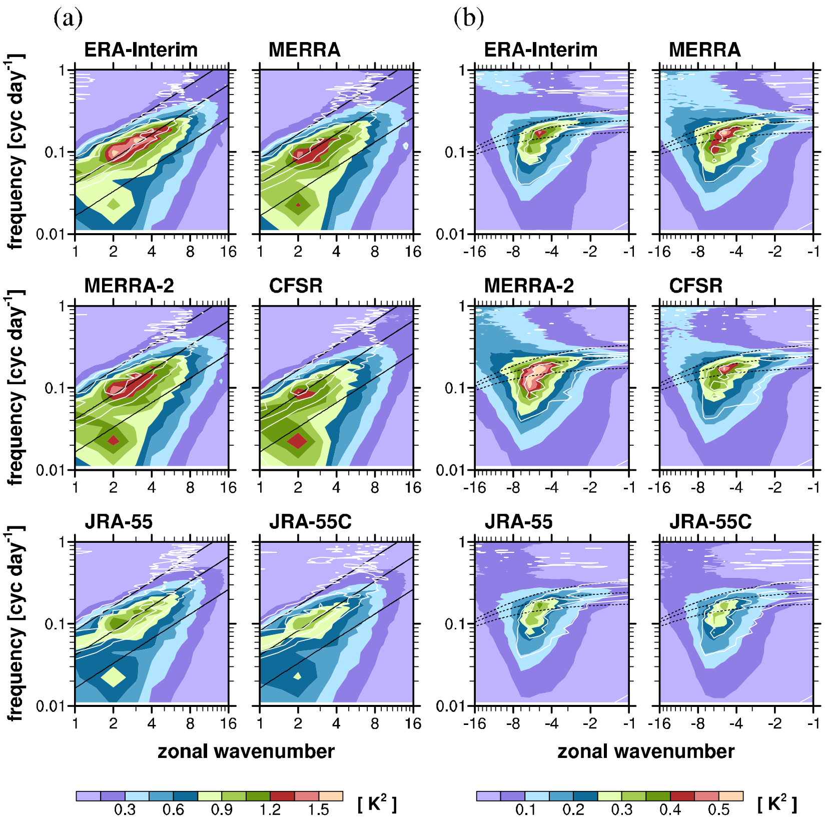


Figure 8.38: *Zonal wavenumber–frequency power spectra of 100 hPa temperature for the (a) symmetric and (b) antisymmetric components averaged over 15°N - 15°S for 1981 - 2010 obtained using ERA-Interim, MERRA, MERRA-2, CFSR, JRA-55, and JRA-55C, along with the dispersion curves of Kelvin waves for the vertical wavelengths of 2.5, 5, and 10 km assuming windless background states (black solid) in (a) and those of mixed Rossby-gravity waves for 2, 4, and 8 km (black dotted) in (b). The ratios to the symmetric and antisymmetric background spectra are also indicated in (a) and (b), respectively, for the values of 1.1, 1.5, and 2 (white solid). The symmetric and antisymmetric spectra are shown only for positive and negative zonal wavenumbers, respectively. (Modified after Kim et al., 2019; see also **Fig. 9.30**).*

Figure 8.38 shows the PSDs of 100 hPa temperature averaged over 1981 - 2010, obtained using ERA-Interim, MERRA, MERRA-2, CFSR, JRA-55, and JRA-55C. The symmetric and antisymmetric spectra are shown for $k>0$ and $k<0$, respectively. In the symmetric PSDs, the spectra from the six reanalyses have a similar shape: the primary peak is at $k=2$ and $\omega=0.09$ - 0.1 cyc day^{-1}, and a large portion of the PSD around this peak appears between the dispersion curves of Kelvin waves for the vertical wavelengths (L_z) of 2.5 km and 10 km. These curves also correspond to the zonal phase speeds of about 9.5 m s^{-1} and 38.2 m s^{-1}, respectively. The peak PSD value is largest in ERA-I (1.64 K^2) among the reanalyses, while the PSD at low phase speeds (around the dispersion curve of L_z=2.5 km) is larger in MERRA-2 than in the others. A secondary peak appears at $k=2$ and $\omega \sim 0.02$ cyc day^{-1} (~50-day period), which is associated with the MJO.

In the antisymmetric PSDs (**Fig. 8.38b**), a large portion of the PSD appears around the dispersion curves of mixed Rossby-gravity (MRG) waves for L_z of 2 - 8 km, with a peak at $k=5$ and $\omega=0.17$ - 0.18 cyc day^{-1}. The PSD values in MERRA-2 are largest in most spectral regions among the reanalyses. A significant portion of the PSD is also distributed in low-frequency ranges ($\omega<0.1$ cyc day^{-1}). This can be due to the westerly background wind in the western hemisphere where the dispersion curves might shift to low-frequency ranges. Consistent to this, the low-frequency wave activity is concentrated on the western hemisphere (**Fig. S2** of *Kim et al.*, 2019). It can also be attributed to the co-existence of free Rossby modes (*e.g.*, *Fujiwara et al.*, 2012; *Madden*, 2007).

Figure 8.39 shows the wave activity (see *Fujiwara et al.* (2012), for the definition of the wave activity) obtained using 100 hPa datasets from the six reanalyses (left six columns). In addition, the wave activity calculated at the native model levels and interpolated to 100 hPa is also shown (rightmost four columns). In comparison of the model-level results, it is found that the ERA-Interim presents the largest Kelvin and MRG wave activity for temperature and vertical-wind components.

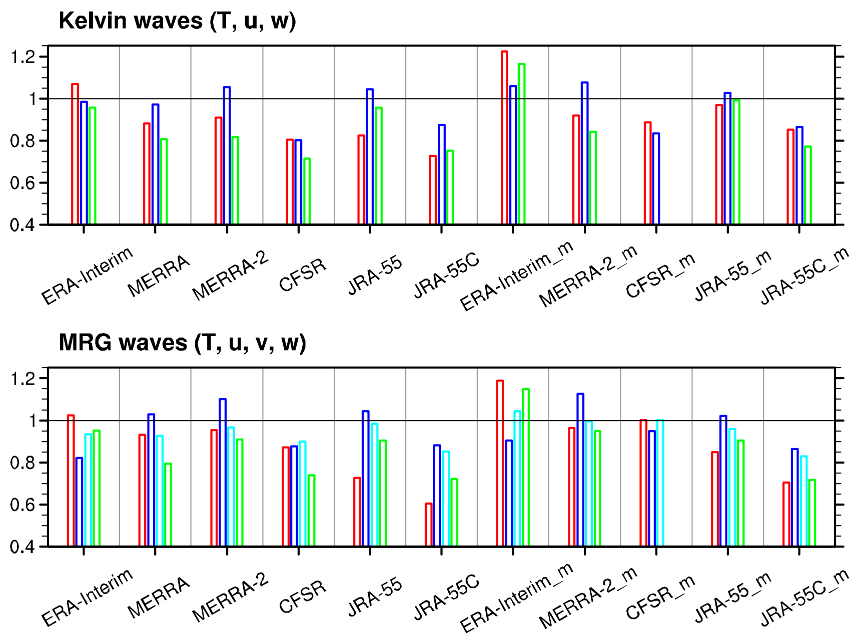


Figure 8.39: *The wave activity normalized by the ensemble average for the model-level results of ERA-Interim, MERRA-2, CFSR, and JRA-55: (upper) Kelvin waves and (lower) mixed Rossby-gravity (MRG) waves, obtained using temperature (red), zonal wind (blue), meridional wind (sky blue), and vertical wind (green). The first six columns present the results using standard pressure-level data, and the remaining using model-level data. See the text for the definition of the activity.*

On the other hand, the zonal-wind component of the wave activity is largest in MERRA-2. JRA-55 presents a moderate amount of the wave activity. The wave activity in JRA-55C is smaller than that in JRA-55, as expected, because of absence of the satellite data assimilation. The difference between the two is roughly 15 - 20 %. The wave activity calculated using the 100 hPa datasets is always smaller than that using the model-level datasets, by approximately 20 - 30 % for ERA-Interim, JRA-55, and JRA-55C. The underestimation is much less for MERRA-2 (and also MERRA, not shown) because they have a model level that is very close to 100 hPa (see *Appendix of Chapter 2*).

Fujiwara et al. (2012) made a similar spectral analysis using seven reanalysis data sets (NCEP-1, NCEP-2, ERA40, ERA-Interim, JRA-25, MERRA, and CFSR) for temperature and horizontal winds at 100 hPa for the period of 1990 - 2000. Their results for ERA-Interim, MERRA, and CFSR are mostly consistent with the ones shown in **Figures 8.38** and **8.39**. The older-generation reanalyses that are not included in **Figures 8.38** and **8.39** show generally lower wave activity. The increase in the wave activity in recent reanalyses could result from many factors including the increase in assimilated data sets and advance of assimilation schemes as well as model vertical resolutions of the reanalyses. Recent numerical modeling studies have reported that representation of the equatorial Kelvin and MRG waves is highly sensitive to the vertical resolution of models (*e.g.*, *Richter et al.*, 2014).

It was reported that the difference in the Kelvin wave variance at 100 hPa between JRA-55 and JRA-55C is persistently larger after the late 1990s than before (*Kim et al.*, 2019, **Fig. 7**), indicating a change in the contribution of satellite data to the reanalysis representation of Kelvin waves. Given the timing, it is attributed to the TOVS–ATOVS transition from 1998 (see *Chapter 2*, *Section 2.4*). While the introduction of new observational instruments generally improves the quality of reanalyses, it may require users of reanalyses to be cautious when they utilize the data for study on long-term variations. For example, the temperature variance of 100 hPa Kelvin waves exhibits a long-term trend from the mid-1990s to 2010 commonly in ERA-Interim, MERRA, MERRA-2, CFSR, and JRA-55, but such a trend is not clear in JRA-55C (*Kim et al.*, 2019). The different result between JRA-55C and JRA-55 manifests an artifact in the trend estimate via the transition of the satellite instruments.

8.6.3 Key findings

- Temperature anomaly patterns at 100 hPa have common features in all reanalyses, including characteristic horseshoe-shaped structures that resemble the stationary wave response to tropical heating. The strength of this structure differs among the reanalyses depending on the aspects of the dynamical model and/or assimilation system. Seasonal variations in the horseshoe-shaped temperature structure are almost identical among reanalyses, with a well-established horseshoe-shaped pattern during northern summer. (*Section 8.6.1*)

- The spectral shapes of low-frequency equatorial waves at 100 hPa are similar among the reanalyses, but their spectral magnitudes differ. Equatorial wave activity tends to be larger in ERA-Interim than in other reanalyses for most variables analyzed. JRA-55C exhibits significantly weaker wave activity than JRA-55 for both Kelvin and mixed Rossby-gravity waves, emphasizing the impact of assimilating satellite observations. (*Section 8.6.2*)

8.7 Width of the TTL

This Section focuses on the changes of the width of the TTL, whereas *Chapter 7* includes basic evaluations of the subtropical jets and tropopause breaks. Multidiagnostic intercomparisons for changes of the tropical belt have been carried out before based on various reanalyses (see *Davis and Rosenlof*, 2012).

The tropical belt has generally been defined as the region straddling the equator that encompasses both the upwelling and subsiding branches of the Hadley cells (*e.g.*, *Birner et al.* 2014). A number of studies have identified evidence that the latitudinal extent of the tropical belt and other features of Earth's atmospheric circulation have expanded poleward over the last several decades (*e.g.*, *Lucas et al.,* 2014 and *Seidel et al.,* 2008). This phenomenon has often been referred to as tropical widening.

The Hadley cell and subtropical jet both provide the physical underpinning for defining a tropical edge latitude. In practice, however, there have been numerous definitions of the tropical edges based on atmospheric phenomena that are assumed to be tied to the Hadley cell or subtropical jet, such as the subtropical break in the height of the tropopause (*e.g.*, *Davis and Rosenlof*, 2012). These disparate definitions of the tropical belt edges have been at least part of the reason that tropical widening estimates span such a large range of values, from statistically insignificant and near zero, to several degrees latitude per decade of highly statistically significant poleward movement.

Using climate model simulations and reanalyses, several recent studies have documented the degree to which tropical edge metrics are temporally correlated with one another (*Waugh et al.*, 2018; *Davis and Birner*, 2017; *Solomon et al.*, 2016). These studies have identified a subset of metrics that are directly correlated with the Hadley cell extent, and another subset of metrics uncorrelated or only weakly correlated with the Hadley cell extent. The latter subset, which includes the subtropical jet (STJ) and tropopause break (TPB) metrics (**Fig. 8.40**), is analyzed in this chapter since these metrics are a more direct measure of the edges of the TTL.

For both types of metrics, we analyze methodologies based on instantaneous longitudinally resolved and zonal-mean annual-mean fields. The instantaneous methodology for the STJ comes from *Manney and Hegglin* (2018) (*Section 8.7.1*), while the zonal-mean STJ methodology is from *Davis and Birner* (2017) (*Section 8.7.3*). For the tropopause break, we use the instantaneous analysis from *Martin et al.* (2019) (*Section 8.7.2*), and the zonal-mean methodology from (*Adam et al.*, 2018) (*Section 8.7.3*).

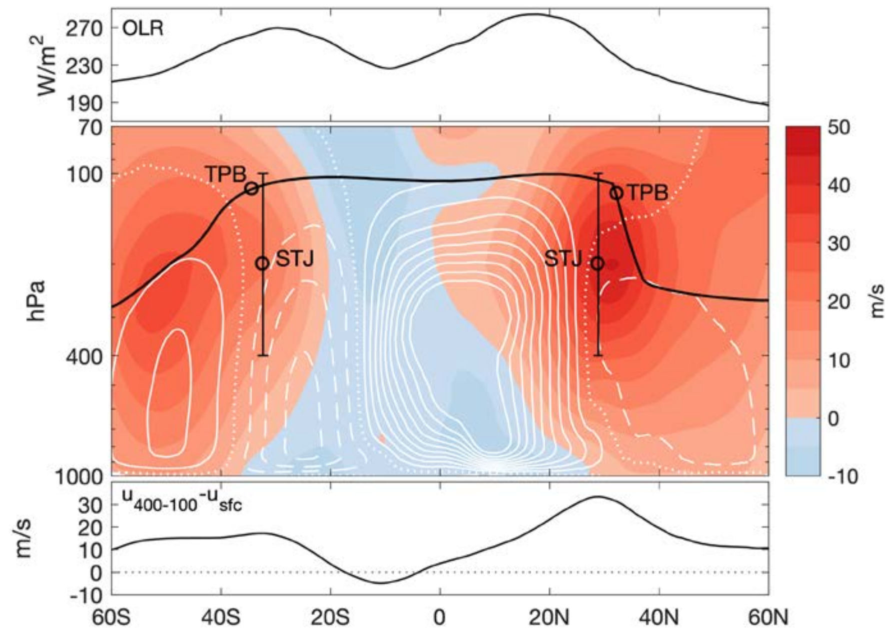


Figure 8.40: *The zonal-mean general circulation for February 2000 from the MERRA-2 reanalysis: (top panel) outgoing longwave radiation; (middle panel) zonal-mean zonal wind (shading, every 5 m s^{-1}) with tropopause pressure (black contour), mean meridional streamfunction (white contours, negative values dashed, every 20x10^9 kg s^{-1}, zero-contour dotted), tropopause break (TPB) and subtropical jet (STJ) metrics; and (bottom panel) the STJ metric field given by the 400 - 100 hPa average zonal-mean zonal wind with the surface zonal-mean zonal wind subtracted.*

8.7.1 Zonally-resolved subtropical jet diagnostic

The locations and characteristics of the upper tropospheric jet and tropopause are determined using the JETPAC (JEt and Tropopause Products for Analysis and Characterization) package, as described by *Manney et al.* (2011, 2014, 2017) and *Manney and Hegglin* (2018). The subtropical jet latitude, altitude, and frequency, among other diagnostics, are presented in *Chapter 7* "ExUTLS". An upper tropospheric jet is identified wherever there is a wind speed maximum greater than 40 m s^{-1}. The boundaries of the jet region are the points surrounding that maximum with wind speed below 30 m s^{-1}. When more than one maximum above 40 m/s appears within a given 30 m s^{-1} contour, they are defined as separate cores if the latitude distance between them is greater than 15 degrees or the decrease in wind speed between them is greater than 25 m s^{-1}. Since the human eye excels at this sort of pattern recognition, these parameters were tuned to approximate the choices made by visual inspection.

The lapse rate tropopause is defined using the WMO definition (*Section 8.2*; a review of issues related to definition of the thermal tropopause is given by *Homeyer et al.* 2010). As in *Manney and Hegglin* (2018), the subtropical jet is defined as the most equatorward westerly jet for which the WMO tropopause altitude at the equatorward edge of the jet is greater than 13.0 km and that tropopause altitude drops by at least 2.0 km from the equatorward to the poleward side of the jet. This definition identifies the jet across which the "tropopause break" occurs.

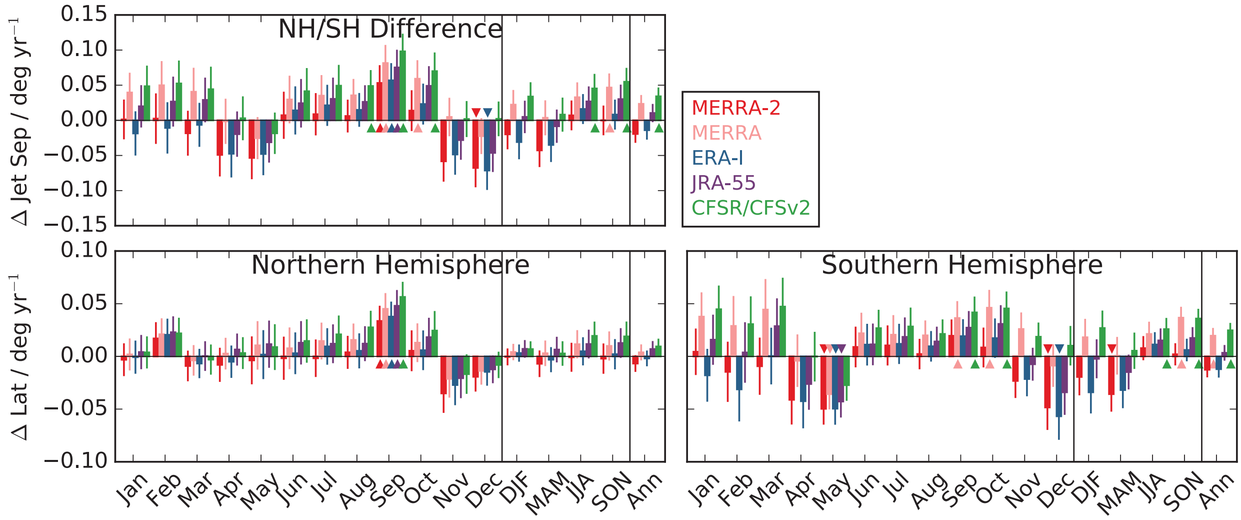


***Figure 8.41:** Bar charts of global subtropical jet latitude and NH/SH subtropical jet separation as a function of month, season, and annual, showing five reanalyses. The bars show the slopes of the linear fits and the error bars (centered about the top of the bars) the 1-sigma uncertainty in those slopes. Note that, in this and similar figures, absolute value of latitude is used, so positive slopes (bars extending upward from the zero line) indicate a poleward shift in both hemispheres. Triangles indicate cases where the permutation analysis shows the slope to be significant at the 95 % confidence level. Adapted from Manney and Hegglin (2018). ©American Meteorological Society. Used with permission.*

A jet intercomparison with respect to the tropical circulation is presented in *Manney et al.* (2017). The authors compared upper tropospheric jets (as well as multiple tropopauses and subvortex jets) in MERRA-2 with those in MERRA, ERA-Interim, JRA-55, and CFSR. Their results show (their **Figure 7**) stronger Walker circulation westerlies in DJF downstream of the Australian Monsoon, and stronger easterlies associated with that monsoon. Likewise, in JJA, the jets bounding the South Asian Summer Monsoon (SASM), particularly the tropical easterly jet, are stronger / more persistent in MERRA-2 and MERRA than in the other reanalyses studied. The Asian monsoon easterlies peak at a lower altitude in CFSR than in the other reanalyses studied (their **Figure 8**). Overall, *Manney et al.* (2017) emphasized that not only the vertical grid spacing, but also differences in the location of the model levels, are important in the reanalyses' representation of jets, including the jets associated with tropical circulations.

Jet latitudes and corresponding latitude shifts are examined based on JETPAC products for the time period 1980-2014 in *Manney and Hegglin* (2018). Their analysis is based on identification of individual jets, thus separating the subtropical and polar jets, and highlights the large regional and seasonal variation in trends in jet location. Therefore, the results sometimes reveal different trends in tropical width than have been shown using zonal or annual mean diagnostics or diagnostics that do not clearly separate the subtropical and polar jets (*Manney and Hegglin*, 2018). A brief summary of the results for subtropical jet latitudes, which are used as a measure of tropical width, is given below.

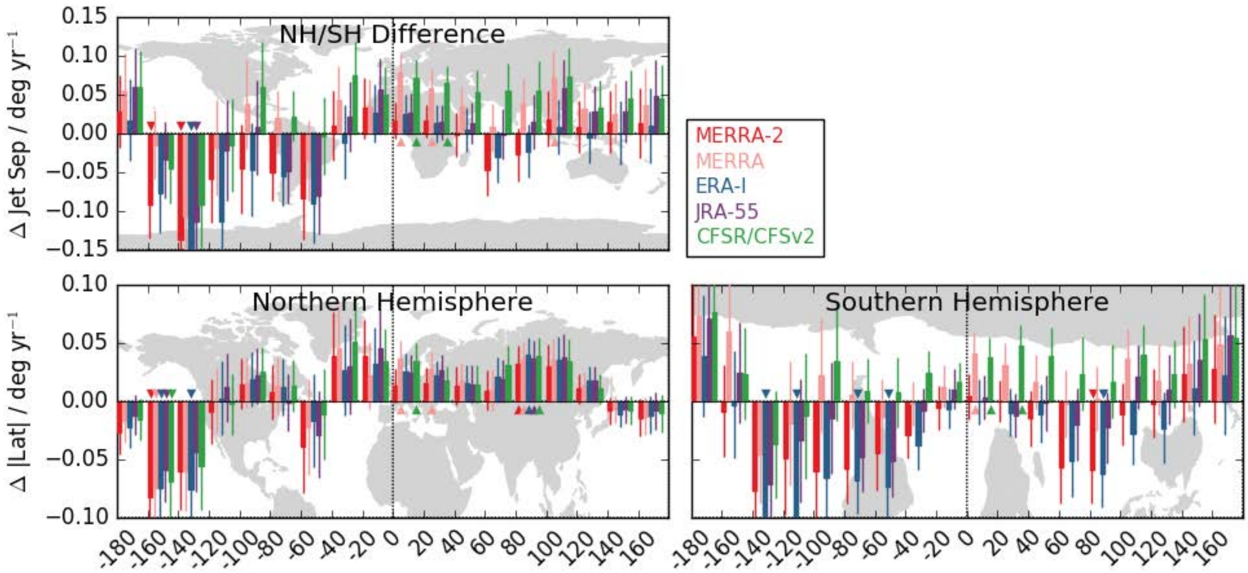


***Figure 8.42:** As in Fig. 8.41, but as a function of longitude during DJF. Adapted from Manney and Hegglin (2018). ©American Meteorological Society. Used with permission.*

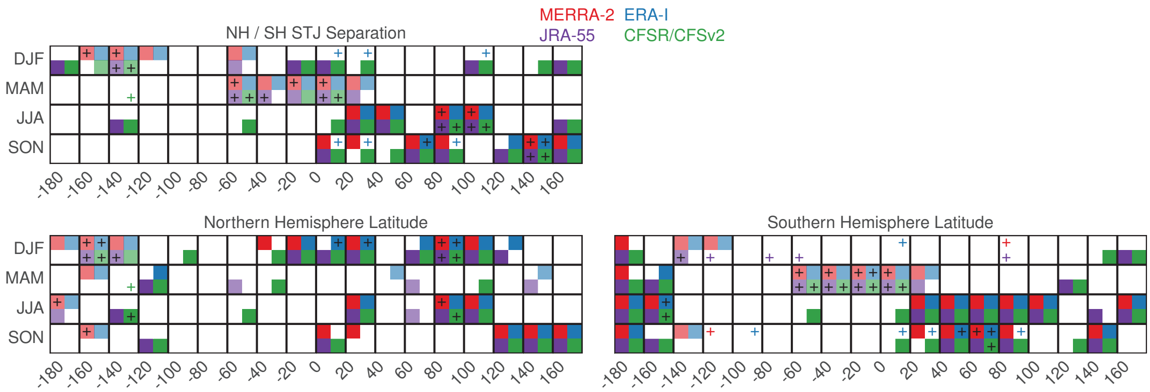


__Figure 8.43:__ Matrix plots summarizing changes in NH and SH subtropical jet latitude and in tropical width. Colored boxes are shown for MERRA-2 (red, upper left of each season / longitude region square), ERA-I (blue, upper right), JRA-55 (purple, lower left), and CFSR (green, lower right) if the sign of the trends agrees among all four reanalyses, and if the trend for that reanalysis is greater than the 1-σ uncertainty in the slope. Plus signs indicate cases where the permutation analysis shows the slope to be significant at the 95 % confidence level. Positive (negative) trends are indicated by bold (pale) colors. The NH (SH) is shown on the left (right). Adapted from Manney and Hegglin (2018). ©American Meteorological Society. Used with permission.

Subtropical jet latitude changes are shown in **Figure 8.41** including the slopes, and ± 1-σ uncertainties from each reanalysis for all months and seasons and annually. Annually, and during some seasons (DJF and MAM) and months (*e.g.*, January to March in both hemispheres) the reanalyses do not agree on the sign of the latitude change over 1980 through 2014. Robust (and sometimes significant) latitude increases are seen in June through October in both hemispheres, and robust NH latitude decreases in November and December. While the sign of these changes agrees among the reanalyses, the magnitude varies strongly.

Longitudinal variations of the jet latitude changes are shown in **Figure 8.42** with DJF from **Figure 8.41** broken down into 20° longitude regions. Robust positive shifts are seen in the NH over Europe and Asia, in the region where a strong, nearly zonal subtropical jet dominates the flow (*e.g.*, *Manney et al.*, 2014). Negative shifts are seen in the eastern Pacific in both hemispheres, and over South America and the western Atlantic in the SH, but the magnitude of the SH shifts varies greatly between the reanalyses, and CFSR shows positive shifts in part of this region. These changes thus result in inconsistent changes in tropical width in DJF, except for clear tropical narrowing over the eastern Pacific.

Regions and seasons that show robust changes in tropical width are summarized in **Figure 8.43**, where boxes are filled only if the trends of all four reanalyses agree in sign and the individual reanalysis's trend is greater than the 1-σ uncertainty. As discussed by *Manney and Hegglin* (2018), the most robust changes are where all reanalyses agree on the sign of the trend, the slope is greater in magnitude than the 1-σ uncertainty, and the permutation analysis shows significance at the 95 % confidence level. Such robust changes in tropical width are seen only in a few regions and seasons: robust tropical widening occurs in JJA over Africa and parts of Asia, and in SON over the western Pacific; robust tropical narrowing occurs in DJF over the eastern Pacific and in MAM over the Atlantic and western Africa. Because these jet-based diagnostics cannot be compared with observations, the agreement among the reanalyses was used as a key factor in assessing the robustness of trends.

8.7.2 Zonally-resolved tropopause break diagnostic

The so-called tropopause break (*i.e.*, the sharp discontinuity in tropopause altitude between the tropics and extratropics) is used as an instantaneous metric for the northern and southern edges of the tropics. To identify tropopause break latitudes, lapse-rate (WMO) tropopause altitudes are computed at each analysis time using model-level temperature and geopotential height fields from each reanalysis. Following tropopause calculation, frequency distributions of tropopause altitudes are computed for each hemisphere to enable identification of the frequency minimum between the high-altitude tropical mode and the low-altitude extratropical mode. The tropopause height corresponding to the frequency minimum in each hemisphere is then used as a threshold for global contouring, which provides the instantaneous latitude of the tropopause break as a function of longitude. Additional detail on this process can be found in *Martin et al.* (2020).

Tropopause break latitudes from 1981 - 2015 are used for trend analysis. In order to examine regional variations in the width of the tropics, the trends are computed every 1 degree in longitude. **Figure 8.44** presents 35-year mean tropopause break locations and their long-term trends from four modern reanalyses: ERA-Interim, JRA-55, CFSR, and MERRA-2. The thick portions of the trend lines are significant at the 3-sigma level. Tropopause break locations are largely consistent amongst the reanalyses, with the largest differences found over the ocean basins. Apart from CFSR, trends in the reanalysis show consistent longitudinal variability, with large and significant narrowing trends over the Pacific Ocean basin in each hemisphere. Weaker significant widening trends are found in

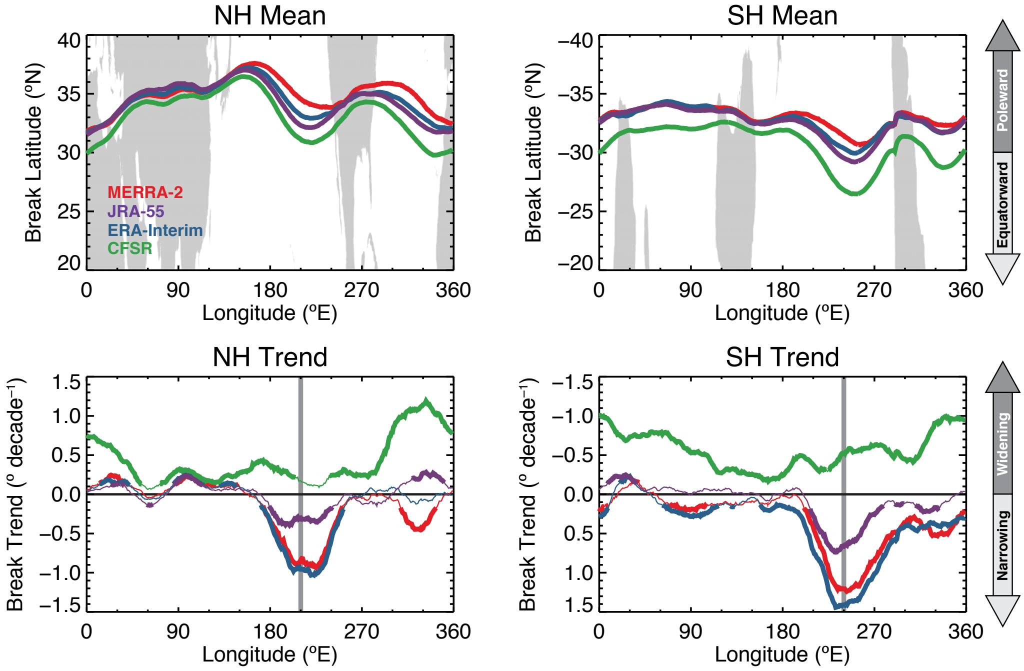


***Figure 8.44:** Mean tropopause break latitudes for 1981 - 2015 (top) and 35-year latitude trends (bottom) from four modern reanalyses (MERRA-2, JRA-55, ERA-Interim, and CFSR) as a function of longitude within the Northern Hemisphere (left) and Southern Hemisphere (right). Thin segments of the trend lines represent insignificant values, while thick segments represent trends that are significant at the 3-σ level. Adapted from Martin et al. (2020) "©American Meteorological Society. Used with permission.*

the NH, especially over northern Africa and eastern Asia. In the SH, trends point to weak narrowing outside of the Pacific, but with less consistency amongst the reanalyses. CFSR largely disagrees with the remaining three reanalyses, showing significant widening throughout much of both hemispheres.

8.7.3 Zonal mean subtropical jet and tropopause break diagnostics

The zonal-mean subtropical jet metric is defined as the latitude of maximum upper-tropospheric zonal-mean zonal wind with the zonal-mean surface zonal wind removed in each hemisphere (*Davis and Birner*, 2017). Here, the upper-tropospheric wind is defined as the 100 hPa to 400 hPa average zonal-mean zonal wind, while the surface wind is defined as the 1000 hPa zonal-mean zonal wind. In many cases, the subtropical and mid-latitude eddy-driven jets are difficult to distinguish in the raw zonal-mean zonal wind field. However, the two jets can be easily distinguished by considering their fundamentally different vertical structures. While the eddy-driven jet is highly barotropic, the subtropical jet is highly baroclinic. Removing the zonal-mean surface zonal wind from the upper-tropospheric zonal-mean zonal wind therefore results in an unambiguous zonal wind maximum in the subtropics characteristic of the subtropical jet. The zonal-mean tropopause break metric is defined as the latitude of the maximum meridional gradient in zonal-mean tropopause height in each hemisphere (*Adam et al.*, 2018), analogous to the zonally-resolved tropopause break metric. For the zonal-mean metric, the tropopause is calculated by applying the standard WMO definition of the tropopause to zonal-mean temperature and geopotential height.

Examination of the times series of the TTL edge latitudes from 1980-2010 (**Fig. 8.45**) reveals a clear differentiation between the zonal-mean subtropical jet and tropopause break metrics. The jet latitudes are generally equatorward of the tropopause break latitudes, especially in the SH. There is overall better agreement among the reanalyses on the jet latitudes than on the tropopause break latitudes. Interestingly, the spread of NH

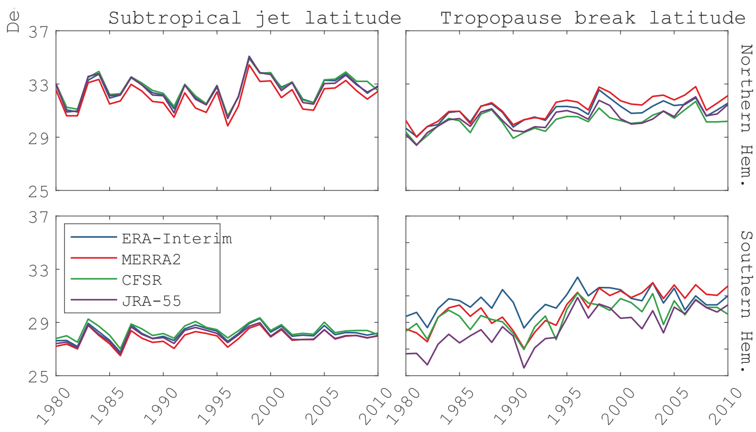


***Figure 8.45:** Time series of the TTL edge latitudes based on the zonal-mean (left column) subtropical jet latitudes and (right column) tropopause break latitudes in the (top row) Northern and (bottom row) Southern Hemispheres.*

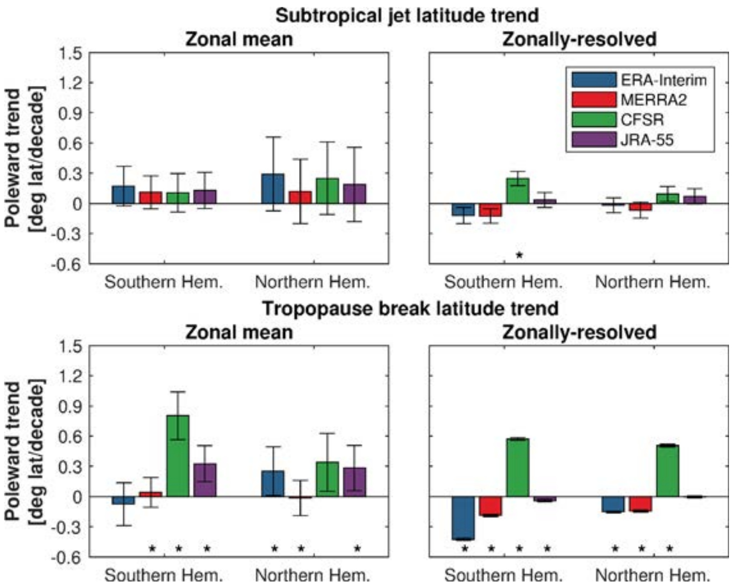


Figure 8.46: Trends in the TTL edge latitudes using the (top row) subtropical jet latitudes and (bottom row) tropopause break latitudes. Trends in the zonal-mean metric are shown in the left column, while the zonal-mean of the zonally-resolved trends are shown in the right column. Whiskers indicate 95 % confidence intervals. Stars indicate trends statistically significant at the 95 % confidence level, except for the zonally-resolved subtropical jet latitude trends which are statistically significant based on the methodology in Manney and Hegglin (2018).

tropopause break latitudes increases into the 2000's.

As might be expected given the good agreement of the zonal-mean subtropical jet latitudes among the reanalyses, their trends are consistent in both hemispheres (**Fig. 8.46**). The trends range from 0.1 ° to 0.3 ° poleward/decade in both hemispheres, but no reanalysis exhibits a statistically significant trend at the 95 % level. On the other hand, there is a relatively large degree of variation among the zonal-mean tropopause break latitude trends (**Fig. 8.46**). MERRA-2 and JRA-55 show significant expansion in both hemispheres, with poleward expansion of approximately 0.8 deg/decade in the SH. The trends in CFSR and ERA-Interim are lower by comparison.

There are some noteworthy differences between these zonal-mean metric trends and the zonal-mean of the zonally-resolved metric trends. The zonally-resolved subtropical jet trends are generally weaker, but like the zonal-mean jet trends are also not significant. On the other hand, the zonally-resolved tropopause break trends tend to have the opposite sign as their zonal-mean counterparts in all reanalyses except CFSR (statistically significant equatorward instead of poleward shifts). CFSR is the only reanalysis to exhibit consistent trends between its zonal-mean and zonally-resolved tropopause break trends.

The differences in the trends in the TTL edge latitudes as measured by the zonally-resolved and zonal-mean metrics warrant further investigation. One reason why trends in the subtropical jet latitudes may be more consistent could be that the zonal wind field is smoothly-varying. Therefore, the zonal-mean of the jet latitudes and their trends should be expected to be representative of the latitude of the zonal-mean jet and its trend. The tropopause break and the tropopause itself are discontinuities, which may be one reason why the trends in the zonal-mean and zonally-resolved metrics disagree.

8.7.4 Key findings and recommendations

Key findings

- Metrics of the width of the TTL based on the zonally-resolved subtropical jet and tropopause break show robust changes in only a few regions and seasons. Both metrics are in agreement on a narrowing of the TTL over the East Pacific and a widening of the TTL over Africa and parts of Asia. Trends in zonal-mean annual-mean values show little agreement among the reanalyses, with significant TTL widening found only for the CFSR tropopause break metric. (*Sections 8.7.1* and *8.7.2*)

- The zonal-mean subtropical jet and tropopause break diagnostics suggest stronger trends in the width of the TTL than their zonally-resolved counterparts. While the subtropical jet trends are not significant, the tropopause break trends show significant widening in both MERRA-2 and JRA-55. The zonal-mean and zonally-resolved subtropical jet diagnostics are more consistent than the tropopause break diagnostics, possibly related to smoother variations in the zonal wind field relative to the tropopause break. (*Section 8.7.3*)

Key recommendations

- Metrics of tropical width based on the subtropical jet or tropopause break are only weakly correlated with the measures of tropical width that are most closely related to changes in surface climate. Questions concerning which aspects of the climate system are measured by a given metric need to be assessed before these metrics are applied. (*Section 8.7*)

- When applying metrics of tropical width based on the subtropical jet or tropopause break, it is recommended to use multiple reanalyses and to be aware of the caveat that the zonal-mean diagnostics suggest stronger trends than their zonally-resolved counterparts. (*Section 8.7*)

8.8 South Asian summer monsoon

Each year, during boreal summer, a strong anticyclonic circulation system emerges in the UTLS over South Asia. This so-called South Asian Summer Monsoon (SASM) anticyclone is a large-scale circulation system (*Mason and*

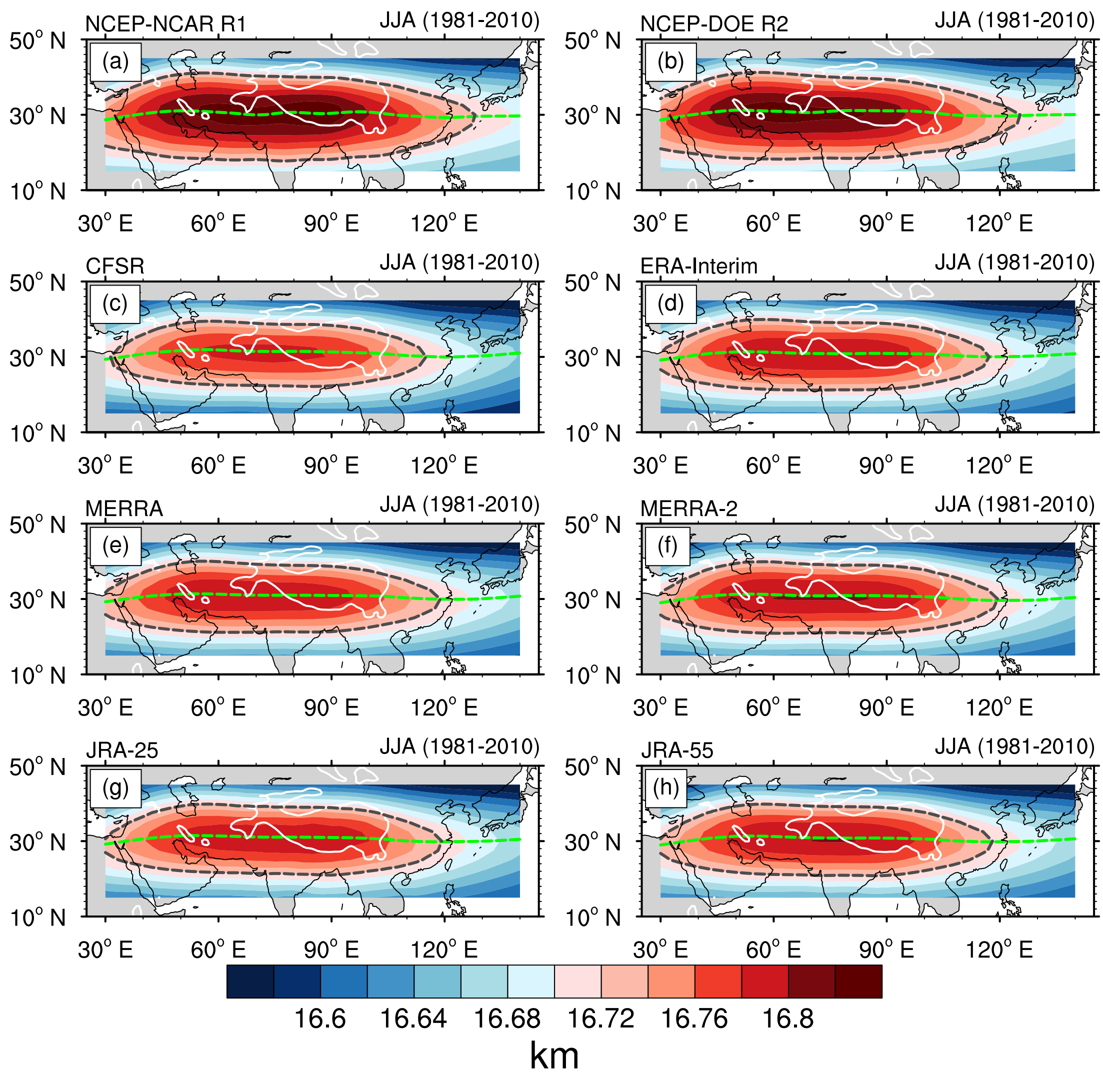


Figure 8.47: *Climatological mean geopotential height (km) at 100 hPa for eight reanalyses during JJA 1981-2010 based on 6-hourly data at 2.5°×2.5° resolution. The green dashed line in each panel indicates the climatological ridgeline based on the corresponding data set (see text for details). Grey dashed contours show the 16.72 km geopotential height contour. White contours show surface elevations greater than 2 km from ERA-Interim (in all panels), thus outlining the Tibetan Plateau (modified from* ***Fig. 4*** *in Nützel et al., 2016).*

Anderson, 1963) characterised by strong dynamic variability (*e.g., Hsu and Plumb*, 2000; *Popovic and Plumb*, 2001). Associated with the anticyclone, enhanced abundances of tropospheric trace gases (*e.g.,* CO, H_2O) are present in the UTLS over the SASM region (*e.g., Santee et al.*, 2017; *Randel and Park*, 2006; *Li et al.*, 2005). However, the contribution of the SASM anticyclone to stratospheric air masses – as discussed by *Dethof et al.* (1999) and *Randel et al.* (2010), among others – remains a current research topic (*e.g., Ploeger et al.*, 2017; *Garny and Randel*, 2016; *Pan et al.*, 2016). A recent study by *von Hobe et al.* (2021) concludes that the interplay of deep convection and subsequent radiatively driven ascent leads to effective transport of air masses from the Asian troposphere into the stratosphere. Given the importance of SASM anticyclone variability to the distribution of trace gases in the SASM UTLS (*e.g., Ploeger et al.*, 2015; *Garny and Randel*, 2013; *Yan et al.*, 2011), we discuss the climatological properties and variability of the SASM and its anticyclone as represented by atmospheric reanalysis systems in the following.

8.8.1 Anticyclone: climatology and variability

As a first analysis of the SASM anticyclone, we show the mean geopotential height at 100 hPa during June-July-August (JJA) 1981-2010 in **Figure 8.47**. The corresponding climatological ridgelines, which mark the position of the minimum absolute zonal wind speeds at each longitude in the SASM region (*cf., Zhang et al.*, 2002), are included as green dashed lines in each panel. All reanalyses indicate that the ridgeline is located at roughly 30°N in the SASM region. The absolute values of 100 hPa geopotential height in the SASM region are similar in ERA-Interim, MERRA-2-ASM, MERRA-ASM, JRA-55, and JRA-25. Relative to these reanalyses, mean 100 hPa geopotential heights are slightly lower in CFSR and slightly higher in NCEP-R1 and NCEP-R2 (by ~20-40 m). Accordingly, the extent of the

SASM anticyclone is largest in NCEP-R1 and NCEP-R2 when a fixed geopotential height threshold of 16.72 km (grey dashed contour) is used to determine its boundary. Nevertheless, despite small differences, all reanalyses generally agree on the climatological mean position of the SASM anticyclone core. These results are in agreement with the findings of *Nützel et al.* (2016).

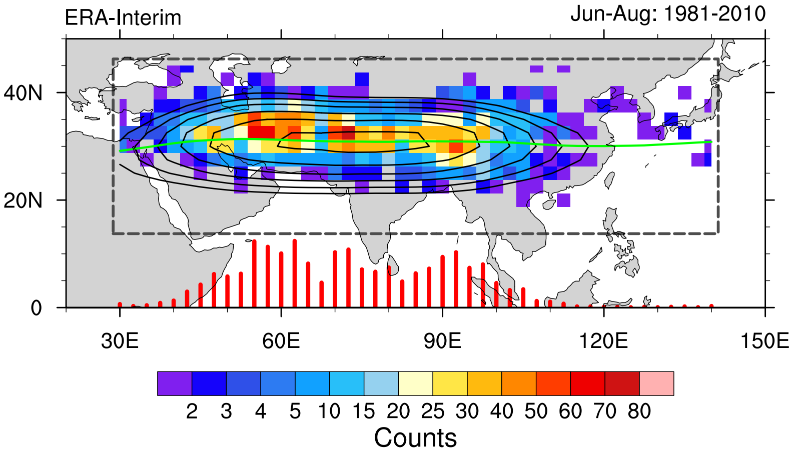


Figure 8.48: *Colour shading indicates the two-dimensional frequency of occurrence of the SASM anticyclone centre at 100 hPa from daily values based on ERA-Interim data for June to August 1981 - 2010 (2.5 x 2.5 bins; note the non-linear colour scale). The box marked by the grey dashed line indicates the range of the data that are used to diagnose the centre. Black contours show the long-term seasonal (JJA, 1981 - 2010) mean of the geopotential height (contour levels starting at 16.72 km and a spacing of 15 m) and the green line shows the long-term mean location of the ridgeline (zero zonal wind) at 100 hPa. Red bars indicate the one-dimensional PDF (bins of 2.5°) of the daily location of the ASM centre over the June - August period 1981 - 2010 with 2° corresponding to 1 % (analysis analogue to* ***Fig. 1b*** *by Nützel et al., 2016).*

Location of the SAS anticyclone centre

It has been suggested that the center of the SASM anticyclone exhibits positional bimodality, characterized by enhanced probabilities for the anticyclone to be centred either over the Iranian Plateau (IP) at 55–65° E or over the Tibetan Plateau at 82.5 - 92.5 ° E (*Zhang et al.*, 2002). The movement of the SASM anticyclone centre is of special interest as the chemical composition of the UTLS in the SASM region is linked to it (*e.g., Yan et al.* 2011). Here, we present frequency distributions of SASM anticyclone centre locations at 100 hPa as derived from eight reanalyses data sets. The analysis and corresponding interpretation are mainly based on results published by *Nützel et al.* (2016); additional details are provided therein.

Following *Zhang et al.* (2002), we identify the centre of the SASM anticyclone at 100 hPa by determining the maximum geopotential height along the ridge line (defined by the minimum absolute zonal wind at each longitude; see green lines in **Figs. 8.47** and **8.48**) within the SASM region (here defined as 15 - 45 °N, 30 - 140 °E). The two-dimensional distribution of the anticyclone centre based on daily data from ERA-Interim during JJA 1981 - 2010 is shown as colour shading in **Figure 8.48**. The red bars along the bottom depict the marginal probability distribution of the SASM anticyclone centre with respect to longitude (see caption for further details).

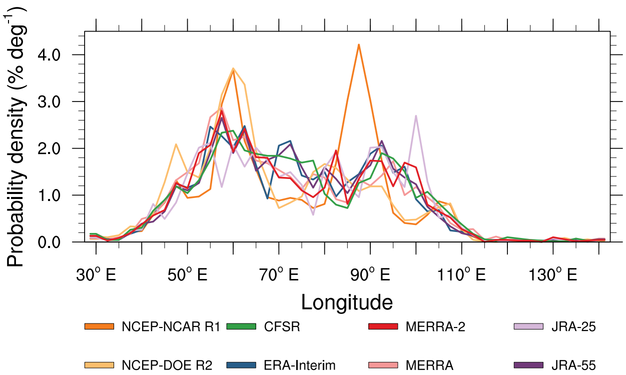


Figure 8.49: *Probability density function (% deg-1) of the SASM anticyclone centre location for daily data during JJA 1981 - 2010 at 100 hPa for eight reanalysis data sets at 2.5 °x 2.5 ° resolution. Daily data has been obtained by averaging 6 hourly data.*

Frequency distributions for the longitudinal location of the SASM anticyclone centre based on daily JJA data from eight reanalyses during the period 1981 - 2010 are shown in **Figure 8.49**. Clear bimodality (*i.e.*, a distinct double peak) is only present in NCEP-NCAR R1. Moreover, the updates introduced between NCEP-NCAR R1 and NCEP-DOE R2 (*Kanamitsu et al.*, 2002) lead to pronounced differences in the distribution of SASM anticyclone centre locations between these two reanalyses. However, NCEP-NCAR R1 and NCEP-DOE R2 agree in producing notable peaks over the IP. The greatest agreement with respect to SASM anticyclone centre distributions is found among CFSR, ERA-Interim, and JRA-55, three relatively

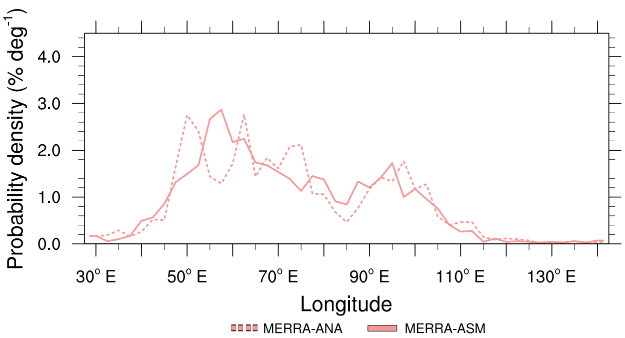


Figure 8.50: *Probability density functions (% deg-1) of SASM anticyclone centre locations on the 100 hPa isobaric surface based on daily data from MERRA-ANA and MERRA(-ASM) during JJA 1981 - 2010 at 2.5 ° × 2.5 ° resolution. Daily data were obtained by averaging 6-hourly data. MERRA-ASM data, which are available at 3-hourly resolution, were subsampled to the 6-hourly resolution of the MERRA-ANA data.*

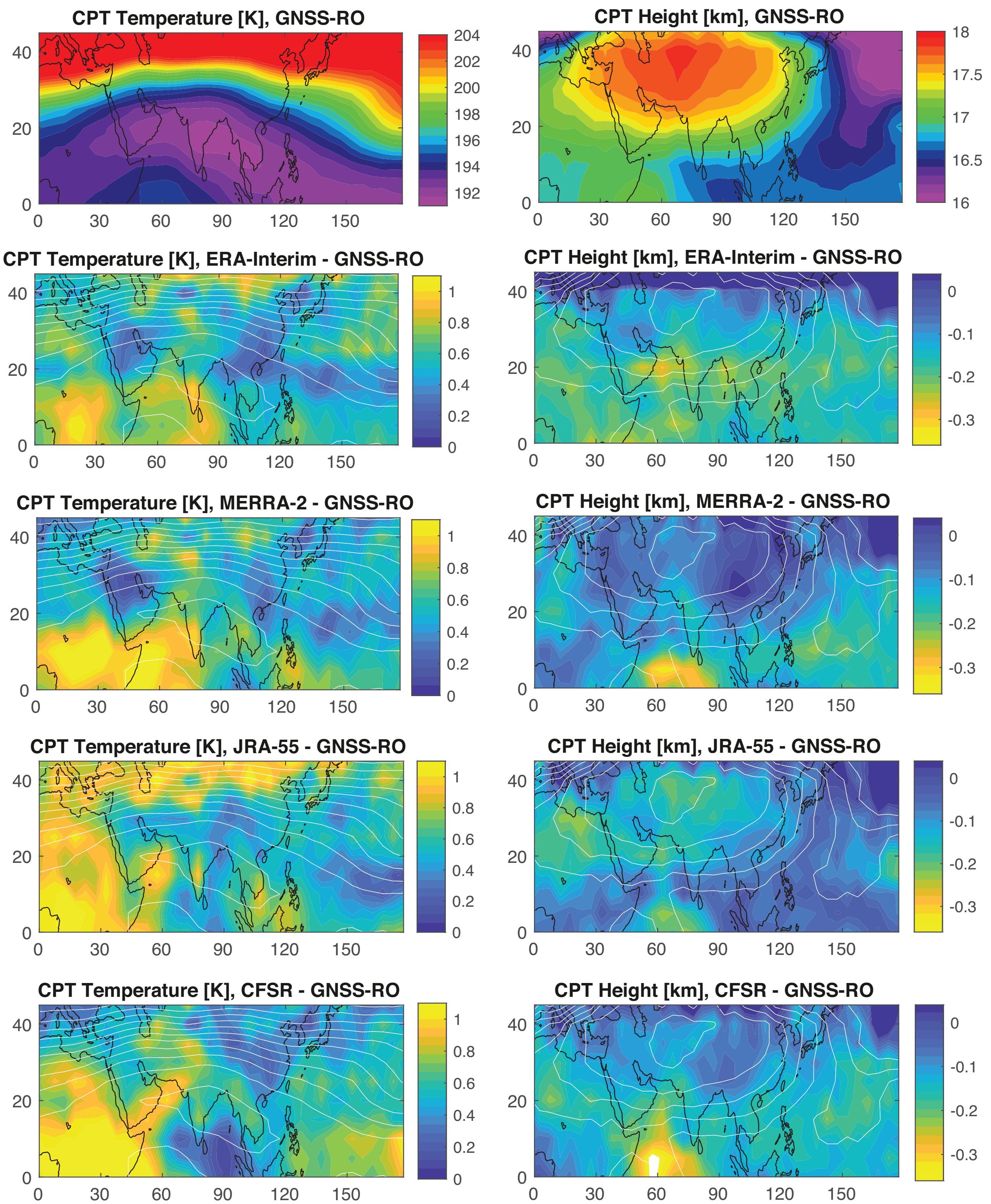


Figure 8.51: *CPT temperature [K] and height [km] for the SASM region showing GNSS-RO observations and anomalies for the reanalyses ERA-Interim, JRA-55, MERRA-2, and CFSR for the time period JJA 2007-2010. For further details see Section 8.2. White contour lines show the climatology data from GNSS-RO; left: 2 K intervals and minimum line is 190 K, and right: 0.25 km intervals starting from 16 km altitude.*

recent reanalyses with horizontal model grid spacings finer than 1° (*cf.*, **Fig. 5** and related text in *Nützel et al., 2016*). *Manney et al.* (2021) also discuss bimodality and show no evidence for it (especially positional as opposed to shape-related) in any of the "modern" reanalyses used in the "SASM anticyclone moments analysis" section below.

As discussed in *Section 2.3.1*, MERRA and MERRA-2 each provide two data assimilation products, respectively referred to as ANA (a standard 3D-FGAT analysis state) and ASM (in which analysis increments are applied gradually via IAU; *Bloom et al.*, 1996). The distributions of SASM anticyclone centre locations for MERRA and MERRA-2 above are based on the corresponding ASM data sets, as these are expected to have a greater degree of physical consistency among all variables (see Discussion in *Section 2.3.1*; see also technical note at **https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/docs/ANAvsASM.pdf**). The ANA products, by contrast, are expected to be in better agreement with observations at the analysis time. In the following we examine differences between the ASM and ANA products for MERRA and MERRA-2.

The distribution of SASM anticyclone centre locations based on MERRA-ANA is evidently different from that based on MERRA-ASM, as the maximum density near 57.5°E in MERRA-ASM is split into two maxima located at 50°E and 62.5°E in MERRA-ANA (**Fig. 8.50**). This difference highlights the importance of the assimilation technique on the distribution of the SASM anticyclone centre. Moreover, while analysing MERRA-2-ANA in the monsoon region in more detail, we found enhanced geopotential height values consistently located along the steep orography of the Himalaya Mountains. This is an artefact of the MERRA-2-ANA data set on pressure levels that was introduced during the conversion from model levels to pressure levels (*personal communication by Krzysztof Wargan, GMAO*). Consequently, we do not show the distribution of SASM anticyclone centre locations for MERRA-2-ANA in this figure here.

Overall, these results suggest that the bimodality of the SASM anticyclone centre location on short time scales (days) as identified in previous studies is mainly a peculiarity of NCEP-NCAR R1. The presented sensitivity of the SASM anticyclone centre in the reanalyses may impact previous findings (*e.g.*, with respect to locations and trace gas distributions) that have been obtained by using older reanalyses in particular NCEP-NCAR R1 (see also *Nützel et al.*, 2016 for further discussion).

SASM anticyclone tropopause

Analysing the tropopause characteristics over the SASM region (**Fig. 8.51**) reveals that the CPT temperature has its minimum (~192 K) over the Indian Subcontinent, Bay of Bengal, and the Indochina Peninsula, where convection is most active. In contrast, the maximum CPT height of ~18 km is found along the northern flank of the SASM anticyclone near 75°E. Differences between the reanalyses ERA-Interim, JRA-55, MERRA-2, and CFSR and GNSS-RO observations reveal systematically higher CPT temperatures (0.1 - 1 K) and lower CPT heights (0.2 km) in all four reanalyses compared to GNSS-RO observations, in accordance with results presented in *Section 8.2*. Different spatial patterns are evident in the CPT height anomaly fields, with JRA-55 having the largest difference (more than 0.2 km) and MERRA-2 the smallest difference (less than 0.1 km) relative to GNSS-RO.

SASM anticyclone moments analysis

A moments analysis (*e.g.*, *Matthewman et al.*, 2009), as well as determination of area and edge characteristics, has been done for the SASM anticyclone as defined by

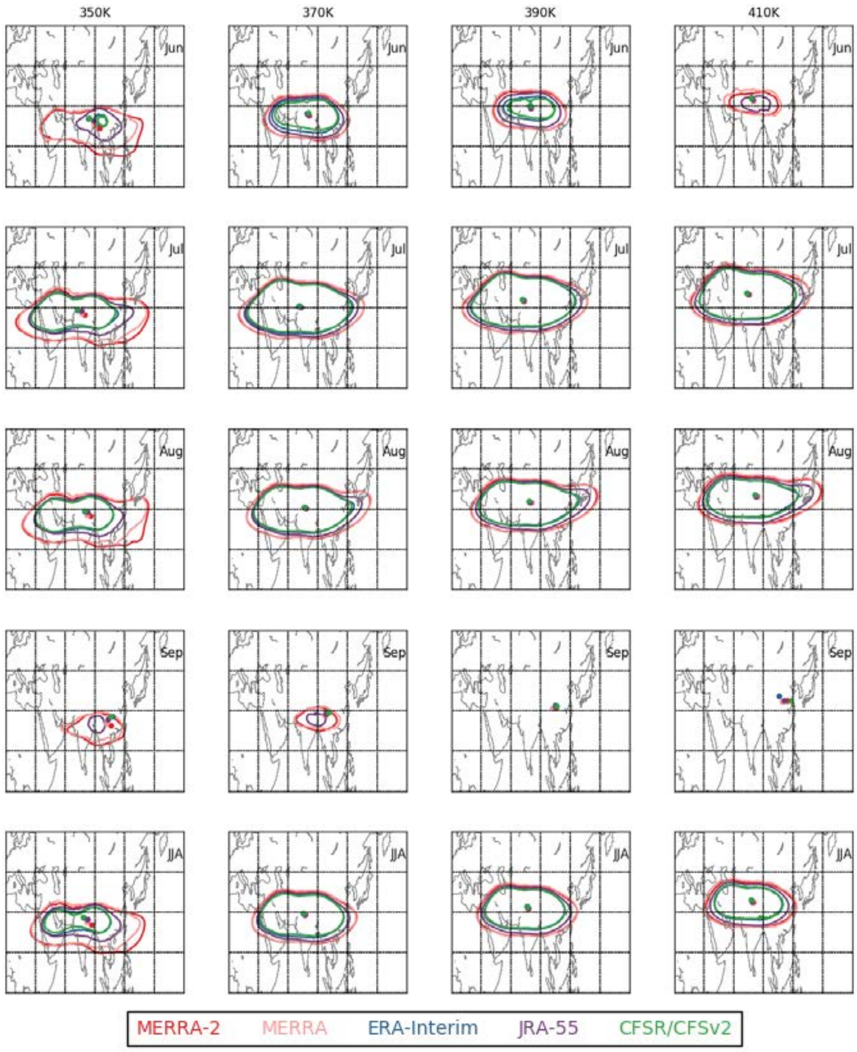


Figure 8.52: *Climatological (1979-2015) means of SASM anticyclone edge (contours) and centroid (symbols) locations for May through September and JJA based on MERRA-2 (red), MERRA (pink), ERA-Interim (blue), JRA-55 (purple), and CFSR/CFSv2 (green). The isentropic levels are (left to right columns) 350, 370, 390, and 410 K.*

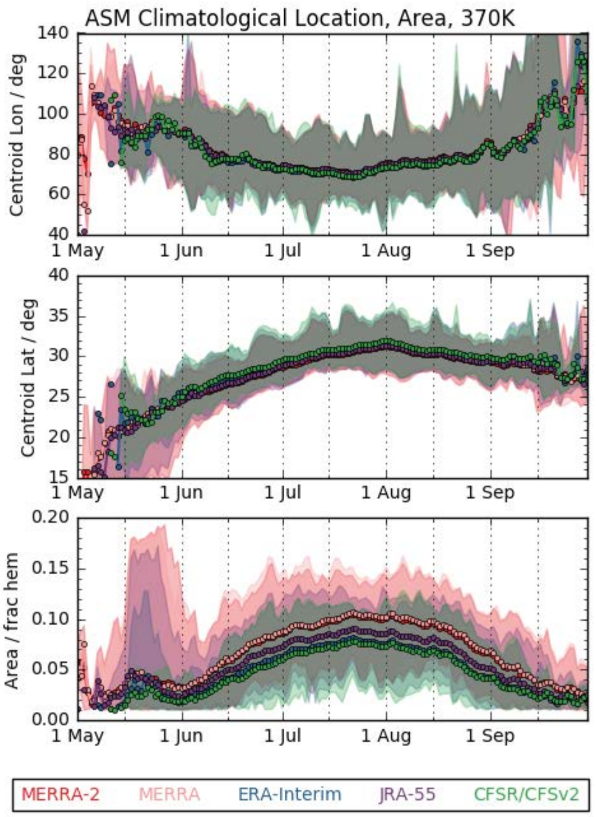


Figure 8.53: *Climatological (1979 - 2015) time series of centroid position and area of the SASM anticyclone at 370 K. Envelopes show the ranges of minimum-maximum values for the corresponding reanalyses.*

Montgomery Streamfunction (MSF; *Montgomery*, 1937) on the 350, 370, 390, and 410 K isentropic surfaces, for MERRA-2, MERRA, ERA-Interim, JRA-55, and CFSR. As discussed by *Santee et al.* (2017), MSF is a streamfunction on isentropic surfaces analogous to geopotential height on isobaric surfaces, and the values used to define the boundary of the SASM at each isentropic level (given by *Santee et al.*, 2017) were determined by examination of their relationship to wind speeds in order to approximate the region where trace gases are relatively confined. We assess the climatology and variability of the SASM anticyclone by analyzing the moments and related diagnostics, including centroid location, angle, aspect ratio, excess kurtosis, area, and edge locations and characteristics. The analysis includes climatology, interannual variability and trends, relationships between diagnostics, onset and decay dates, relationships to upper tropospheric jet streams, and relationships of SASM changes to natural modes of variability such as ENSO. A paper has been submitted on this material (*Manney et al.*, 2021), with some example figures presented below. The results are consistent with those illustrated below when time series for MERRA-2, ERA-Interim, and JRA-55 are extended through 2018.

Figure 8.52 gives a climatological (1979 through 2015) overview of the monthly and seasonal (JJA) SASM anticyclone edge and centroid locations based on MSF on the isentropic surfaces listed above. MERRA and MERRA-2 show larger SASM anticyclones than the other reanalyses, with most of the difference being on the equatorward boundary in the region of monsoon easterlies. The differences amongst analyses are largest at 350 K and decrease for each higher level. These differences appear to be consistent with the stronger monsoon easterlies in MERRA and MERRA-2 found by *Manney et al.* (2017). The mean centroid locations are generally very close in all reanalyses, though these locations are shifted to slightly lower latitudes in MERRA and MERRA-2 in some cases More discussion on the possible cause of this can be found in *Section 8.8.3*.

Figure 8.53 shows climatological time series of the SASM centroid location and area at 370 K. The centroid locations usually agree quite well among the reanalyses (as do higher-order moments shown by *Manney et al.*, 2021). Substantially larger areas are seen in MERRA and MERRA-2 than in the other reanalyses, with CFSR/CFSv2 and ERA-Interim showing the smallest areas. At this level, MERRA, MERRA-2 and JRA-55 have slightly lower centroid latitudes than ERA-I and CFSR, consistent with **Figure 8.52**.

Figure 8.54 shows climatological seasonal frequency distributions of the centroid location and area at 370 K. The most striking difference among the reanalyses is larger areas for MERRA and MERRA-2 than for the other reanalyses. The slightly lower centroid latitudes in MERRA, MERRA-2, and JRA-55 are again apparent. Consistent with **Figure 8.52**, the differences are larger at 350 K and smaller at the higher levels (*Manney et al.*, 2021).

Figure 8.55 shows time series and trends of the JJA mean SASM anticyclone areas for 1979 through 2015. In addition to

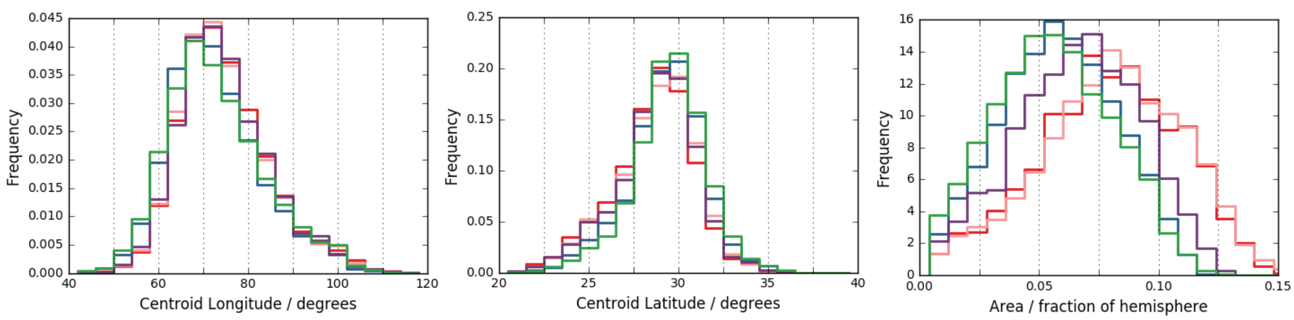


Figure 8.54: *Histograms of climatological seasonal (JJA) SASM anticyclone diagnostics at 370 K, left to right: centroid longitude, centroid latitude, and area.*

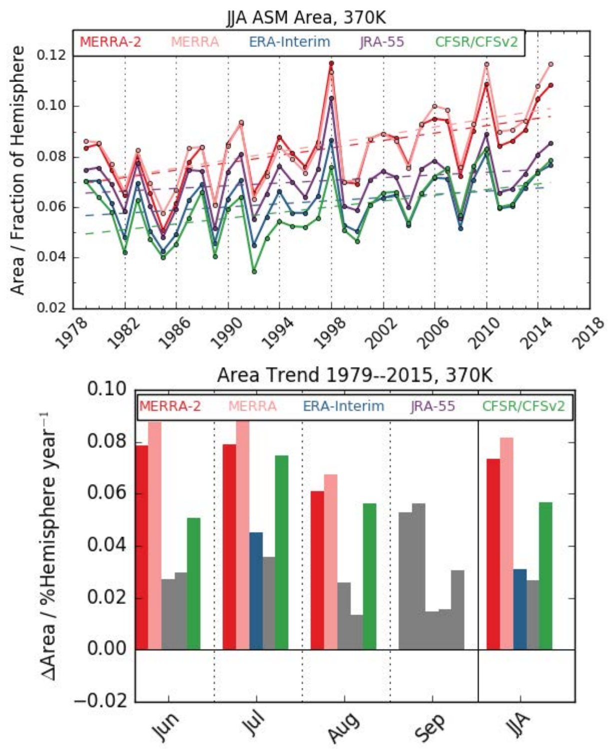


***Figure 8.55:** Time series (top) and trend (bottom) of JJA-mean SASM area at 370 K during 1979 - 2015. Top: Dashed lines show linear fits; calculations are based on ordinary least squares with permutation analysis. Bottom: Bars show the slopes of the linear fits (top figure), colored according to the key at the top when fits are significant at the 90 % confidence level.*

the overall larger areas seen for MERRA and MERRA-2 in the previous figures, those reanalyses also show stronger trends in the SASM anticyclone area than the other reanalyses. An increasing trend is seen in all of the reanalyses, but these trends are only marginally significant (according to permutation analysis using 100,000 re-samplings). ERA-Interim typically shows the weakest trends. Similar patterns are seen at the other levels, with area being the only diagnostic that consistently exhibits significant trends (that is, the higher order moments shown by *Manney et al,* 2021, generally do not show robust trends).

Figure 8.56 shows the start and end dates of the SASM defined as the first and last periods with an SASM anticyclone area greater than 1% of a hemisphere for at least 20 consecutive days on the corresponding isentropic level. The duration (end minus start date) of the SASM season is also shown. Consistent with the larger SASM anticyclone areas in MERRA and MERRA-2, these reanalyses show earlier start dates, and longer durations than the other reanalyses, with the largest differences at 350 K. Interannual variability in these diagnostics is generally consistent among the reanalyses, except at 350 K, where MERRA and MERRA-2 show some unique fluctuations (*e.g.*, around 2000 to 2006 in both start dates and durations).

8.8.2 Vertical velocity

Climatologically, upward motion prevails in the UT on the eastern side of the SASM, whereas downward motion prevails on the western side (*Pan et al.*, 2016; their **Fig. 10a** and references therein). Here, we analyse differences amongst reanalysis vertical velocity products in the UT during the SASM. Vertical velocities in this Section are expressed in the pressure vertical coordinate (ω; *Section 8.4*) and are computed from the analysed horizontal winds via the continuity equation.The four recent reanalyses agree well on the overall

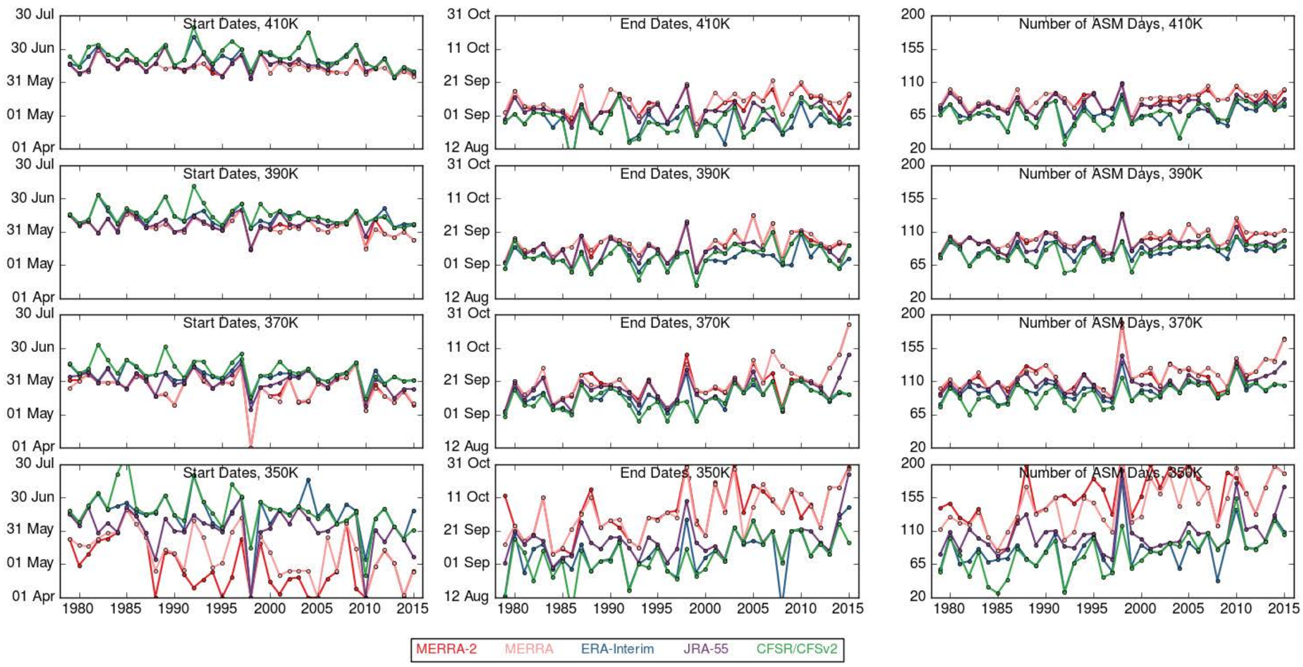


***Figure 8.56:** Time series of SASM start dates (left), end dates (center), and durations (right) during 1979 - 2015. SASM start (end) dates are defined by the appearance (disappearance) of a SASM anticyclone with area greater than 1 % of a hemisphere for at least 20 consecutive days.*

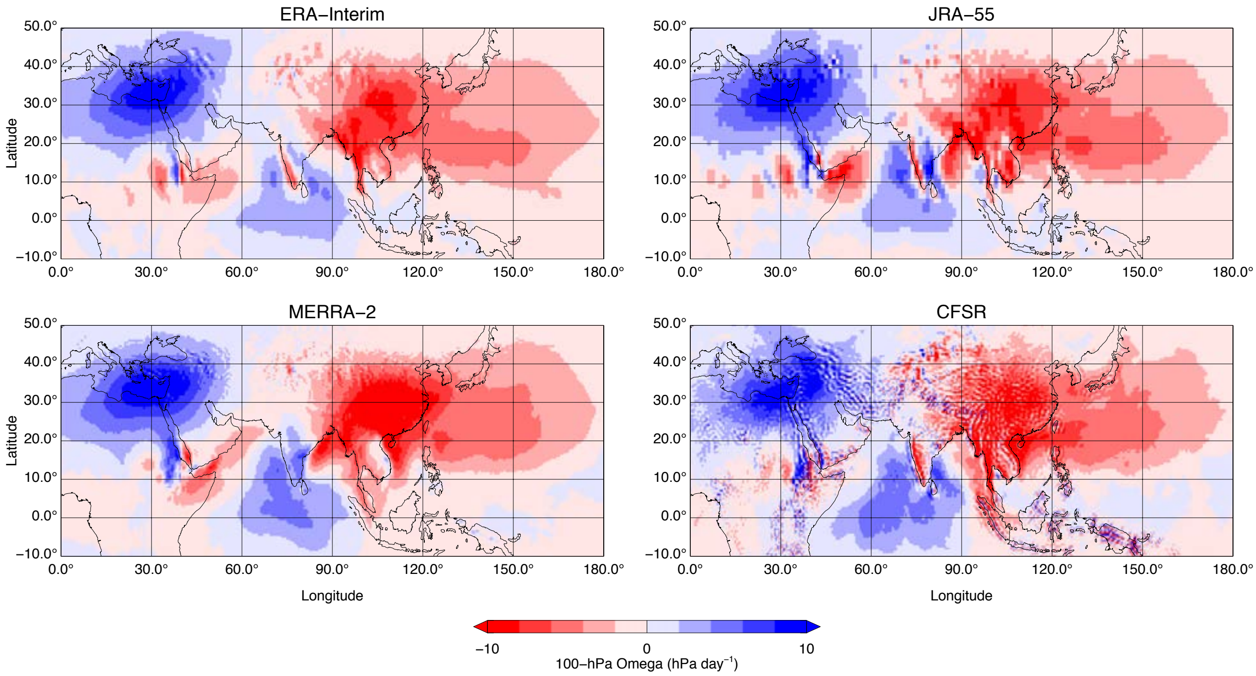


***Figure 8.57:** Omega (Pa/s, shading) at 100 hPa for the SASM region (ERA-I, JRA-55, MERRA-2, CFSR) during June-August 1980-2010.*

spatial pattern, with rising motion along the eastern flank and sinking motion along the western flank of the SASM anticyclone on the 100 hPa isobaric surface (**Fig. 8.57**). There are some regional differences, however, such as over the western coast of India (where all reanalyses indicate rising air except for MERRA-2) and above the Bay of Bengal (where the reanalyses indicate different spatial distributions and lateral gradients of vertical motion). Vertical velocities at 100 hPa are noticeably noisier in CFSR than in the other reanalyses, which may be due to the relatively fine horizontal resolution and topography effects in CFSR.

Intercomparison of ω within the SASM region (**Fig. 8.58**) reveals fewer differences among the modern reanalyses in contrast to heating rates (*Section 8.8.3*, **Fig.8.59**). It follows that more consistent results can be expected

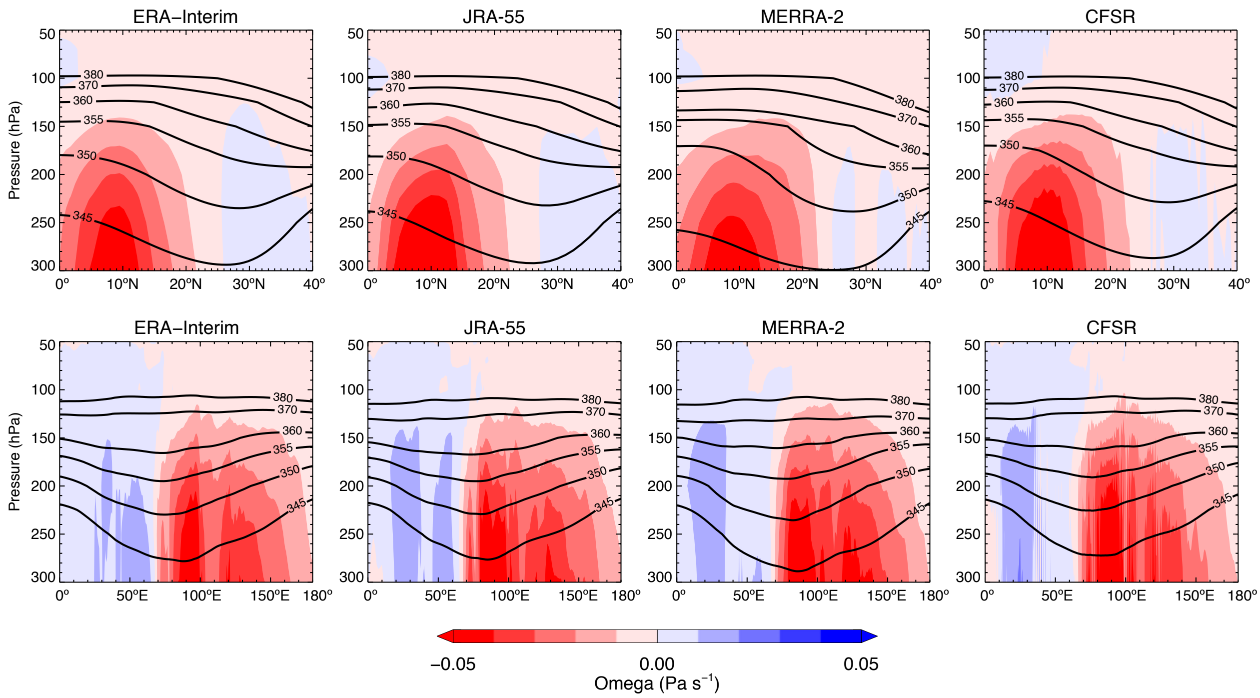


***Figure 8.58:** Zonal and meridional mean omega (Pa s⁻¹, shading) and potential temperature (K, contours) during JJA 1980-2010 within the SASM region. Zonal means are calculated over 0-180°E; area-weighted meridional means over 0-40°N.*

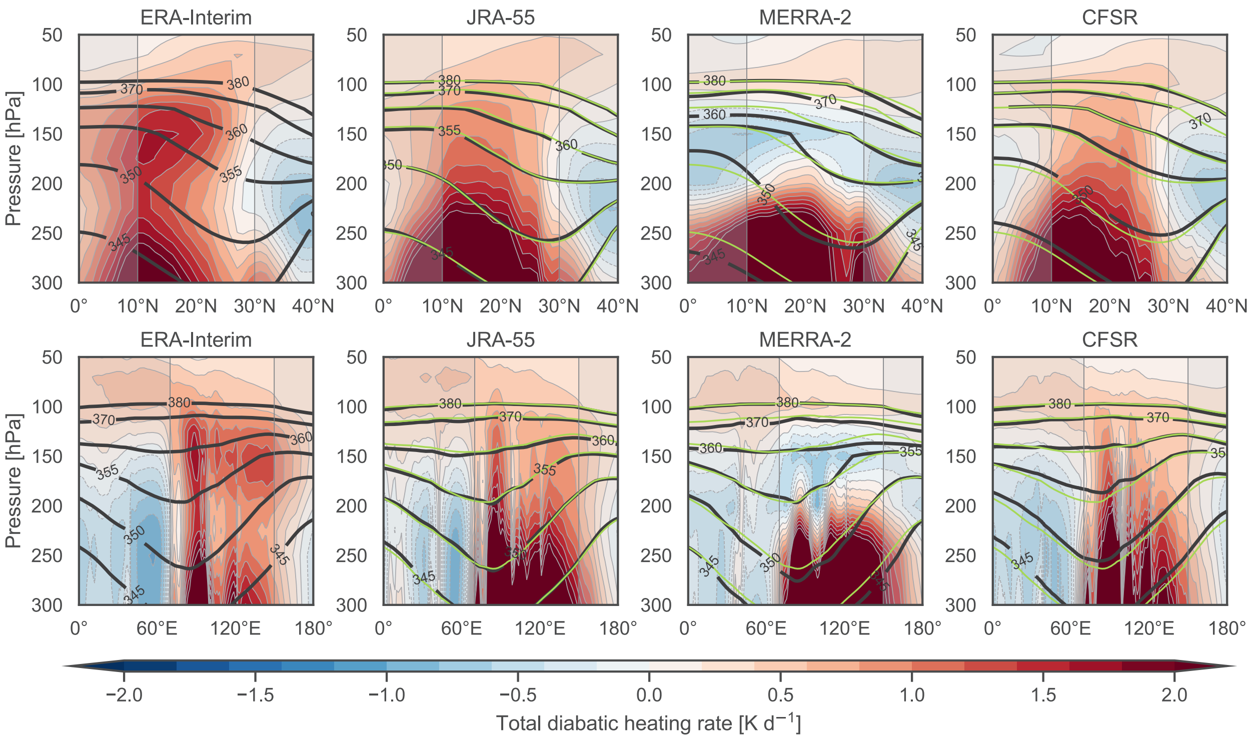


Figure 8.59: *Zonal and meridional mean total diabatic heating rates due to parameterized physics (K day^{-1}, shading) and potential temperature (K, contours) during JJA 1980 - 2010 within the SASM region. Zonal means are calculated over 70 - 150 °E (vertical lines in lower panels); area-weighted meridional means over 10 - 30 - °N (vertical lines in upper panels). ERA-Interim potential temperature contours are shown in light green on the JRA-55, MERRA-2, and CFSR panels for ease of comparison.*

on average when using ω from different reanalyses to drive chemical transport models (often referred to as the kinematic approach) in this region than when using diabatic heating rates (the diabatic approach). However, it should be noted that kinematic transport calculations tend to be noisier than diabatic transport calculations for the stratosphere, TTL, and the SASM anticyclone (*Garny and Randel*, 2016; *Krüger et al.*, 2008; *Schoeberl et al.*, 2003); thus, we may still expect a large spread between results based on kinematic and diabatic simulations that use products from the same reanalysis (*e.g.*, *Bergman et al.*, 2013; *Ploeger et al.*, 2010). Recent results suggest that the latter is much improved in ERA5, with greater consistency between the diabatic and kinematic approaches relative to ERA-Interim (*Legras and Bucci*, 2020; *Hoffman et al., 2019*).

8.8.3 Diabatic heating

Figure 8.59 illustrates the diabatic heating distribution within the UTLS above the SASM region. Here, diabatic heating corresponds to the total diabatic heating as introduced in *Section 8.3*, and includes radiative transfer, moist physics, and other parameterized processes that affect the temperature budget. Please also see the footnote on diabatic heating rates in reanalyses in *Section 12.1.3*. Large differences exist among the four modern reanalyses. Perhaps the most striking difference concerns the location and magnitude of positive heating rates within the zonal-mean diabatic 'chimney', which connects the convective detrainment zone in the UT to dynamical ascent balanced by radiative heating in the LS. The maximum values in this chimney are located near 10 - 15 °N in ERA-Interim, but are more widespread and shifted progressively further northward in JRA-55 and CFSR. The local maximum in heating at 200 hPa is also shifted northward in MERRA-2 (~ 20 °N) relative to ERA-Interim (~ 11 °N); however, the diabatic chimney is completely missing in the MERRA-2 time mean above this level. The layer of time-mean diabatic cooling that overlays the convective core of the monsoon in MERRA-2 is related to cloud radiative effects (*cf.*, **Figs. A8.7** and **A8.8**), as discussed also for the full tropical domain in *Section 8.4*. Cloud radiative effects enhance heating at 250 hPa, especially around 20 °N; however, LW cooling above deep convective clouds coupled with shallower convective heating (see also **Fig. A8.9**) inhibit diabatic ascent above this level. This inhibition has also been shown to affect kinematic ascent based on pressure vertical velocities in the earlier MERRA reanalysis (*Bergman et al.*, 2013).

Not only do the diabatic heating biases in MERRA-2 restrict vertical transport between the convective detrainment layer and the LS, but they also result in an evident deformation of the SASM upper tropospheric high (as indicated here by differences in the potential temperature contours in **Fig. 8.59**). Relative to ERA-Interim, MERRA-2 shows greater heating between 250 hPa and 300 hPa, including a secondary centre near 30 °N, as well as greater

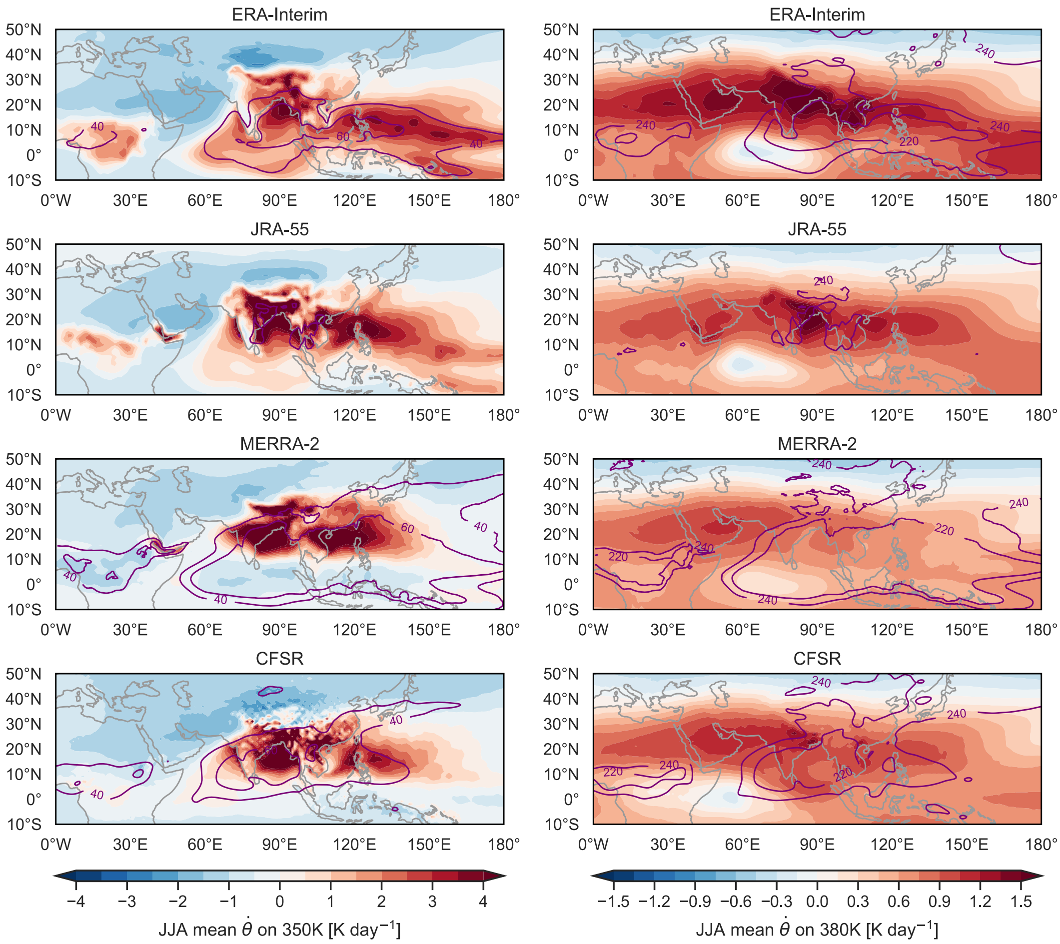


Figure 8.60: *Left panels: Diabatic heating on the 350 K isentropic surface (K day^{-1}, shading) and LWCRE at the TOA (W m^{-2}, purple contours at 40 W m^{-2} and 60 W m^{-2}) during JJA 1980 - 2010. Right panels: same as left panels, but for diabatic heating on the 380 K isentropic surface (shading) and OLR at the TOA (W m^{-2}, purple contours at 220 W m^{-2} and 240 W m^{-2}).*

cooling between 150 hPa and 200 hPa south of 10 °N. These differences bend the 350 K isentropic surface downward toward larger pressures between about 15 - 30 °N and upward toward smaller pressures south of about 12 °N. Together with the nearly isobaric nature of model levels at these altitudes (see *Chapter 2*, *Appendix A*), this deformation may help explain why the SASM anticyclone based on isentropic MSF is relatively distinctive in MERRA-2 (**Fig. 8.52**), while the SASM anticyclone based on isobaric geopotential height is quite consistent between MERRA-2 and other reanalyses (**Fig. 8.47**).

Other differences in the vertical distributions of diabatic heating within the SASM region include the existence of a secondary maximum in diabatic heating of more than 1.5 K day^{-1} in ERA-Interim at the 150 hPa pressure level. This feature, which is not reproduced by the other reanalyses shown in **Figure 8.59** is related mainly to cloud radiative effects and latent heating associated with cloud formation in the lower TTL in ERA-Interim (as discussed for the entire tropical domain in *Sect. 8.4*; see also **Figs. A8.7 - A8.9**). In addition to the local maximum at this level in the latitudinal distribution (10 - 15 °N), similar features are evident in the longitudinal distribution around 80 - 90 °E (Bay of Bengal) and between 120 - 150 °E (western North Pacific). Centres of convective heating are also evident at 300 hPa in ERA-Interim over these two regions. All of the reanalyses reproduce these two centres of convective heating, but with substantial differences in the depth of the heating (deepest in JRA-55; shallowest in MERRA-2) and some differences in the precise east–west location, especially for the centre over the Bay of Bengal (furthest east in ERA-Interim; furthest west in MERRA-2).

To further explore these differences, **Figure 8.60** shows maps of diabatic heating rates on the 350 K (left) and 380 K (right) isentropic surfaces, together with distributions of OLR (right) and the LWCRE (left) at the nominal TOA. At 350 K, positive heating rates within the broader Asian monsoon region (comprising the South Asian, East Asian, and western North Pacific monsoons) are centred more toward the tropics in ERA-Interim. The distribution is shifted northward in JRA-55, with enhancements relative to ERA-Interim over the South China Sea, Southeast Asia, and the south slope of the Himalayas, but weaker heating south of about 10 °N. The northward shift relative to ERA-Interim is even more pronounced in MERRA-2 and CFSR, which show larger and more organized heating rates over China (suggesting that effects of the East Asian monsoon rain band extend to higher altitudes in these reanalyses) and an evident northward shift in the southern boundary of positive heating rates. The area of strong positive heating rates over the SASM is bounded to the north and west by relatively strong negative heating rates at 350 K. Despite some differences in magnitude and the precise distribution, cooling to the north of the SASM is broadly similar among the reanalyses. That to the west is less consistent. ERA-Interim and JRA-55 both show relatively strong cooling over the southern portion of the Arabian Peninsula. This centre is displaced to the north and east in CFSR and MERRA-2, and is particularly weak in the latter.

At 380 K, the reanalyses all show zonally-elongated bands of positive diabatic heating rates centred near 20 ° N. This band is roughly collocated with the tropical easterly jet along the southern edge of the SASM anticyclone (see, *e.g.*, **Fig. 8.48**). Heating rates within this band of relatively strong heating are larger on average in ERA-Interim than in the other three reanalyses (**Fig. 8.60**), although all four reanalyses show relatively strong heating around 50 - 60 ° E over the southern portion of the Arabian Peninsula. Based on **Figure 8.59**, convection and associated anvil clouds are relatively infrequent in this region in comparison to 70 - 150 ° E, meaning larger upwelling LW fluxes from the troposphere. The resulting enhancement in the convergence of LW radiation in the LS causes stronger radiative heating at 380 K. Differences are more pronounced to the east of this feature, where ERA-Interim, JRA-55, and, to a lesser extent, CFSR show locally enhanced heating around 70 - 90 ° E, while MERRA-2 shows a local minimum. Whereas the enhanced heating in this region arises mainly from cloud radiative effects in ERA-Interim, non-radiative heating at the tops of very deep convection plays a more consequential role in JRA-55 and CFSR (**Fig. 8.59**; see also *Sect. 8.4* and **Fig. A8.10**). The local minimum in MERRA-2 is also linked to cloud radiative effects, namely the attenuation of upwelling LW radiation by extensive convective anvil clouds. Another notable difference among the reanalysis diabatic heating rates at 380 K is the local minimum centred over the equatorial western Indian Ocean near 60 ° E (**Fig. 8.60**). This feature is rooted in the effects of parameterized turbulent mixing (see also **Fig. A8.10** in the *Appendix*), and is therefore strongest in ERA-Interim (which has the largest temperature tendencies due to parameterized turbulence; see **Fig. 8.27** and related discussion) and weakest in MERRA-2 (which has the smallest).

Among the most important differences for diabatically-driven transport studies are the locations of strong UTLS heating associated with the SASM and the western North Pacific monsoon, as different reanalyses are known to imply very different distributions of convective sources for cross-tropopause transport into the stratosphere from the Asian monsoon region (*e.g.*, *Wright et al.*, 2011). For South Asia, ERA-Interim produces maximum heating at 350 K near the northern and northeastern coastlines of the Bay of Bengal (BoB). In JRA-55 and MERRA-2 this heating is displaced more toward the northwestern coastline of the BoB, while in CFSR it is centred over the BoB itself. JRA-55 also shows strong heating over the south slope of the Himalayas, whereas this heating is shifted further north over the southern Tibetan Plateau in ERA-Interim and MERRA-2. This difference is also evident in the zonal-mean distributions shown in **Figure 8.59**, where ERA-Interim and MERRA-2 show a clearer separation between enhanced heating at 30 ° N and that at lower latitudes than JRA-55 (note also that the local maximum in pressure along the 350 K isentropic contour is located near 30 ° N in this region during JJA, indicating a local minimum in altitude). The distribution at 350 K in CFSR is much noisier (**Fig. 8.60**), but appears to be more consistent with JRA-55 in that the largest heating rates are centred over the south slope of the Himalayas. The noisiness of the diabatic heating distribution in CFSR even after taking the 1980 - 2010 climatological mean suggests that the distribution of deep convection in CFSR may be very sensitive to the complex topography of this region. Over the western North Pacific, the primary difference is in the latitude of enhanced heating at 350 K. Whereas the strongest heating in this region is at approximately the same latitude as the Philippines in ERA-Interim, it is centred north of the Philippines in JRA-55 and MERRA-2, with the distribution in CFSR located between the two. Differences at 380 K may also be influential in diagnosing the distribution of convective sources for air crossing the stratosphere, particularly that ERA-Interim, JRA-55, and CFSR show local maxima of varying magnitudes at this level near the most active convective regions while MERRA-2 shows local minima (right panels of **Fig. 8.60**).

Diabatic heating distributions outside of the core Asian monsoon domain also show substantial differences, especially at 350 K (left panels of **Fig. 8.60**). Whereas positive heating rates extend southward across the equator over the tropical Indian and Pacific Oceans in ERA-Interim and JRA-55, these features are missing in MERRA-2 and CFSR. This is despite the fact that MERRA-2 evidently produces strong convection in these regions, as indicated by large values of LWCRE. Indeed, whereas positive heating rates at 350 K are tightly collocated with large values of LWCRE in ERA-Interim, they are limited to the northwestern edge of large values of LWCRE in MERRA-2. This difference may again be understood in terms of shallower convective anvil clouds and associated radiative effects in MERRA-2 (**Fig. 8.24** and related discussion). Toward the western edge of the domain, ERA-Interim produces strong heating at 350 K over much of equatorial Africa. This feature is present but weaker in JRA-55, largely absent in CFSR, and replaced by substantial cooling in MERRA-2. Meanwhile, JRA-55 and MERRA-2 have centres of strong heating over the southern part of the Red Sea and the Gulf of Aden that are absent from ERA-Interim and CFSR.

8.8.4 Transport

The differences in diabatic heating rates shown above manifest in differences in Lagrangian transport calculations

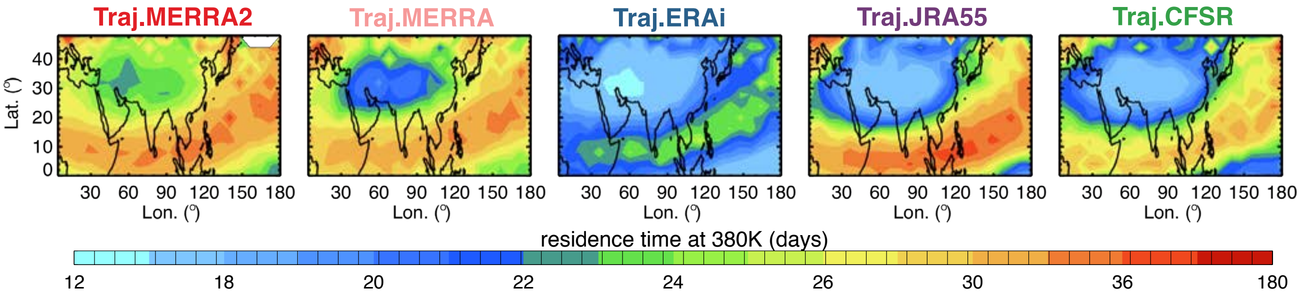


***Figure 8.61:** Residence time (days) between 370 K and 380 K, displayed at 380 K, during JJA 2005 - 2015 (for details see **Fig. 8.31**; Section 8.5.2).*

driven by diabatic vertical velocities within the SASM region. As an extreme example, the negative heating rates below 370 K in MERRA-2 mean that diabatic transport calculations are impractical unless they are initialized at the 370 K potential temperature or above (*Section 8.5.2*). **Figure 8.61** shows the residence time for parcels traveling between the 370 K and 380 K isentropic levels within the SASM region during JJA (see *Section 8.5.2* for more details). The differences can be directly linked to differences in diabatic heating rates. For example, ERA-Interim, which has the strongest heating rates in the TTL (**Fig. 8.59**), shows the shortest residence times (often less than 15 days). Conversely, MERRA-2, which has the weakest heating rates within the TTL, shows the longest residence times. The minimum residence time within the SASM domain based on MERRA-2 is ~22 days, and many locations within the anticyclone show mean residence times greater than 25 days. Such long residence times are only found along the southeastern edge of the anticyclone in ERA-Interim. Meanwhile, CFSR and JRA-55 show relatively homogeneous residence time distributions throughout the SASM anticyclone region, whereas MERRA-2, MERRA, and ERA-Interim show local residence time minima (indicating faster uplift) in the western flank of the anticyclone.

8.8.5 Ozone

A pronounced local minimum in total column ozone during boreal summer has led researchers to dub the SASM region an 'ozone valley' (*Bian et al.*, 2011; *Zhou et al.*, 1995). Much of this regional-scale minimum in total column ozone is due to low ozone mixing ratios within the UTLS anticyclone (*Santee et al.*, 2017; *Park et al.*, 2007). The low ozone concentrations are thought to result from extensive convective detrainment of ozone-poor tropospheric air and subsequent confinement within the SASM anticyclone. Ozone is parameterized and assimilated in reanalyses, as outlined and evaluated in *Chapter 2* and *Chapter 4* of this report.

Observational data sets

SWOOSH is an observationally-based analysis of ozone and water vapor based on a limb-sounding and solar occultation instruments from the 1980s until now. For the period we use (2005 - 2018) it is almost exactly the same as Aura MLS. The data set itself has been described by *Davis et al.* (2016).

Ozone

Figure 8.62 shows climatological spatial distributions of ozone volume mixing ratios at 100 hPa in the SASM region during JJA. ERA5, ERA-Interim, JRA-55, MERRA-2, and CFSR all show relatively low ozone concentrations above the SASM region, although the magnitudes and spatial distributions of ozone within the anticyclone vary. Averages within the area bounded by 30 - 120° E and 20 - 40° N range from approximately 190 ppbv (JRA-55) to 325 ppbv (CFSR). All are larger than the average based on Aura MLS during 2005 - 2018 (150 ppbv), as illustrated in **Figure 8.62a** by the SWOOSH distribution (*Davis et al.*, 2016). For the period used here (2005 - 2018), the SWOOSH distribution for the 2005 - 2018 period shown in **Figure 8.62a** is almost entirely determined by Aura MLS. It is therefore important to note that Aura MLS ozone retrievals have been assimilated during recent years by ERA-Interim, ERA5, and MERRA-2 (*Chapter 4*; **Fig 4.2**). ERA5 and ERA-Interim show elongated minima in ozone mixing ratios along the southern edge of the anticyclone, as does SWOOSH. By contrast, JRA-55, MERRA-2, and CFSR produce minima centred more over the Bay of Bengal, to the southeast of the anticyclone.

Figure 8.63 shows latitude–pressure cross-sections of ozone anomalies within the SASM region (30 - 120° E) relative to zonal-mean volume mixing ratios within the same latitude band. This view provides more information on the vertical and meridional structure of the SASM 'ozone valley' within the UTLS. Negative ozone anomalies correspond well to positive anomalies in geopotential height, with the largest anomalies typically located in the upper portion of the anticyclone and slightly to the south of its centre. The ozone valley remains least pronounced in CFSR, for which the largest anomalies are farther south of the anticyclone centre and at a slightly lower altitude than in the other reanalyses. However, comparison with SWOOSH again suggests that all five reanalyses underestimate the amplitude of negative anomalies associated with this feature. ERA5, ERA-Interim and JRA-55 show substantial negative

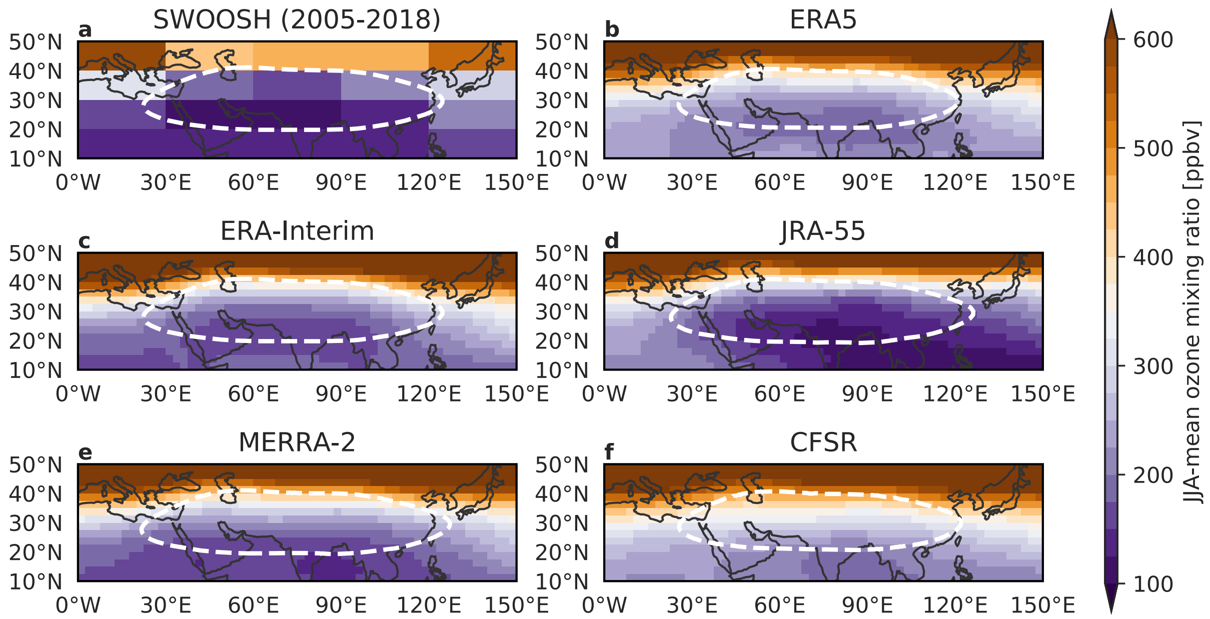


Figure 8.62: *Spatial distributions of JJA-mean ozone mixing ratio [ppbv] on the 100 hPa isobaric surface based on (a) SWOOSH (Davis et al., 2016), (b) ERA5, (c) ERA-Interim, (d) JRA-55, (e) MERRA-2, and (f) CFSR. Reanalysis ozone products are averaged over 1980-2010; SWOOSH data are averaged over 2005-2018. The 16700 m contour in 100-hPa geopotential height based on the corresponding data sets is shown as a white dashed line in each panel for context. Geopotential height in panel (a) is from ERA-Interim.*

ozone anomalies (-10% or larger) extending downward to 250 or even 300 hPa within the SASM region, whereas anomalies are more confined to the lower TTL ($p \leq 200$ hPa) in MERRA-2. None of the reanalyses reproduce observed positive anomalies relative to the zonal mean at lower altitudes, which are located both below and to the south of the anticyclone core according to the SWOOSH distribution. These positive anomalies may be related to anthropogenic emissions of ozone precursor species and subsequent convective transport. As these emissions are not represented in the simple ozone schemes used in the forecast models (*Chapter 2*; **Table 2.10**), such effects could only enter the reanalysis products through the data assimilation. Although temporal variations in SASM ozone are not evaluated here, users should be aware that changes in assimilated ozone data over time (*Chapter 4*; **Figs. 4.1** and **4.2**), especially vertically-resolved profile data (Fig. 4-2), may lead to discontinuities in reanalysis representations of the SASM ozone valley.

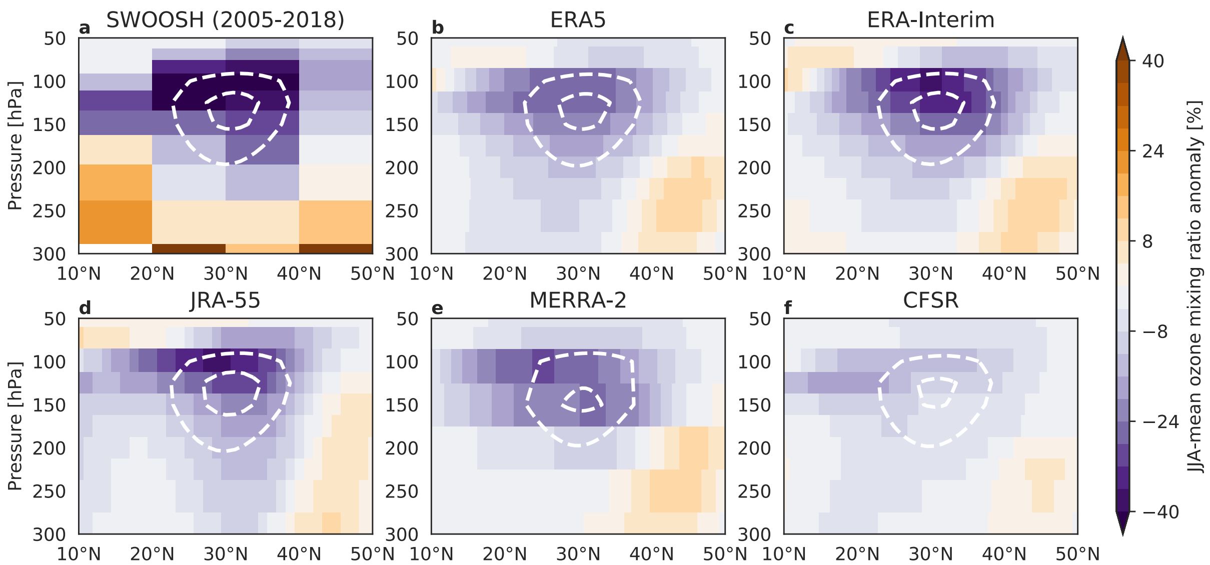


Figure 8.63: *Latitude–pressure distributions of normalized anomalies [%] in JJA-mean ozone mixing ratios within 30-120°E relative to zonal-mean values for the corresponding zonal bands derived using (a) SWOOSH, (b) ERA5, (c) ERA-Interim, (d) JRA-55, (e) MERRA-2, and (f) CFSR. Reanalysis ozone products are averaged over 1980-2010; SWOOSH data are averaged over 2005-2018. Absolute geopotential height anomalies in the 30-120°E band relative to the zonal mean are shown as white dashed contours at values of 100 and 125 m. Geopotential height anomalies relative to the zonal mean in panel (a) are from ERA-Interim.*

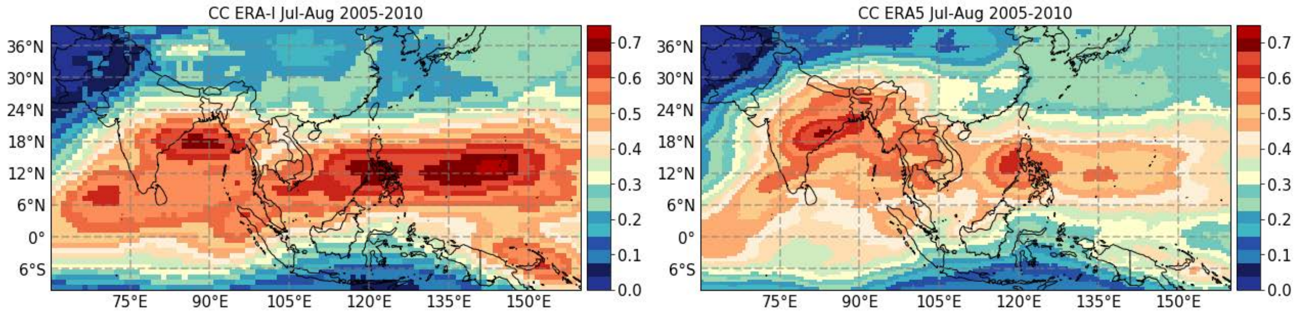


Figure 8.64: *Cloud cover in [fraction] at the level of the maximum cloud in the SASM domain for ERA-Interim (left) and ERA5 (right), July - August 2005 - 2010.*

8.8.6 Regional analysis of clouds and radiative effects

In this Section, we focus on a regional analysis of cloud and radiative properties in the SASM, which is recognized as the location of the largest discrepancies among reanalyses and climate models (*Tissier and Legras*, 2016; *Johansson et al.*, 2015; *Heath et al.,* 2014).

We compare the five reanalyses CSFR, JRA-55, MERRA-2, ERA-Interim and ERA-5 and use satellite products as references. The comparison includes the ERA5 reanalysis of ECMWF (*Hersbach et al.*, 2020), which is a new reanalysis based on a new generation of the ECMWF Integrated Forecasting System (IFS) model (cycle CY41R2) with a gap of more than 10 years with respect to the previous ERA-Interim reanalysis.

Observational data sets

For comparison the reanalyses are compared with different satellite products which contain strong assumptions.

We use the 2B-FLXHR-LIDAR radiative heating product version 4 (FLXHR) that combines cloud data from the A-train satellite instruments CLOUDSAT, CALIPSO and MODIS to calculate radiative heating (*L'Ecuyer et al.*, 2015; *Henderson et al.*, 2013)[1]. This product depends on a number of other products, retrieval algorithms and assumptions, and temperature profiles from the ECMWF AUX product; it is therefore liable to biases and errors. Besides this, the A-train satellites are helio-synchronous and therefore the daytime and nighttime observations occur at fixed hours (close to 1:30 and 13:30 in local time) and do not sample properly the daily cycle of convection, especially over land where convection has its maximum in late afternoon. Nevertheless, FLXHR is based on comprehensive observations rather than modelled properties of clouds and represents a state-of-the-art estimate of the radiative effect.

The 2B-CWC-RVOD retrieves ice water path from the CLOUDSAT radar reflectivity and the visible optical depth from MODIS. The 2C-ICE product retrieves the ice water path from the radar reflectivity and the backscatter coefficient of the CALIOP lidar. They both use Rodgers optimal estimation in the retrieval. Total condensates are available from the 2B-CWC-RVOD product version 4 (*Austin et al.*, 2009), which is used in FLXHR version 4, and the ice profile from the 2C-ICE product version 4 (*Deng et al.*, 2013, 2015).

Clouds and radiative effects

The cloud properties differ quite significantly among the reanalyses and the radiative properties vary accordingly. **Figure 8.64** shows that the maximum cloud cover in the ERA5 is smaller than in the ERA-Interim in the monsoon region, especially in the maritime regions that surround Asia. The altitude of the maximum cloud cover, not shown, is also lower by about 3 K on the average in potential temperature. As a result of these changes in the high clouds, the cloud radiative effect is also strongly modified. **Figure 8.65** illustrates the cloud radiative properties in the SASM longitude range (73 - 97 ° E) for the five reanalyses investigated in this Section. The two ECMWF reanalysis differ by the fact that the ERA5 cloud is smaller and located at a lower altitude than in the ERA-Interim, especially over the oceans (see cloud cover Section below). Therefore, the maximum of the cloud radiative effect is shifted downward by about 2 km, and the mean zero level of net radiative heating is left rather unperturbed by the clouds except over 20 - 40 ° N where continental convection dominates and it will be seen below that this is mostly an effect of the Tibetan plateau. Two other reanalyses, CFSR and JRA-55, display cloud radiative heating patterns that are in fairly good agreement with ERA5 but with much reduced amplitude for JRA-55. MERRA-2 exhibits a very different pattern from other reanalysis with a strong radiative heating in the 0 - 20 ° N latitude range and from 6 km to 12 km and a strong cooling above from 12 km to 16 km. As a consequence, an island of positive all sky radiative heating is observed between 8 km and 10 km and the zero level of radiative heating is shifted upward by one kilometer by the clouds between 10 ° S and 30 ° N.

[1] T2B-FLXHR-LIDAR and all the CLOUDAT/CALIPSO products mentioned in this study (2B-GEOPROF, 2B-CWC-RVOD, 2C-ICE) are available at **http://www.cloudsat.cira.colostate.edu/data-products**

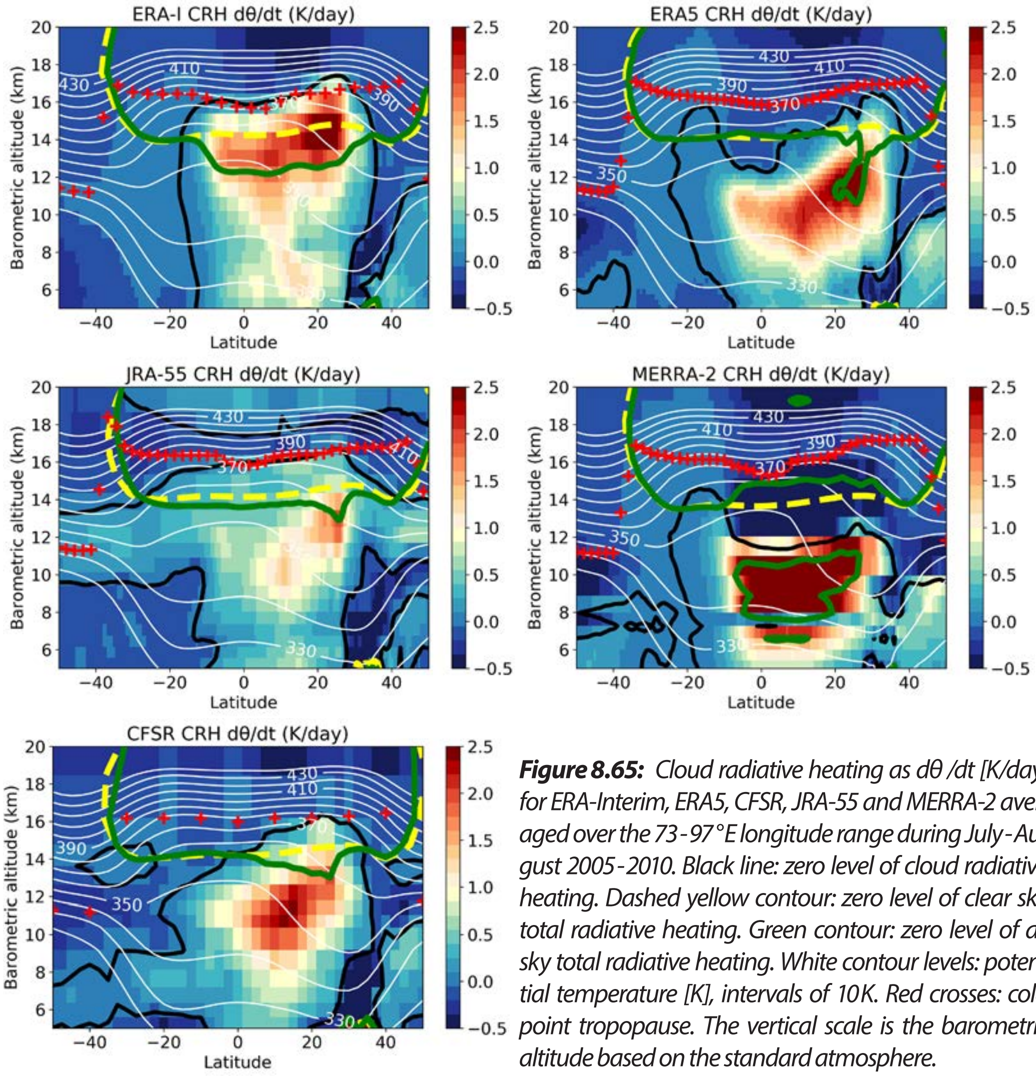


Figure 8.65: *Cloud radiative heating as dθ/dt [K/day] for ERA-Interim, ERA5, CFSR, JRA-55 and MERRA-2 averaged over the 73-97°E longitude range during July-August 2005-2010. Black line: zero level of cloud radiative heating. Dashed yellow contour: zero level of clear sky total radiative heating. Green contour: zero level of all sky total radiative heating. White contour levels: potential temperature [K], intervals of 10K. Red crosses: cold point tropopause. The vertical scale is the barometric altitude based on the standard atmosphere.*

This is contrary to the ERA-Interim case where it is shifted downward by about 2 kilometers over the same latitude range. In the following, we investigate more details about these discrepancies and their causes.

In order to separate land from ocean and, among land, the high orography of the Tibetan plateau from the rest of Asia, we divide the SASM domain into a set of regions as indicated in **Figure 8.66**. We focus on six regions that encompass most of the convective activity during SASM and its variability: Bay of Bengal (BoB), Indian Subcontinent (Indian Sub), South China, Sea of China and Philippine Sea (SCSPhi), Indochinese Peninsula (Pen) and the Tibetan-Plateau.

From the heating archive of the five reanalyses, the cloud heating has been obtained by removing the clear-sky contribution from the all-sky value. As the clear-sky heating rates are not available for JRA-55 and CFSR, we use ERA5 as a reference. It has been checked with ERA5, ERA-Interim and MERRA-2 that the discrepancies among clear sky radiative heating rates are at least one order of magnitude smaller than the all sky discrepancies, except near the ground over land where differences in albedo induce also differences in shortwave heating. The total shortwave heating rate is calculated using the clear sky sun variation as integrator.

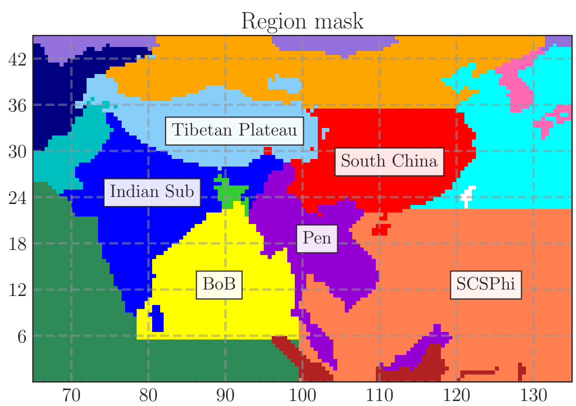


Figure 8.66: *Longitude-latitude distribution of considered SASM regions including two maritime regions: Bay of Bengal (BoB) and Sea of China and Philippine Sea (SCSPhi); and four continental regions: Indian Subcontinent (Indian Sub), South China, Indochina Peninsula (Pen) and the Tibetan Plateau. The Tibetan Plateau is defined as the region of altitude higher than 3800m. Other regions seen in this map are not used in this study.*

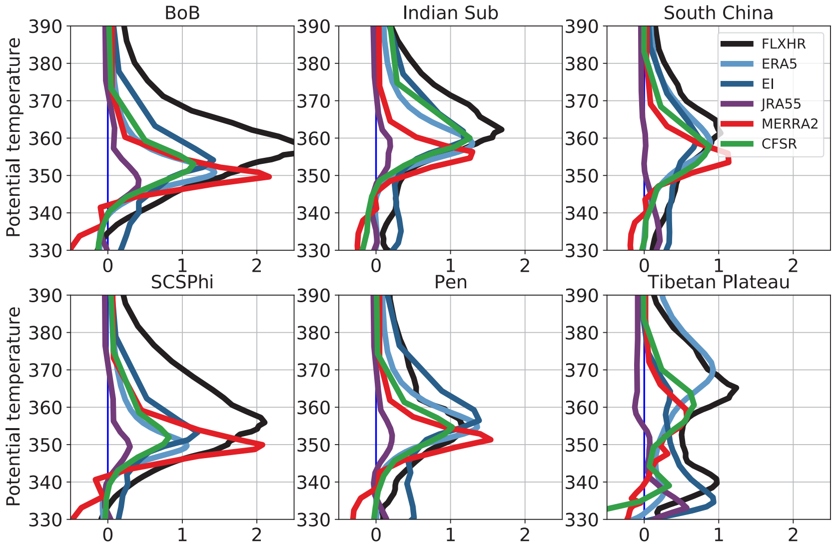


Figure 8.67: *Cloud short-wave radiative heating as dθ/dt [K/day] for ERA5, ERA-Interim, JRA-55, CFSR and MERRA-2 as potential temperature tendencies for the six regions shown in* ***Fig. 8.66****: BoB, Indian Sub, South China, SCSPhi, Pen and Tibetan Plateau as a function of altitude in potential temperature. The average is performed over July - August of 2007 - 2010. Black curve shows the FLXHR satellite profile, and other colors the reanalyses as indicated in legend.*

Tibetan Plateau is not well sampled by the A-train satellites. It is noticeable that, except over this region, the CFSR and ERA5 curves are very close, closer than ERA5 and ERA-Interim.

Figure 8.68 shows the cloud longwave heating. In the maritime regions, CFSR and ERA5 are still very close and follow closely the FLXHR curve with small cooling above 350 K and warming below. JRA-55 and ERA-Interim form another group with warming all the way down from 370 K. MERRA-2 exhibits a very strong cooling-warming pattern, typical of the effect of fat convective anvils, with a crossover at 350 K. Over Indian Sub and South China, the agreement persists between CFSR and ERA5 on one side and between ERA-Interim and JRA-55 on the other side but now FLXHR agrees better with the second pair above 370 K where it produces heating instead of cooling. MERRA-2 displays the same pattern than over the ocean but attenuated. As for shortwave heating, Indochinese Peninsula (Pen) shows intermediate patterns.

Over the Tibetan Plateau, CFSR agrees with ERA-Interim but not with ERA5. No reanalysis agrees well with FLXHR and the MERRA-2 curve shows multiple crossings with the zero line. The fact that reanalyses disagree even on the sign of the cloud longwave radiative effect above 350 K is not totally surprising as the antagonist warming effect of cirrus and cooling effect of the underlying thick anvils largely balance in this region (*Johansson et al.*, 2015).

In order to explore the origin of such discrepancies, **Figure 8.69** shows the cloud cover profiles for the reanalyses and for the 2B-GEOPROF-LIDAR product used in FLXHR

The comparison is based on July and August months of the 2007-2010 period, for which the FLXHR product is available for both night-time and day-time orbits.

Figure 8.67 shows that JRA-55 differs from the other reanalyses in producing very small cloud shortwave heating. Over maritime regions (BoB and SCSPhi), the reanalyses have maximum radiative heating near 350 K with fast decay above, while FLXHR displays a maximum higher up at ~ 358 K and larger values than all reanalyses in the 360 - 390 K range. MERRA-2 is the reanalysis with the lowest and narrowest maximum. The discrepancy from FLXHR is the largest over the maritime regions, where FLXHR samples convection near its mid-day maximum. Our calculation might therefore generate a positive bias with a large maximum at the altitude of maximum cloud cover (see also **Fig. 8.69**, below). The discrepancy is strongly reduced over China and India with respect to ECMWF reanalysis and CFSR while MERRA-2 retains its characters. For FLXHR and all reanalyses, the radiative heating maxima are shifted upward over land with respect to ocean. The Indochinese Peninsula (Pen) region presents intermediate patterns between land and ocean. Over the Tibetan Plateau, a diversity of patterns is obtained and the sole reanalysis that displays the neat double peak structure of FLXHR is ERA5, the lower peak being due to low level clouds (as 330 K is close to the surface in this region). Notice, however, that the typically late afternoon convection of the

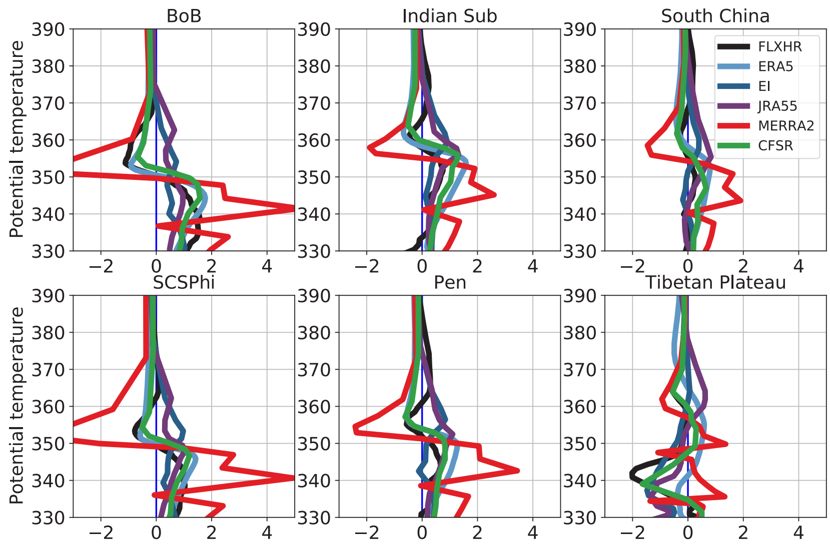


Figure 8.68: *Same as* ***Fig. 8.67*** *but for cloud long-wave heating [K/day].*

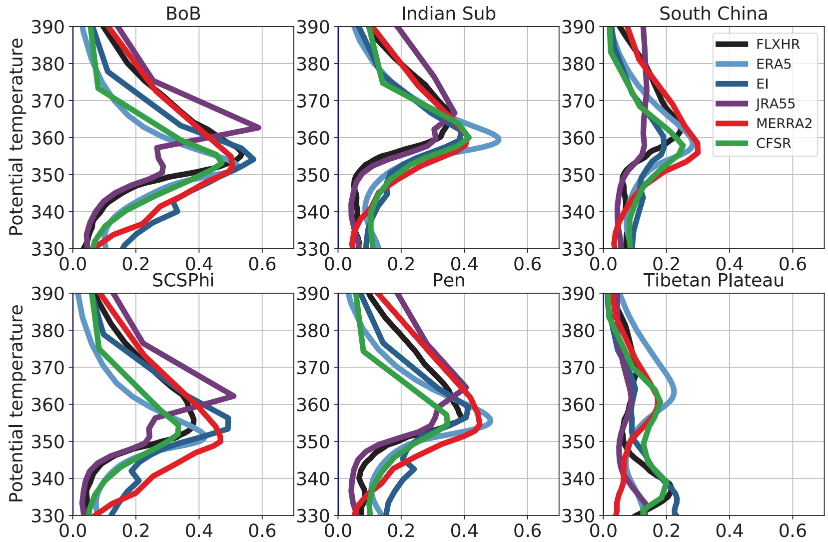


Figure 8.69: *Same as* ***Fig. 8.67*** *but for cloud cover [fraction].*

(*Mace and Zhang*, 2014). Such comparison must be considered carefully as the notion of cloud cover is not necessarily defined in the same way between observations and models. Nevertheless, we do not see here any pattern that would explain MERRA-2 differences with the other reanalyses. The reanalyses and 2B-GEOPROF-LIDAR cloud cover are in good agreement over South China and Indian Sub, but for the tendency of JRA-55 to maintain significant cloud cover at very high altitude. The dispersion is larger over the maritime regions and not surprisingly over the Tibetan Plateau where, however, all reanalyses except MERRA-2 show a double maximum structure with a layer of low clouds. In all cases, the higher cloud cover over continental regions is consistent with the cloud radiative profiles.

The water condensate profiles of the reanalyses and two A-train satellite products (2B-CWC-RVOD and 2C-ICE) are shown in **Figure 8.70**. An evaluation of 2C-ICE against other satellite products and ground observations can be found in *Deng et al.* (2013, and 2015). These curves display the non-precipitating component which is usually the one used for radiative calculations. Again, ERA5 and CFSR are very close over the maritime region but CFSR exceeds ERA5 by about 70 % above 340 K over the land. There is a very small amount of condensates in JRA-55 which drops rapidly to zero at high levels. Therefore, the large cloud cover found in **Figure 8.69** is of no consequence and this explains the overall low cloud radiative effect of JRA-55. On the contrary, MERRA-2 exhibits a large maximum in the condensates between 340 K and 350 K over the maritime region, due to thick anvils mainly consisting of ice, which is clearly correlated with the warming layer in the longwave heating and to the sharp maximum in the shortwave heating. The strong longwave cooling above is due to the small emission of this thick opaque ice layer. The same pattern is seen over land but weaker and at higher altitude, again in good correlation with the radiative heating. The level of zero crossing in the longwave heating is also located just at the top of the condensate layer in CFSR and ERA5. The smaller amount of condensates in the anvil layers of ERA-Interim and JRA-55 is a good candidate to explain why heating by cirrus clouds overwhelms the cooling effect of anvils above 350 K. It is quite certain that these water condensates profiles are the main explanation of the discrepancies visible in the radiative heating above the SASM region and that the competition between the signatures of the convective cloud anvils and the cirrus clouds is the key factor as already shown by *Johansson et al.* (2015).

The satellite products can hardly be compared to the non-precipitating water condensates in models since they measured both non-precipitating and precipitating condensates together; separating the two requires ad hoc filtering and corrections. As the ERA5 archived data includes also the rain and snow variables we also compare in **Figure 8.70** the total condensate profile of ERA5 with two A-train satellite products. The agreement between the three curves is best over the continental regions (outside the Tibetan Plateau). The whole profiles of ERA5 and 2B-CWC-RVOD are very

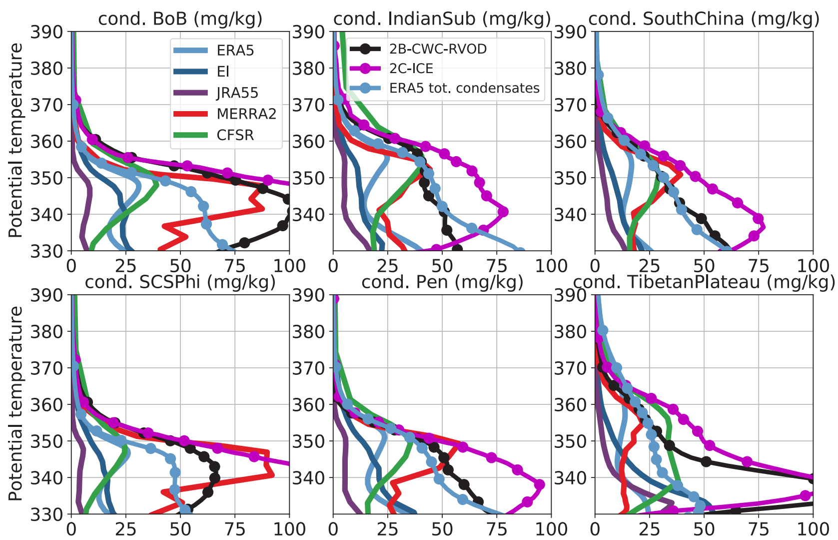


Figure 8.70: *Non-precipitating cloud condensates (ice and water; in [mg/kg]) for the four reanalyses; plain solid curves. Additional, total cloud condensates are shown for ERA5, including snow and rain (light blue dotted curve), and for the 2B-CWC-RVOD (black dotted curve) and 2C-ICE (magenta dotted curve) A-train satellite products; (all in [mg/kg]). Otherwise same period and regions as for* ***Fig. 8.67****.*

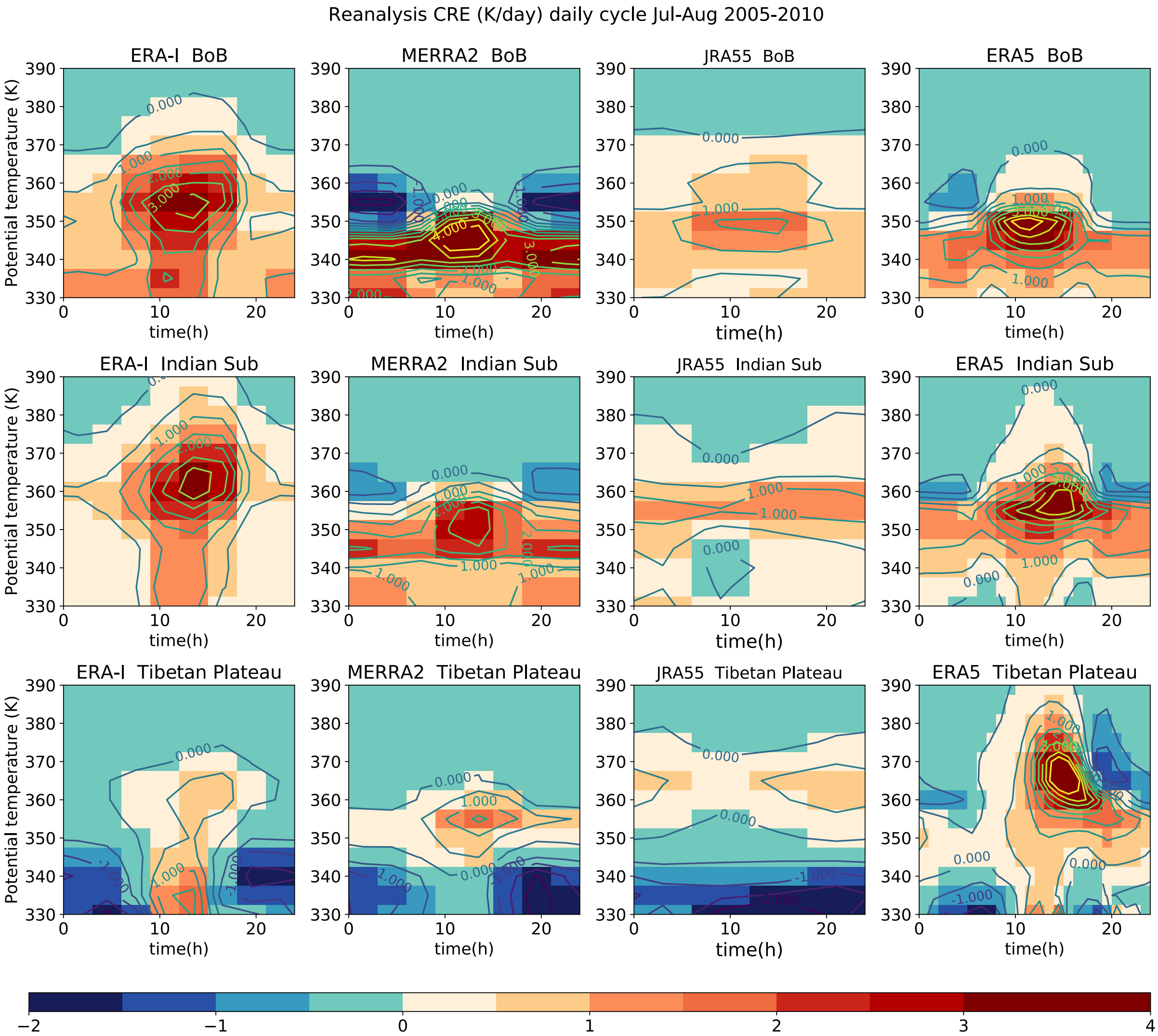


Figure 8.71: *Daily cycle of cloud radiative heating as d /dt [K/day] calculated over July - August 2005 - 2010 for ERA-Interim, MERRA-2, JRA-55 and ERA5 (columns) and for the three regions BoB, Indian Sub and Tibetan-Plateau (rows). The figure is based on hourly data for ERA5, 3-hourly data for MERRA-2 and ERA-Interim, and 6-hourly date for JRA-55.*

close over Indian Sub and South China while 2C-ICE is larger below the top of the convective cloud anvils. Over the maritime regions, ERA5 has less condensates than the satellites products and the separation increases below 350 K where 2C-ICE provides also much larger values than 2B-CWC-RVOD.

Over the Tibetan-Plateau, ERA5 displays a large deficit of condensates with respect to the satellite products around 340 K. Only the ERA-Interim and JRA-55 show a profile with a strong maximum in this region in agreement with the cloud cover but with a much too weak value. Even if the A-train satellite products are likely to contain some biases over the Tibetan-Plateau, the discrepancies with respect to analyses in this region might be for a large part due to a general under-representation of the low-level convection (*Li et al.*, 2016, 2017).

Finally, **Figure 8.71** shows the daily cycle of the cloud radiative heating for three regions – maritime (BoB), continental (Indian Sub) and the Tibetan-Plateau and four reanalyses (excluding CFSR). The contrast is strong between JRA-55 which shows a very weak daily cycle and MERRA-2 which shows a very strong cycle with intense nocturnal cooling between 345 K and 365 K (~11 km and 15 km) due to persistent thick convective cloud anvils. In the BoB region, the diurnal maximum of MERRA-2 is located at the lowest level at 345 K (~ 11 km) while the ERA-Interim is at the highest level at 355 K (~ 13.5 km). Over India, the diurnal radiative heating in MERRA-2 is attenuated and slightly shifted towards afternoon with respect to BoB. There is no attenuation respective to maritime value but a noticeable afternoon shift in ERA-Interim and ERA5. In addition, the vertical location of the maximum rises by about 10 K in both reanalyses but remains lower in ERA5. Over the Tibetan-Plateau, where other reanalyses show weak radiative heating ERA5 exhibits a strong maximum that reaches 380 K (~ 17 km), which is associated with high penetrative convection in this region, a distinguished signature of ERA5.

8.8.7 Key findings and recommendations

Key findings

- Modern reanalyses agree well regarding the climatological position and extent of the SASM anticyclone, although there are notable differences in the distribution of SASM anticyclone centre locations among different reanalyses. Distinct bimodality of the SASM anticyclone centre location based on daily data is only present in NCEP-NCAR R1. (*Section 8.8.1*)

- Reanalyses indicate slightly higher CPT temperatures and lower CPT heights in the SASM anticyclone compared to GNSS-RO satellite observations. (*Section 8.8.1*)

- Climatologies of SASM anticyclone moments (centroid location, aspect ratio, angle, excess kurtosis) computed using MSF on isentropic surfaces to define the SASM anticyclone edge show good agreement among the MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 reanalyses. Good qualitative agreement is seen in the evolution of SASM anticyclone area defined using MSF, but MERRA and MERRA-2 show larger areas and consequently longer monsoon seasons, along with more significant apparent increasing trends in SASM anticyclone area. (*Section 8.8.1*)

- Omega fields from ERA-Interim, JRA-55, MERRA-2, and CFSR reanalyses agree well on the overall spatial pattern within the SASM domain. However, regional discrepancies exist, especially over locations of frequent convection such as the western coast of India and the Bay of Bengal. (*Section 8.8.2*)

- Diabatic heating distributions within the SASM region differ significantly among reanalyses, especially with respect to the mean latitudinal location of the maximum heating rates connecting the convective detrainment layer to the lower stratosphere. This maximum is located progressively further north in ERA-Interim, JRA-55, and CFSR, and is missing entirely from MERRA-2. These differences can be attributed to differences in the dominant diabatic processes in the UT above the SASM region: cloud-induced radiative heating in ERA-Interim, convective heating in JRA-55 and CFSR, and cloud-induced radiative cooling in MERRA-2. (*Section 8.8.3*)

- The depth and location of convection within the SASM and surrounding regions varies among the reanalyses. These differences impact implied convective transport over the Bay of Bengal, the Himalayan South Slope, the south-eastern Tibetan Plateau, and the western North Pacific, as well as over East Asia and equatorial Africa. Differences in the distribution and magnitude of diabatic heating near the tropopause (380 K) are likewise strongly affected by differences in the distribution of convection and related clouds. (*Section 8.8.3*)

- Residence times based on diabatic Lagrangian transport calculations are shortest in the centre of the SASM anticyclone for all reanalyses. The absolute magnitudes of residence time in the SASM anticyclone show large differences, varying from a minimum of 14 days (ERA-Interim) to a maximum of 22 days (MERRA-2) during the JJA season. (*Section 8.8.4*)

- Despite differences in magnitude and in the locations of local extrema, distributions of ozone volume mixing ratios within the SASM anticyclone are qualitatively consistent among reanalyses and broadly consistent with observations. However, none of the evaluated reanalyses are able to fully reproduce the low ozone mixing ratios within the SASM 'ozone valley'. (*Section 8.8.5*)

- Cloud properties differ greatly among reanalyses as these properties are weakly constrained by assimilated observations. The radiative effect of clouds is used here as a metric in the SASM domain. In all reanalyses, maximum cloud cover is found between 350 K and 355 K over maritime regions and near 360 K over land. The maximum shortwave cloud heating rates essentially follow the maximum cloud cover. The longwave cloud heating rate combines the effect of thick anvils (warming below and cooling above) and the warming effect of overlying cirrus. The balance above clouds can be either positive or negative depending on the reanalysis. (*Section 8.8.6*)

- MERRA-2 displays a very strong anvil signature in contrast to all other reanalyses and satellite products. The discrepancies in the long-wave cloud heating rates are mostly explained by the ice content of high clouds in the reanalyses. Overall, the heating rates based on ERA5 are closest to the FLXHR satellite product. ERA5 is distinguished by lower cloud maxima on the average but stronger penetrative convection, in particular over the Tibetan Plateau. (Section 8.8.6)

Key recommendations

- For subsequent analyses involving the position of the SASM anticyclone centre it is recommended to use more recent reanalyses. In particular, researchers are encouraged to avoid NCEP-NCAR R1 and NCEP-DOE R2 when possible, proceed with caution if it is necessary to use one of these two reanalyses, and assess the sensitivity of results to the choice of reanalysis regardless of which reanalysis is used. (*Section 8.8.1*)

- The geopotential height field from the MERRA-2-ANA pressure-level data set features spurious enhancements over the steep orography of the Himalaya Mountains due to a conversion error. For analyses that are sensitive to this issue, any use of MERRA-2-ANA data should rely on the model-level products only. MERRA-2-ASM products are unaffected by this issue. (*Section 8.8.1*)

- Transport simulations for the SASM domain that use diabatic heating rates to represent vertical motion should use multiple reanalyses if possible and carefully consider the representation of convective sources to the TTL. MERRA-2 diabatic heating rates should only be used at 370 K potential temperature level and above. (*Sections 8.8.3* and *8.8.4*)

- Reanalyses capture the existence of the ozone minimum in the UTLS above the SASM but do not reliably reproduce its observed distribution or magnitude. Use of reanalysis ozone products in this region is appropriate for evaluation of internal reanalysis behaviour. Other applications should keep in mind the relatively simple formulation of the ozone models (see *Chapters 2* and *4*) and include careful validation against observations. (*Section 8.8.5*)

- Cloud and radiative heating for SASM regions should be used with caution for all reanalyses. However, ERA5 is most consistent with the satellite-based FLXHR product. (*Section 8.8.6*)

8.9 Summary, key findings, and recommendations

Chapter 8 investigates the extent to which reanalysis data sets are able to reproduce key characteristics of the TTL. Representations of the cold point and lapse rate tropopause are evaluated based on comparison of tropopause zonal mean profiles and time series to observational records from radio occultation and radiosonde data. The vertical structure of the TTL is then assessed by comparing reanalysis temperature profiles at model-level resolution to high-resolution GNSS-RO temperature profiles.

Basic dynamical processes and circulation patterns are evaluated by comparing zonal-mean and tropical-mean distributions of diabatic heating, as well as by means of Lagrangian trajectory simulations of transport within the TTL. Final dehydration locations and temperatures as well as TTL residence times derived from these trajectory simulations are compared among the reanalyses and validated against vertical velocity estimates derived from satellite observations of water vapor. Large-scale wave forcing is analysed based on the characteristic horseshoe-shaped structure that results from the superposition of Rossby and Kelvin responses to intense convective heating. Comparison against NOAA outgoing longwave radiation allows an assessment of spatiotemporal variability in this wave response. A zonal wavenumber-frequency spectral analysis is also carried out to describe and evaluate equatorial wave activity in the reanalyses. Long-term changes in the width of the tropical belt are derived based on the tropical jet and tropopause break positions, two metrics which are known to correlate only weakly with Hadley cell extent. Tropical-width metrics calculated based on instantaneous longitudinally-resolved and zonal-mean annual-mean fields are compared with each other. Owing to the lack of observational data for validating the tropical width diagnostics, the extent to which changes may be considered robust is determined based on statistical methods and consistency among the reanalyses. Finally, analysis of the upper troposphere and lower stratosphere above the South Asian Summer Monsoon (SASM) highlights some key differences in reanalysis performance within the TTL via focus on regional and seasonal aspects of the SASM anticyclone.

Key findings

- Advances in reanalysis and observational systems over recent years have led to a clear improvement in TTL reanalysis products over time. In particular, the reanalyses ERA-Interim, ERA5, MERRA2, CFSR, and JRA-55 show very good agreement after 2002 in terms of the vertical TTL temperature profile, meridional tropopause structure, and inter-annual variability. Long-term temperature trends from reanalyses and adjusted radiosonde data indicate significant cooling in the upper TTL during 1979 - 2005 (above the cold point). (*Section 8.2*)

- While climatological TTL temperatures from reanalyses agree very well with observations with relatively small low biases, the cold point and lapse rate tropopause show warm biases, most likely related to the fact that the discrete values corresponding to reanalysis model levels are unable to reproduce the observed minimum temperature as recorded in a near-continuous profile. (*Section 8.2.2*)

- Cloud fields in the tropical UTLS vary greatly in both magnitude and vertical distribution across reanalyses. Differences in cloud fraction and cloud water content impact the radiation budget both at the top-of-atmosphere and within the UTLS, and the effects of differences in cloud and convection parameterizations can be identified in vertical profiles of temperature and humidity in the tropical troposphere. (*Section 8.3*)

- There are large differences among reanalysis diabatic heating products within the TTL, which are known to influence transport statistics and rates of ascent in trajectory simulations of cross-tropopause transport in this region. Differences among reanalysis diabatic heating rates in the tropical UTLS are not limited to any one component: longwave, shortwave, and non-radiative components all show substantial discrepancies. (*Section 8.4*)

- Lagrangian transport studies demonstrate large differences in reanalysis temperatures at the dehydration point and in TTL residence times. However, the data sets agree on the spatial distribution of dehydration locations and produce roughly similar distributions, seasonal cycles, and interannual variations of TTL residence time. (*Section 8.5*)

- Equatorial wave activity and corresponding temperature anomaly patterns at 100 hPa are similar among the reanalyses, including the characteristic horseshoe-shaped structures that resemble the stationary wave response to tropical heating. However, the strength of the wave activities, their spectral magnitudes, and the intensity of temperature response differ among the reanalyses, with the latter differences depending on the aspects of the dynamical model and/or assimilation system. (*Section 8.6*)

- Metrics of the width of the TTL based on the zonally-resolved subtropical jet and tropopause break show robust changes in only a few regions and seasons and poor agreement of the resulting zonal-mean annual-mean values. The diagnostics based on the zonal-mean subtropical jet and tropopause break, on the other hand, suggest stronger trends in the width of the TTL than their zonally-resolved counterparts. Overall, the two subtropical jet diagnostics are more consistent than the two tropopause break diagnostics, possibly related to smoother variations in the zonal wind field relative to the tropopause break. (*Section 8.7*)

- Modern reanalyses agree well regarding the climatological position and evolution of area extent and moments of the SASM anticyclone, although there are notable differences in the distribution of SASM anticyclone centre locations. All of the reanalyses indicate slightly higher CPT temperatures and lower CPT heights in the SASM anticyclone compared to GNSS-RO satellite observations. (*Section 8.8.1*)

- Distributions of ozone volume mixing ratios within the SASM anticyclone are qualitatively consistent among reanalyses and broadly consistent with observations. However, none of the evaluated reanalyses are able to reproduce the low ozone mixing ratios within the SASM anticyclone. (*Section 8.8.5*)

- Cloud properties, convection, radiative heating, and omega fields for the SASM UTLS differ significantly among reanalyses on a regional scale as these properties are only weakly constrained by assimilated observations. These differences impact derived transport processes in the UTLS, and residence times based on diabatic Lagrangian transport calculations reveal large differences. (*Sections 8.8.2, 8.8.3, 8.8.4, 8.8.6*)

Key recommendations

- In the TTL, temperature on native model levels should be used rather than the standard pressure-surface data sets. Various diagnostics such as the cold point and lapse rate tropopause and the analysis of equatorial waves are demonstrably improved when model-level data are used. For a more realistic representation of the tropical tropopause levels, data sets that combine low temperature biases with high vertical resolution should be used. (*Sections 8.2* and *8.6*)

- Long-term drifts in high cloud fraction, OLR, and LWCRE are present in almost all reanalyses, and often disagree in terms of sign, timing, or magnitude. These products should generally not be used for trend or time series analysis without independent verification. Among the reanalyses, ERA5 shows greater stability in time and stronger correlations with observed variability for these cloud and radiation metrics and may therefore offer a more reliable characterization of long-term variations in related metrics relative to earlier reanalyses. (*Section 8.3*)

- Given large differences in reanalysis diabatic heating products and related metrics within the tropical UTLS, researchers using these fields to drive or nudge model simulations of this region should use multiple reanalyses whenever possible. (*Sections 8.4* and *8.5*)

- When applying metrics of tropical width based on the subtropical jet or tropopause break, it is recommended to use multiple reanalyses and to be aware of the caveat that the zonal-mean diagnostics suggest stronger trends than their zonally-resolved counterparts. (*Section 8.7*)

- For analyses involving the SASM anticyclone it is recommended to use more recent reanalyses. In particular, researchers are encouraged to avoid NCEP-NCAR R1 and NCEP-DOE R2 data sets and the geopotential height field of the MERRA-2-ANA pressure-level data when possible. (*Section 8.8.1*)

- Transport simulations for the SASM domain that use diabatic heating rates to represent vertical motion should use multiple reanalyses if possible and carefully consider the representation of convective sources to the TTL. MERRA-2 diabatic heating rates should only be used at 370 K potential temperature level and above. (*Sections 8.8.3, 8.8.4*)

- Ozone in the UTLS above the SASM should be carefully validated against observations, and cloud and radiative heating should be used with caution for all reanalyses. (*Sections 8.8.5, 8.8.6*)

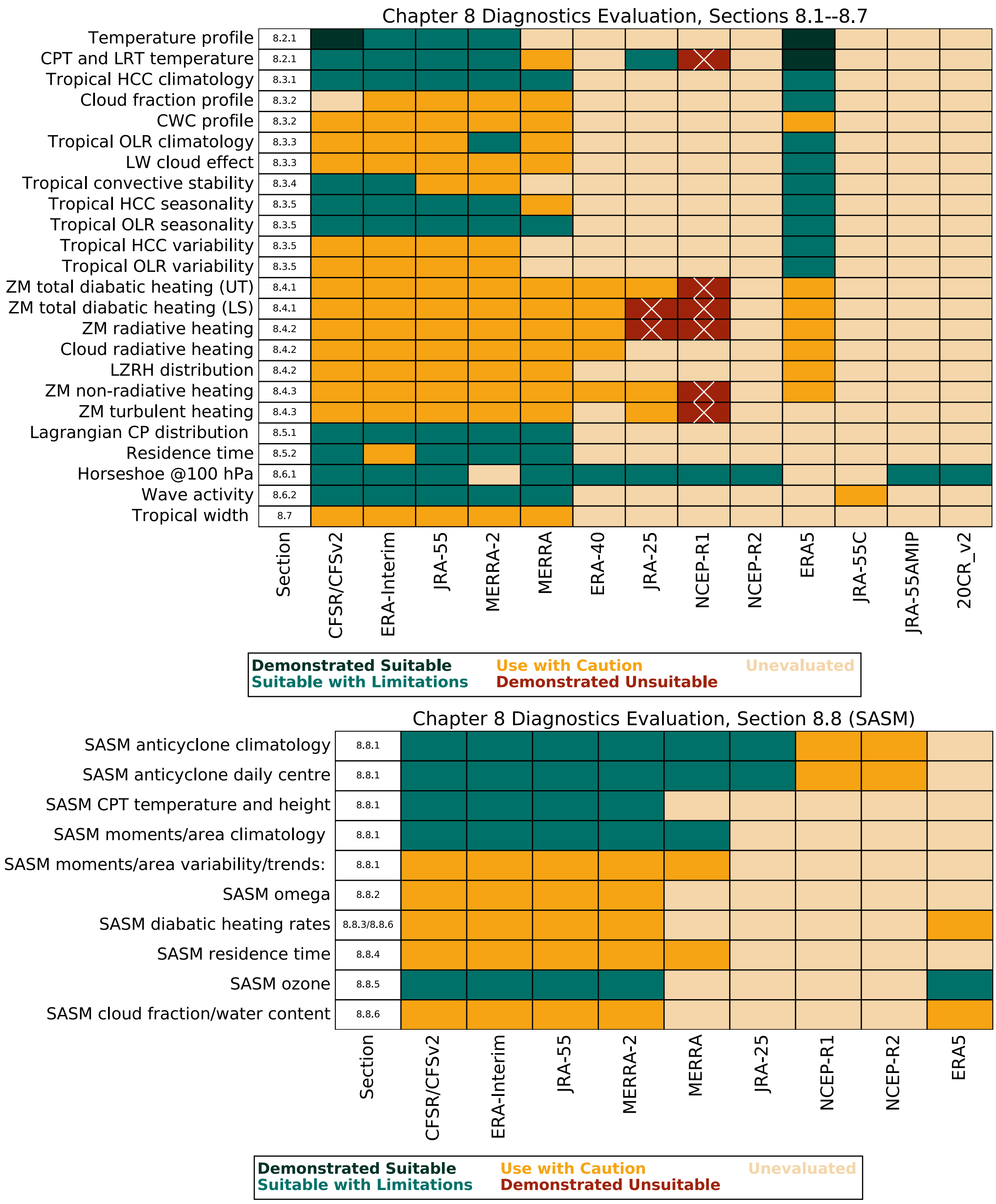


Figure 8.72: *Diagnostic evaluation for Chapter 8 on the Tropical Tropopause Layer (TTL). Top: Climatological and dynamical TTL characteristics (Sections 8.2-8.7) and bottom: South Asian Summer Monsoon (SASM) characteristics (Section 8.8). CPT: Cold Point Tropopause, LRT: Lapse Rate Tropopause, HCC: High Cloud Cover, CWC: Cloud Water Content, OLR: Outgoing Longwave Radiation, LW: Long Wave, ZM: Zonal Mean, UT: Upper Troposphere, LS: Lower Stratosphere, LZRH: Level of Zero Radiative Heating, CP: Cold Point.*

Author contributions

ST and KK wrote Chapter 8 with contributions from all coauthors. The data, figures and text for the individual sections were compiled by the following coauthors listed by sections.

Section 8.1: ST, KK

Section 8.2: ST, SD, II, BL, RPK, JSWang, TW

Section 8.3: JSWright (JSW)

Section 8.4: JSW

Section 8.5: 8.5.1: TW; 8.5.2: TW; 8.5.3: JSW, TB

Section 8.6: 8.6.1: EN, MF; 8.6.2: YHK, MF

Section 8.7: NAD, SD; 8.7.1: GLM; 8.7.2: CRH, 8.7.3: TB, NAD, GLM, CRH

Section 8.8: 8.8.1 MN, GLM, ST, KK; 8.8.2: CRH, KK; 8.8.3: JSW, KK; 8.8.4: TW, KK; 8.8.5: JSW; 8.8.6: BL

Section 8.9: All authors

Data availability

(8.8.1) For some of the analyses, NCEP-DOE R2, CFSR, JRA-25 and JRA-55 data were provided by the Research Data Archive (RDA) of the Computational and Information Systems Laboratory at the National Center for Atmospheric Research via **https://rda.ucar.edu/**. NCAR is supported by grants from the National Science Foundation. NCEP-NCAR R1, ERA-Interim and MERRA/MERRA-2 data were obtained from NOAA/OAR/ESRL PSD at **https://www.esrl.noaa.gov/psd/**, from ECMWF at **https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=pl/** and from GES DISC at **https://disc.gsfc.nasa.gov/** (partly via **mirador.gsfc.nasa.gov**), respectively. ERA5 diagnostics have been produced using data extracted from Copernicus Climate Change Service. A-train data have been obtained and processed at ICARE/AERIS Data Centre.

(8.8.6) The satellite data are provided by the AERIS/ICARE data center **http://www.icare.univ-lille1.fr**. ERA5 data were generated using Copernicus Climate Change Service Information.

Acknowledgements

We acknowledge the institutions involved in the production and dissemination of the reanalysis data NCAR is supported by grants from the National Science Foundation.

(8.3, 8.4, 8.8.3, 8.8.5) Contributions by JSW to these Sections were supported by funding from the National Natural Science Foundation of China (grant number 41761134097).

(8.3) We thank Xianglei Huang for advice on interpreting satellite observations of outgoing LW radiation, Peter Bechtold, Andrea Molod, and Guang J. Zhang for clarifying aspects of the cloud, convection, and turbulence parameterizations in the reanalysis forecast models, Xiaoyi Sun for conducting the initial data processing for high cloud fields and OLR, and Yi Huang for recommendations on methodology and presentation.

(8.4) We thank Stephan Fueglistaler for providing the turbulent mixing temperature tendency for ERA-Interim, SHADOZ data, ERA-Interim ozone climatology, and many helpful suggestions.

(8.8.1) The research leading to these results has received funding from the European Communities Seventh Framework Programme (FP7/2007 - 2013) under grant agreement No. 603557.

(8.8.5) We thank Xiaolu Yan for providing context on the Asian monsoon ozone valley and its origins.

(8.8.6) BL acknowledges support from the TTL-Xing ANR-17-CE01-0015 project of Agence Nationale de la Recherche.

Figures 8.2, 8.3, 8,4, 8.5, 8.6, 8.7, 8.8, and **8.9** are adapted from *Tegtmeier et al.* (2020). **Figures 8.10, 8.11, 8.12, 8.13, 8.14, 8.15, 8.17, 8.24**, and **8.25** are adapted or reproduced from *Wright et al.* (2020). **Figures 8.20, 8.21**, and **8.26** are updated from *Wright and Fueglistaler* (2013). **Fig. 8.38** is modified after *Kim et al.* (2019). **Figure 8.47** is modified from *Nützel et al.* (2016).

References

Abalos, M., B. Legras, F. Ploeger, and W.J. Randel, 2015: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979–2012. *J. Geophys. Res.*, **120**, 7534 - 7554, ooi: 10.1002/2015JD023182.

Adam, O., *et al.*, 2018: The TropD software package (v1): standardized methods for calculating tropical-width diagnostics. *Geosci. Model Dev.*, **11**, 4339 - 4357, doi: 10.5194/gmd-11-4339-2018.

Anthes, R.A., *et al.*, 2008: The COSMIC/FORMOSAT-3 Mission: Early Results. *B. Am. Meteor. Soc.*, **89**, 313 - 333, doi: 10.1175/BAMS-89-3-313,.

Austin, R.T., A.J. Heymsfield, and G.L. Stephens, 2009: Retrieval of ice cloud microphysical parameters using the CloudSat millimeter-wave radar and temperature. *J. Geophys. Res.*, **114**, doi: 10.1029/2008JD010049.

Bechtold, P., *et al.*, 2008: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Q. J. R. Meteor. Soc.*, **134**, 1337 - 1351, doi: 10.1002/qj.289.

Bechtold, P., *et al.*, 2014: Representing equilibrium and nonequilibrium convection in large-scale models. *J. Atmos. Sci.*, **71**, 734 - 753, doi: 10.1175/JAS-D-13-0163.1.

Bergman, J.W., *et al.*, 2013: Boundary layer sources for the Asian anticyclone: Regional contributions to a vertical conduit. *J. Geophys. Res. Atmos.*, **118**, 2560 - 2575, doi: 10.1002/jgrd.50142.

Bian, J., *et al.*, 2011: Formation of the summertime ozone valley over the Tibetan Plateau: The Asian summer monsoon and air column variations. *Adv. Atmos. Sci.,* **28**, 1318, doi: 10.1007/s00376-011-0174-9.

Beyerle, G., *et al.*, 2005: GPS radio occultation with GRACE: Atmospheric profiling utilizing the zero difference technique. *Geophys. Res. Lett.*, **32**, L13806, doi: 10.1029/2005GL023109.

Beyerle, G., *et al.*, 2001: First results from the GPS atmosphere sounding experiment TOR aboard the TerraSAR-X satellite. *Atmos. Chem. Phys.*, **11**, 6687 - 6699, doi: 10.5194/acp-11-6687-2011.

Birner, T., S.M. Davis, and D.J. Seidel, 2014: The changing width of Earth's tropical belt. *Phys. Today*, **67**, 38 - 44, doi: 10.1063/PT.3.2620.

Bloom, S.C., L.L. Takacs, A.M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256 - 1271, doi: 10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.

Bodas-Salcedo, A., *et al.*, 2015: COSP: Satellite simulation software for model assessment. *Bull. Am. Meteor. Soc.*, **92**, 1023 - 1043, doi: 10.1175/2011BAMS2856.1.

Bonazzola, M., and P.H. Haynes, 2004: A trajectory-based study of the tropical tropopause region. *J. Geophys. Res.*, **109**, D20112, doi: 10.1029/2003JD004356.

Bowman, K.P., 1993: Large-scale isentropic mixing properties of the Antarctic polar vortex from analysed winds. *J. Geophys. Res.*, **98**, 23013 - 23027, doi: 10.1029/93JD02599.

Bowman, K.P., *et al.*, 2013: Input data requirements Lagrangian Trajectory Models. *Bull. Am. Meteor. Soc.*, **94**, 1051 - 1058, doi: 10.1175/BAMS-D-12-00076.1.

Brewer, A.W., 1949: Evidence for a world circulation provided by the measurements of helium and water vapor distribution in the stratosphere. *Q. J. R. Meteorol. Soc.,* **75**, 351 - 363, doi: 10.1002/qj.49707532603.

Chepfer, H., *et al.*, 2010: The GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP). *J. Geophys. Res. Atmos.*, **115**, D00H16, doi: 10.1029/2009JD012251.

Chipperfield, M.P., 1999: Multiannual simulations with a three-dimensional chemical transport model. *J. Geophys. Res.*, **104**, 1781 - 1805, doi: 10.1029/98JD02597.

Corti, T., *et al.*, 2005: Mean radiative energy balance and vertical mass fluxes in the equatorial upper troposphere and lower stratosphere. *Geophys. Res. Lett.*, **32**, L06802, doi:10.1029/2004GL021889.

Davis, N. and T. Birner, 2017: On the Discrepancies in Tropical Belt Expansion between Reanalyses and Climate Models and among Tropical Belt Width Metrics. *J. Climate*, **30**, 1211 - 1231, doi: 10.1175/JCLI-D-16-0371.1.

Davis, S.M., and K.H. Rosenlof, 2012: A multidiagnostic intercomparison of tropical-width time series using reanalyses and satellite observations. *J. Climate*, **25**, 1061 - 1078, doi: 10.1175/JCLI-D-11-00127.1.

Davis, S.M., *et al.*, 2016: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: A long-term database for climate studies. *Earth Syst. Sci. Data*, **8**, 461 - 490, doi:10.5194/essd-8-461-2016.

Deng, M., G.G. Mace, Z. Wang, and R.P. Lawson, 2013: Evaluation of Several A-Train Ice Cloud Retrieval Products with In Situ Measurements Collected during the SPARTICUS Campaign. *J. App. Meteor. Clim.*, **52**, 1014 - 1030, doi: 10.1175/JAMC-D-12-054.1.

Deng, M., G.G. Mace, Z. Wang, E. and Berry, 2015: CloudSat 2C-ICE product update with a new Z_e parameterization in lidar-only region: CLOUDSAT 2C-ICE PRODUCT UPDATE. *J. Geophys. Res. Atmos.*, **120**, 12,198 - 12,208, doi: 10.1002/2015JD023600.

Dessler, A.E. and P. Yang, 2003: The Distribution of Tropical Thin Cirrus Clouds Inferred from Terra MODIS Data. *J. Clim.*, **16**, 1241 - 1247, doi: 10.1175/1520-0442(2003)16<1241:TDOTTC>2.0.CO;2.

Dethof, A., A. Oneill, J.M. Slingo, and H.G.J. Smit, 1999: A mechanism for moistening the lower stratosphere involving the Asian summer monsoon. *Q. J. Roy. Meteor. Soc.*, **125,** 1079 - 1106, doi: 10.1002/qj.1999.49712555602.

Doelling, D., 2017: CERES SYN1DEG-DAYTerra+Aqua - Edition 4A, available at: **https://doi.org/10.5067/terra+aqua/ceres/syn1degday_l3.004a**, accessed: 2018-06-17.

Doelling, D.R., 2019: CERES Energy Balanced and Filled (EBAF) TOA and Surface Monthly means data in netCDF Edition 4.1. *NASA Atmospheric Science Data Center (ASDC)*, Hampton, VA, USA, available at: **https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF_L3B.004.1**.

Durre, I., R.S. Vose, and D.B. Wuertz, 2006: Overview of the Integrated Global Radiosonde Archive. *J. Climate,* **19**, 53 - 68, doi: 10.1175/JCLI3594.1.

Emanuel, K., *et al.*, 2013: Influence of Tropical Tropopause Layer Cooling on Atlantic Hurricane Activity. *J. Climate*, **26**, 2288 - 2301, doi: 10.1175/JCLI-D-12-00242.1.

Flannaghan, T.J. and S. Fueglistaler, 2011: Kelvin waves and shear-flow turbulent mixing in the TTL in (re-)analysis data. *Geophys. Res. Lett.*, **38**, L02801, doi: 10.1029/2010GL045524.

Folkins, I., *et al.*, 1999: A barrier to vertical mixing at 14 km in the tropics (1999). Evidence from ozonesondes and aircraft measurements. *J. Geophys. Res. Atmos.*, **104**, 22095 - 22102, doi: 10.1029/1999JD900404.

Fortuin, J.P. and U. Langematz, 1994: An update on the current ozone climatology and on concurrent ozone and temperature trends. *Proc. SPIE*, **2311**, 207 - 216, doi:10.1117/12.198578.

Free, M., *et al.*, 2005: Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): A new data set of large-area anomaly time series. *J. Geophys. Res.*, **110**, D22101, doi: 10.1029/2005JD006169, 2005.

Fu, R., W.T. Liu, and R.E. Dickinson, 1996: Response of Tropical Clouds to the Interannual Variation of Sea Surface Temperature. *J. Clim.*, **9**, 616 - 634, doi: 10.1175/1520-0442(1996)009<0616:ROTCTT>2.0.CO;2.

Fueglistaler, S., H. Wernli, and T. Peter, 2004:Tropical troposphere-to-stratosphere transport inferred from trajectory calculations. *J. Geophys. Res.*, **109**, D03108, doi: 10.1029/2003JD004069.

Fueglistaler, S. and Q. Fu, 2006: Impact of clouds on radiative heating rates in the tropical lower stratosphere. *J. Geophys. Res.*, **111**, D23202, doi: 10.1029/2006JD007273.

Fueglistaler, S., *et al.*, 2009a: Tropical tropopause layer. *Rev. Geophys.*, **47**, RG1004, doi: 10.1029/2008RG000267.

S. Fueglistaler, *et al.*, 2009b:The diabatic heat budget of the upper troposphere and lower/mid stratosphere in ECMWF reanalyses. *Q. J. R. Meteorol. Soc.* **135**, 21 - 37, doi: 10.1002/qj.361.

Fueglistaler, S., P.H. Haynes, and P.M. Forster, 2011: The annual cycle in lower stratospheric temperatures revisited. *Atmos. Chem. Phys.*, **11**, 3701 - 3711, doi: 10.5194/acp-11-3701-2011.

Fujiwara, M., *et al.*, 2012: Wave activity in the tropical tropopause layer in seven reanalysis and four chemistry climate model data sets. *J. Geophys. Res.*, **117**, D12105, doi: 10.1029/2011JD016808.

Fujiwara, M., *et al.*, 2015: Global temperature response to the major volcanic eruptions in multiple reanalysis data sets. *Atmos. Chem. Phys.*, **15**, 13507 - 13518, doi: 10.5194/acp-15-13507-2015.

Fujiwara, M., *et al.*, 2017: Introduction to the SPARC Reanalysis IntercomparisonProject (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp-17-1417-2017.

Garny, H. and W.J. Randel, 2013: Dynamic variability of the Asian monsoon anticyclone observed in potential vorticity and correlations with tracer distributions. *J. Geophys. Res. Atmos.*, **118**, 13,421 - 13,433, doi: 10.1002/2013JD020908.

Garny, H. and W.J. Randel, 2016: Transport pathways from the Asian monsoon anticyclone to the stratosphere. *Atmos. Chem. Phys.*, **16**, 2703 - 2718, doi: 10.5194/acp-16-2703-2016.

Gettelman, A., *et al.*, 2010: Multimodel assessment of the upper troposphere and lower stratosphere: Tropics and global trends. *J. Geophys. Res.*, **115**, D00M08, doi: 10.1029/2009JD013638.

Gill, A.E., 1980: Some simple solutions for heat-induced tropical circulation. *Q. J. R. Meteorol. Soc.*, **106**, 447 - 462, doi:10.1002/qj.49710644905.

Haimberger, L., 2007: Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics. *J. Climate*, **20**, 1377 - 1403, doi: 10.1175/JCLI4050.1, 2007.

Hajj, G.A., *et al.*, 2004: CHAMP and SAC-C atmospheric occultation results and intercomparisons. *J. Geophys. Res.-Atmos.*, **109**, D06109, doi: 10.1029/2003JD003909.

Heath, N.K. and H.E. Fuelberg, 2014: Using a WRF simulation to examine regions where convection impacts the Asian summer monsoon anticyclone. *Atmos. Chem. Phys.*, **14**, 2055 - 2070, doi: 10.5194/acp-14-2055-2014.

Henderson, D.S., *et al.*, 2013: A Multisensor Perspective on the Radiative Impacts of Clouds and Aerosols. J. Appl. Meteor. Climat., **52**, 853 - 871, doi: 10.1175/JAMC-D-12-025.1.

Hersbach, H., *et al.*, 2020: The ERA5 global reanalysis. *Q.J.R. Meteorol. Soc.*, **146**, 1999 - 2049, doi: 10.1002/qj.3803.

Highwood, E.J. and B.J. Hoskins, 1998: The tropical tropopause. *Q.J.R. Meteorol. Soc.*, **124**, 1579 - 1604. doi: 10.1002/qj.49712454911.

Ho, S.-P., L. Peng, and H. Vömel, H.: Characterization of the long-term radiosonde temperature biases in the upper troposphere and lower stratosphere using COSMIC and Metop-A/GRAS data from 2006 to 2014. *Atmos. Chem. Phys.*, **17**, 4493 - 4511, doi: 10.5194/acp-17-4493-2017.

Hoffmann, L., *et al.*, 2019: From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations, *Atmos. Chem. Phys.*, **19**, 3097 - 3124, doi: 10.5194/acp-19-3097-2019.

Holton, J.R. and A. Gettelman, 2001: Horizontal transport and the dehydration of the stratosphere. *Geophys. Res. Lett.*, **28**, 2799 - 2802, doi: 10.1029/2001GL013148.

Homeyer, C.R., K.P. Bowman, and L.L. Pan, 2010: Extratropical tropopause transition layer characteristics from high-resolution sounding data. *J. Geophys. Res.*, **115**, D13108, doi: 10.1029/2009JD013664.

Hsu, C.J. and R.A. Plumb, 2000: Nonaxisymmetric Thermally Driven Circulations and Upper-Tropospheric Monsoon Dynamics. *J. Atmos. Sci.*, **57**, 1255 - 1276, doi: 10.1175/1520-0469(2000)057<1255:NTDCAU>2.0.CO;2.

Johansson, E., *et al.*, 2015: The vertical structure of cloud radiative heating over the Indian subcontinent during summer monsoon. *Atmos. Chem. Phys.*, **15**, 11557 - 11570, doi: 10.5194/acp-15-11557-2015.

Kanamitsu, M., *et al.*, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Am. Meteorol. Soc.*, **83**, 1631 - 1643, doi: 10.1175/BAMS-83-11-1631.

Kay, J.E. and A. Gettelman, 2009: Cloud influence on and response to seasonal Arctic sea ice loss. *J. Geophys. Res. Atmos.*, **114**, D18204, doi: 10.1029/2009JD011773.

Kiladis, G.N., *et al.*, 2009: Convectively coupled equatorial waves. *Rev. Geophys.*, **47**, RG2003, doi: 10.1029/2008RG000266.

Kim, J. and S. Son, 2012: Tropical Cold-Point Tropopause: Climatology, Seasonal Cycle, and Intraseasonal Variability Derived from COSMIC GPS Radio Occultation Measurements. *J. Clim.*, **25**, 5343 - 5360, doi: 10.1175/JCLI-D-11-00554.1.

Kim, Y.-H., *et al.*, 2019: Comparison of equatorial wave activity in the tropical tropopause layer and stratosphere represented in reanalyses. *Atmos. Chem. Phys.*, **19**, 10027 - 10050, doi: 10.5194/acp-19-10027-2019.

Krüger, K., Tegtmeier, S., and Rex, M., 2008: Long-term climatology of air mass transport through the Tropical Tropopause Layer (TTL) during NH winter, *Atmos. Chem. Phys.*, **8**, 813 - 823, doi: 10.5194/acp-8-813-2008.

Krüger, K., S. Tegtmeier, and M. Rex, 2009: Variability of residence time in the Tropical Tropopause Layer during Northern Hemisphere winter. *Atmos. Chem. Phys.*, **9**, 6717 - 6725, doi: 10.5194/acp-9-6717-2009.

Lawrence, Z.D., G.L. Manney, and K. Wargan, 2018: Reanalysis intercomparisons of stratospheric polar processing diagnostics. *Atmos. Chem. Phys.*, **18**, 13547 - 13579, doi: 10.5194/acp-18-13547-2018.

L'Ecuyer, T. S., *et al.*, 2008: Impact of clouds on atmospheric heating based on the R04 CloudSat fluxes and heating rates data set. *J. Geophys. Res.*, **113**, doi: 10.1029/2008JD009951.

Legras, B. and S. Bucci, 2020: Confinement of air in the Asian monsoon anticyclone and pathways of convective air to the stratosphere during the summer season. *Atmos. Chem. Phys.*, **20**, 11045 - 11064, doi: 10.5194/acp-20-11045-2020.

Li, Y. and M. Zhang, 2016: Cumulus over the Tibetan Plateau in the Summer Based on CloudSat–CALIPSO Data. *J. Clim.*, **29**, 1219 - 1230, doi: 10.1175/JCLI-D-15-0492.1.

Li, Y. and M. Zhang, 2017: The Role of Shallow Convection over the Tibetan Plateau. *J. Clim.*, **30**, 5791 - 5803, doi: 10.1175/JCLI-D-16-0599.1.

Li, Q., *et al.*, 2005: Convective outow of South Asian pollution: A global CTM simulation compared with EOS MLS observations. *Geophys. Res. Lett.*, **32**, L14826, doi: 10.1029/2005GL022762.

Liebmann, B. and C.A. Smith, 1996: Description of a complete (interpolated) outgoing LW radiation dataset. *Bull. Am. Meteor. Soc.*, **77**, 1275 - 1277, 1996.

Linz, M., *et al.*, 2019: The global diabatic circulation of the stratosphere as a metric for the Brewer–Dobson circulation, *Atmos. Chem. Phys.*, **19**, 5069 - 5090, doi: 10.5194/acp-19-5069-2019.

Liu, C., and E.J. Zipser, 2005: Global distribution of convection penetrating the tropical tropopause. *J. Geophys. Res.*, **110**, D23104, doi: 10.1029/2005JD006063.

Loeb, N.G., *et al.*, 2020: Toward a Consistent Definition between Satellite and Model Clear-Sky Radiative Fluxes. *J. Clim.*, **33** 61–75, doi: 10.1175/JCLI-D-19-0381.1.

Long, C.S., *et al.*, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14593 - 14629, doi: 10.5194/acp-17-14593-2017.

Louis, J.-F., 1979: Parametric model of vertical eddy fluxes in the atmosphere. *Boundary Layer Meteorol.*, **17**, 187 - 202, doi: 10.1007/BF00117978.

Louis, J.-F., M. Tiedtke, and J.-F. Geleyn, 1982: A short history of the PBL parameterization at ECMWF. *Workshop on Planetary Boundary Layer Parameterization*, 25-27 November 1981, Shinfield Park, Reading, UK, available at **https://www.ecmwf.int/en/elibrary/10845-shorthistory-pbl-parameterization-ecmwf** (accessed August 2018).

Lucas, C., B. Timbal, and H. Nguyen, 2014: The expanding tropics: A critical assessment of the observational and modeling studies. *Wiley Interdiscip. Rev.: Climate Change*, **5**, 89 - 112, doi: 10.1002/wcc.251.

Mace, G.G. and Q. Zhang, 2014: The CloudSat radar-lidar geometrical profile product (RL-GeoProf): Updates, improvements, and selected results: CLOUDSAT RADAR-LIDAR GEOMETRICAL PROFILE. *J. Geophys. Res. Atmos.*, **119**, 9441 - 9462, doi: 10.1002/2013JD021374.

Madden, R.A., 2007: Large-scale, free Rossby waves in the atmosphere - an update. *Tellus A*, **59**, 571 - 590, doi: 10.1111/j.1600-0870.2007.00257.x.

Madden, R.A., and P.R. Julian, 1994: Observations of the 40 - 50-day tropical oscillation - A review. *Mon. Weather Rev.*, **122**, 814 - 837, doi: 10.1175/1520-0493(1994)122<0814:OOTDTO>2.0.CO;2.

Manney, G.L., and M.I. Hegglin, 2018: Seasonal and Regional Variations of Long-Term Changes in Upper Tropospheric Jets from Reanalyses. *J. Clim.*, **31**, 423 - 448, doi: 10.1175/JCLI-D-17-0303.1.

Manney, G.L., *et al.*, 2014, Climatology of Upper Tropospheric/Lower Stratospheric (UTLS) Jets and Tropopauses in MERRA, *J. Clim.*, **27**, 3248 - 3271, doi: 10.1175/JCLI-D-13-00243.1.

Manney, G.L., *et al.*, 2017, Reanalysis comparisons of upper tropospheric/lower stratospheric jets and multiple tropopauses. *Atmos. Chem. Phys.*, **17**, 11,541 - 11,566, doi: 10.5194/acp-17-11541-2017.

Manney, G.L., *et al.*, 2011: Jet characterization in the upper troposphere/lower stratosphere (UTLS): Applications to climatology and transport studies. *Atmos. Chem. Phys.*, **11**, 6115 - 6137, doi: 10.5194/acp-11-6115-2011.

Manney, G.L., *et al.*, 2021: A Moments View of Climatology and Variability of the Asian Summer Monsoon Anticyclone. *J. Clim.*, in press.

Martin, E.R., *et al.*, 2020: Regionally Varying Assessments of Upper-Level Tropical Width in Reanalyses and CMIP5 Models Using a Tropopause Break Metric. *J. Clim.*, **33**, 5885 - 5903, doi: 10.1175/JCLI-D-19-0629.1.

Martineau, P., J.S. Wright, N. Zhu, and M. Fujiwara, 2018: Zonal-mean data set of global atmospheric reanalyses on pressure levels. *Earth Syst. Sci. Data*, **10**, 1925 - 1941, doi: 10.5194/essd-10-1925-2018.

Mason, R.B. and C.E. Anderson, 1963: The development and decay of the 100-mb. summertime anticyclone over southern Asia. *Mon. Weather Rev.*, **91**, 3 - 12, doi: 10.1175/1520-0493(1963)091<0003:TDADOT>2.3.CO;2.

Matthewman, N.J., J.G. Esler, A.J. Charlton-Perez, and L.M. Polvani, 2009: A new look at stratospheric sudden warmings. Part III: Polar vortex evolution and vertical structure. *J. Clim.*, **22**, 1566 - 1585, doi: 10.1175/2008JCLI2365.1.

Matsuno, T., 1966: Quasi-Geostrophic Motions in the Equatorial Area. *J. Meteorol. Soc. Japan*, **44**, 25 - 43, doi: 10.2151/jmsj1965.44.1_25.

Ming, A., P. Hitchcock, and P. Haynes, 2016: The Double Peak in Upwelling and Heating in the Tropical Lower Stratosphere. *J. Atmos. Sci.*, **73**, 1889 - 1901, doi: 10.1175/jas-d-15- 0293.1.

Montgomery R., 1937: A suggested method for representing gradient flow in isentropic surfaces. *Bull. Am. Meteorol. Soc.*, **18**, 210 - 212, doi: 10.1175/1520-0477-18.6-7.210.

Mote, P.W., *et al.*, 1996: An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor. *J. Geophys. Res.*, **101**, 3989 - 4006, doi: 10.1029/95JD03422.

Nishimoto, E., and M. Shiotani, 2012: Seasonal and interannual variability in the temperature structure around the tropical tropopause and its relationship with convective activities. *J. Geophys. Res.*, **117**, D02104, doi: 10.1029/2011JD016936.

Nützel, M., M. Dameris, and H. Garny, H., 2016: Movement, drivers and bimodality of the South Asian High. *Atmos. Chem. Phys.*, **16**, 14755 - 4774, doi: 10.5194/acp-16-14755-2016.

Pan, L.L., *et al.*, 2016: Transport of chemical tracers from the boundary layer to stratosphere associated with the dynamics of the asian summer monsoon. *J. Geophys. Res. Atmos.*, **121**, 14,159 - 14,174. doi: 10.1002/2016JD025616.

Park, M., *et al.*, 2007: Transport above the Asian summer monsoon anticyclone inferred from Aura Microwave Limb Sounder tracers. *J. Geophys. Res. Atmos.*, **112**, D16309, doi: 10.1029/2006JD008294.

Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., and Hofmann, R. J. P.: Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators, J. Climate, 25, 4699–4720, doi:10.1175/JCLI-D-11-00267.1, 2012.

Platnick, S., 2015: MODIS Atmosphere L3 Daily Product. doi:10.5067/modis/mod08_d3.061.

Ploeger, F., *et al.*, 2010, Impact of the vertical velocity scheme on modeling transport in the tropical tropopause layer. *J. Geophys. Res.*, **115**, D0330, doi: 10.1029/2009JD012023.

Ploeger, F., *et al.*, 2012: Horizontal transport affecting trace gas seasonality in the Tropical Tropopause Layer (TTL). *J. Geophys. Res.*, **117**, D09303, doi: 10.1029/2011JD017267.

Ploeger, F., *et al.*, 2015: A potential vorticity-based determination of the transport barrier in the Asian summer monsoon anticyclone. *Atmos. Chem. Phys.*, **15**, 13145 - 13159, doi: 10.5194/acp-15-13145-2015.

Ploeger, F., P. Konopka, K. Walker, and M. Riese, 2017: Quantifying pollution transport from the Asian monsoon anticyclone into the lower stratosphere. *Atmos. Chem. Phys.*, **17**, 7055 - 7066, doi: 10.5194/acp-17-7055-2017.

Popovic, J.M. and R.A. Plumb, 2001: Eddy Shedding from the Upper-Tropospheric Asian Monsoon Anticyclone. *J. Atmos. Sci.*, **58**, 93 - 104, doi: 10.1175/1520-0469(2001)058<0093:ESFTUT>2.0.CO;2.

Randel, W.J. and M. Park, 2006: Deep convective influence on the Asian summer monsoon anticyclone and associated tracer variability observed with Atmospheric Infrared Sounder (AIRS). *J. Geophys. Res.*, **111**, D12314, doi: 10.1029/2005JD006490.

Randel, W.J., *et al.*, 2009: An update of observed stratospheric temperature trends. *J. Geophys. Res.*, **114**, D02107, doi: 10.1029/2008JD010421, 2009.

Randel W.J., and E. Jensen, 2013, Physical processes in the tropical tropopause layer and their roles in a changing climate. *Nat Geosci.* **6,** 169 - 176, doi: 10.1038/ngeo1733.

Randel, W.J., and F. Wu, 2015: Variability of Zonal Mean Tropical Temperatures Derived from a Decade of GPS Radio Occultation Data. *J. Atmos. Sci.*, **72**, 1261-1275, doi: 10.1175/JAS-D-14-0216.1.

Randel, W.J., *et al.*, 2004, Interannual Changes of Stratospheric Water Vapor and Correlations with Tropical Tropopause Temperatures. *J. Atmos. Sci.*, **61**, 2133 - 2148, doi: 10.1175/1520-0469.

Randel, W.J., *et al.*, 2010: Asian Monsoon Transport of Pollution to the Stratosphere. *Science*, **328**, 611 - 613, doi: 10.1126/science.1182274.

Richter, J.H., A. Solomon, and J.T. Bacmeister, 2014: On the simulation of the quasi-biennial oscillation in the Community Atmosphere Model, version 5. *J. Geophys. Res. Atmos.*, **119**, 3045 - 3062, doi: 10.1002/2013JD021122.

Rossow, W.B. and R.A. Schiffer, 1991: ISCCP cloud data products. *Bull. Am. Meteor. Soc.*, **72**, 2 - 20, doi: 10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2.

Rossow, W. B. and Schiffer, R. A.: Advances in understanding clouds from ISCCP, Bull. Am. Meteor. Soc., 80, 1999.

Rossow, W.B., *et al.*, 2017: International Satellite Cloud Climatology Project (ISCCP) Climate Data Record, H-Series. *NOAA National Centers for Environmental Information*, USA, doi: 10.7289/V5QZ281S.

Saha, S., *et al.*, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Am. Meteorol. Soc.*, **91**, 1015 - 1057, doi: 10.1175/2010BAMS3001.1.

Santee, M.L., et al., 2017: A comprehensive overview of the climatological composition of the Asian summer monsoon anticyclone based on 10 years of Aura Microwave Limb Sounder measurements. *J. Geophys. Res.*, **122**, 5491 - 5514, doi: 10.1002/2016JD026408.

Schneider, T., Bischoff, T., and Haug, G. H.: Migrations and dynamics of the intertropical convergence zone, Nature, 513, 45–53, doi:10.1038/nature13636, 2014.

Schoeberl, M., A. Douglass, Z. Zhu, and S. Pawson, 2003: A comparison of the lower stratospheric age spectra derived from a general circulation model and two data assimilation systems. *J. Geophys.Res.*, **108**, 4113, doi: 10.1029/2002JD002652.

Schoeberl, M.R. and A.E. Dessler, 2011: Dehydration of the stratosphere. *Atmos. Chem. Phys.*, **11**, 8433 - 8446, doi: 10.5194/acp-11-8433-2011.

Schoeberl, M.R., A.E. Dessler, and T. Wang, 2012: Simulation of stratospheric water vapor and trends using three reanalyses, *Atmos. Chem. Phys.*, **12**, 6475 - 6487, doi: 10.5194/acp-12- 6475-2012.

Schoeberl, M.R., A.E. Dessler, and T. Wang, 2013: Modeling upper tropospheric and lower stratospheric water vapor anomalies. *Atmos. Chem. Phys.*, **13**, 7783 - 7793, doi: 10.5194/acp-13- 7783-2013.

Seidel, D.J., Q. Fu, W.J. Randel, and T.J. Reichler, 2008: Widening of the tropical belt in a changing climate. *Nat. Geosci.*, **1**, 21 - 24, doi: 10.1038/ngeo.2007.38.

Seidel, D.J., and W.J. Randel, 2006: Variability and trends in the global tropopause estimated from radiosonde data. *J. Geophys. Res.*, **111**, D21101, doi: 10.1029/2006JD007363.

Seidel, D.J., R.J. Ross, J.K. Angell, and G.C. Reid, 2001: Climatological characteristics of the tropical tropopause as revealed by radiosondes. *J. Geophys. Res.*, **106**, 7857 - 7878, doi: 10.1029/2000JD900837.

Solomon, A., L.M. Polvani, D.W. Waugh, and S.M. Davis, 2016: Contrasting upper and lower atmospheric metrics of tropical expansion in the Southern Hemisphere. *Geophys. Res. Lett.*, **43**, 10,496 - 10,503, doi: 10.1002/2016GL070917.

Stengel, M., *et al.*, 2018: Comparing ERA-Interim clouds with satellite observations using a simplified satellite simulator. *Atmos. Chem. Phys.*, **18**, 17601 - 17614, doi: 10.5194/acp-18-17601-2018.

Stevens, B. and S.E. Schwartz, 2012: Observing and modeling Earth's energy flows. *Surv. Geophys.*, **33**, 779 - 816, doi: 10.1007/s10712-012-9184-0.

Su, H., *et al.*, 2011: Comparison of regime-sorted tropical cloud profiles observed by CloudSat with GEOS5 analyses and two general circulation models. *J. Geophys. Res. Atmos.*, **116**, D09104, doi:10.1029/2010JD014971.

Tao, M., *et al.*, 2019: Multitimescale variations in modeled stratospheric water vapor derived from three modern reanalysis products. *Atmos. Chem. Phys.*, **19**, 6509 - 6534, doi: 10.5194/acp-19-6509-2019.

Tegtmeier, S., *et al.*, 2020: Temperature and tropopause characteristics from reanalyses data in the tropical tropopause layer. *Atmos. Chem. Phys.*, **20**, 753 - 770, doi: 10.5194/acp-20-753-2020.

Texeira, A.S.T., 2013: Aqua AIRS Level 3 Standard Daily Product using AIRS and AMSU without HSB V6. doi: 10.5067/aqua/airs/data301, accessed: 2017-03-13.

Thompson, A.M., *et al.*, 2003: Southern Hemisphere Additional Ozoneson- des (SHADOZ) 1998 - 2000 tropical ozone climatology 1. Comparison with Total Ozone Mapping Spectrometer (TOMS) and ground-based measurements. *J. Geophys. Res. Atmos.*, **108**, 8238, doi: 10.1029/2001JD000967.

Thorne, P.W., *et al.*, 2005: Revisiting radiosonde upper air temperatures from 1958 to 2002. *J. Geophys. Res.*, **110**, D18105, doi: 10.1029/2004JD005753, 2005.

Tian, B., *et al.*, 2013: AIRS/AMSU/HSB Version 6 Level 3 Product User Guide. *Tech. rep.*, Jet Propulsion Laboratory, Pasadena, California.

Tissier, A.-S. and B. Legras, 2016: Convective sources of trajectories traversing the tropical tropopause layer. *Atmos. Chem. Phys.*, **16**, 3383 - 3398, doi: 10.5194/acp-16-3383-2016.

von Engeln, A., Y. Andres, C. Marquardt, and F. Sancho, 2001: GRAS radio occultation on-board of Metop. *Adv. Space Res.*, **47**, 336 - 347, doi: 10.1016/j.asr.2010.07.028.

von Hobe, *et al.*, 2021: Upward transport into and within the Asian monsoon anticyclone as inferred from StratoClim trace gas observations. *Atmos. Chem. Phys.*, **21**, 1267 - 1285, doi: 10.5194/acp-21-1267-2021.

Wang, T., *et al.*, 2015: The impact of temperature vertical structure on trajectory modeling of stratospheric water vapor. *Atmos. Chem. Phys.*, **15**, 3517 - 3526, doi: 10.5194/acp-15-3517-2015.

Wang, T., *et al.*, 2014: Trajectory model simulations of ozone (O_3) and carbon monoxide (CO) in the lower stratosphere. *Atmos. Chem. Phys.*, **14**, 7135 - 7147, doi: 10.5194/acp-14-7135-2014.

Wang, J.S., D.J. Seidel,and M. Free, 2012, How well do we know recent climate trends at the tropical tropopause? *J. Geophys. Res.*, **117**, D09118, doi: 10.1029/2012JD017444.

Waugh, D.W., *et al.*, 2018: Revisiting the Relationship among Metrics of Tropical Expansion. *J. Clim.*, **31**, 7565 - 7581, doi: 10.1175/JCLI-D-18-0108.1.

Wickert, J., *et al.*, 2001: Atmosphere sounding by GPS radio occultation: First results from CHAMP. *Geophys. Res. Lett.*, **28**, 3263 - 3266, doi: 10.1029/2001GL013117.

World Meteorological Organization, 1957: Definition of the tropopause. *Bull. World Met. Organ.*, **6**, 136 - 137.

Wright, J.S., *et al.*, 2011: The influence of summertime convection over Southeast Asia on water vapor in the tropical stratosphere. *J. Geophys. Res.*,**116**, D12302, doi: 10.1029/2010JD015416.

Wright, J.S. and S. Fueglistaler, 2013: Large differences in reanalyses of diabatic heating in the tropical upper troposphere and lower stratosphere. *Atmos. Chem. Phys.*, **13**, 9565 - 9576, doi: 10.5194/acp-13-9565-2013.

Wright, J.S., *et al.*, 2020: Differences in tropical high clouds among reanalyses: origins and radiative impacts. *Atmos. Chem. Phys.*, **20**, 8989 - 9030, doi: 10.5194/acp-20-8989-2020.

Yan, R., J. Bian, and Q. Fan, 2011: The Impact of the South Asia High Bimodality on the Chemical Composition of the Upper Troposphere and Lower Stratosphere. *Atmos. Ocean. Sci. Lett.*, **4**, 229 - 234, doi: 10.1080/16742834.2011.11446934.

Yang, Q., Q. Fu, and Y. Hu, 2010: Radiative impacts of clouds in the tropical tropopause layer.*J. Geophys. Res.*,**115**,D00H12, doi: 10.1029/2009JD012393.

Zelinka, M.D., S.A. Klein, and D.L. Hartmann, 2012: Computing and partitioning cloud feedbacks using cloud property histograms. Part II: Attribution to changes in cloud amount, altitude, and optical depth. *J. Clim.*, **25**, 3736 - 3754, doi: 10.1175/JCLI-D-11-00249.1.

Zhang, K., W.J. Randel, and R. Fu, 2017: Relationships between outgoing LW radiation and diabatic heating in reanalyses. *Clim. Dyn.*, **49**, 2911 - 2929, doi: 10.1007/s00382-016-3501-0.

Zhang, Q., G. Wu, and Y. Qian, 2002: The Bimodality of the 100 hPa South Asia High and its Relationship to the Climate Anomaly over East Asia in Summer. *J. Meteorol. Soc. Jpn.*, **80**, 733 - 744, doi: 10.2151/jmsj.80.733.

Zhang, T., *et al.*, 2013: The NASA GEWEX surface radiation budget project: Dataset validation and climatic signal identification. *AIP Conference Proceedings*, **1531**, 636 - 639, doi: 10.1063/1.4804850.

Zhou, X.J., C. Luo, W.L. Li, and J.E. Shi, 1995: Variations of total ozone amount in China and the ozone low center over the Tibetan Plateau. *China Sci. Bull.*, **40**, 1396 - 1398, 1995 (in Chinese).

Zhou, X., M. Geller, and M. Zhang, 2001: Tropical cold point tropopause characteristics from ECMWF reanalyses and soundings. J. *Climate*, **14**, 1823 - 1838, doi: 10.1175/1520-0442(2001)014<1823:TCPTCD>2.0.CO;2.

Appendix A: Supplementary material

A8.1 Supplementary material for Section 8.2

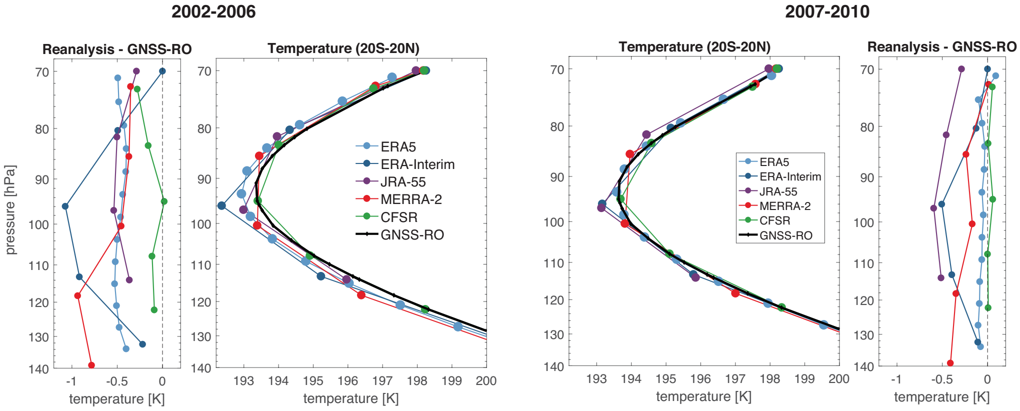


Figure A8.1: *Tropical mean (20 °S - 20 °N) temperature profiles at reanalyses model levels between 140 hPa and 70 hPa and difference between reanalyses and GNSS-RO temperatures. Left panels for 2002 - 2006 and right panels for 2007 - 2010.*

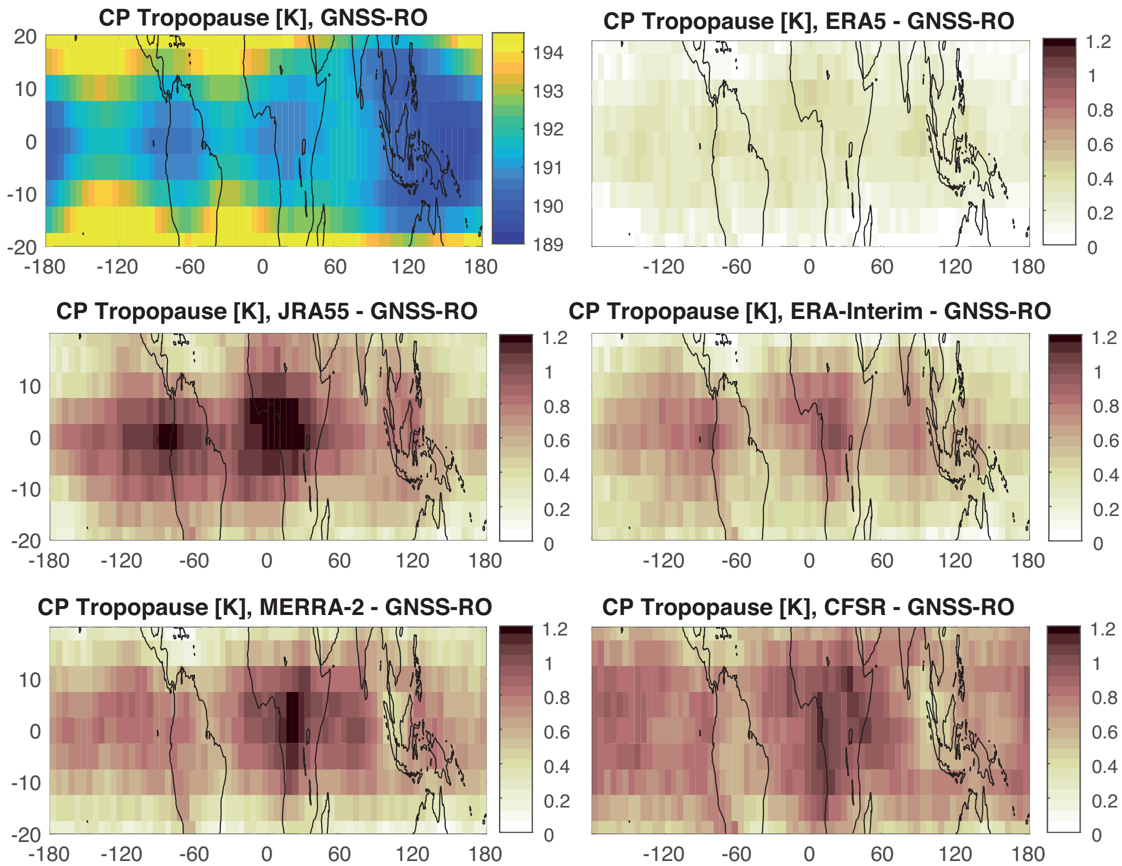


Figure A8.2: *Latitude–longitude distributions of annual mean GNSS-RO cold point (CP) tropopause temperatures (upper left) and differences between cold point tropopause temperatures from individual reanalyses and those from GNSS-RO during 2007 - 2010 (lower panels).*

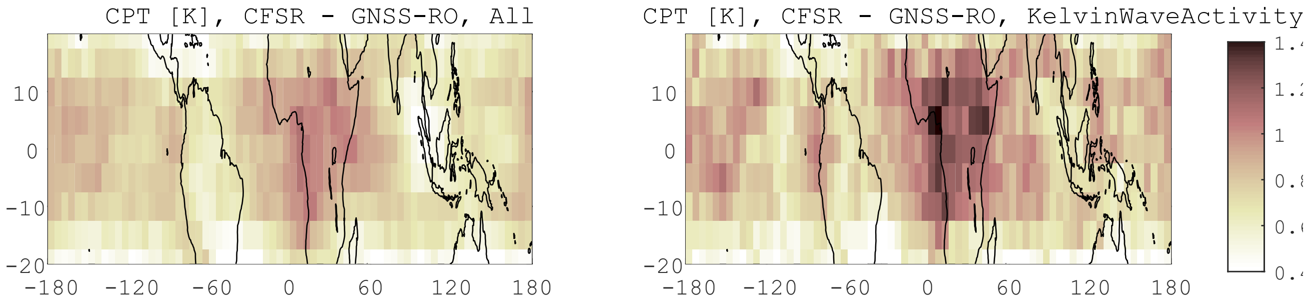


Figure A8.3: *Latitudinal-longitude sections of the differences between GNSS-RO and CFSR cold point temperatures for 2007 - 2010 (left panel) and for time periods of high wave activity (right panel).*

A8.2 Supplementary material for Section 8.4

Figures A8.4 and **A8.5** illustrate aspects of the seasonal cycle of diabatic heating within the TTL and tropical LS. **Figure A8.4** shows zonal-mean distributions of diabatic heating and potential temperature for the DJF and JJA solstice seasons based on ERA-Interim, JRA-55, MERRA-2, and CFSR over 1980 - 2010. It may be compared with the upper row of the annual-mean zonal-mean distributions shown in **Figure 8.20** of the *Chapter 8* main text. **Figure A8.5** shows mean annual cycles of total diabatic heating based on daily-mean data. Unlike the other figures included in *Section 8.4*, **Figure A8.5** uses potential temperature as the vertical coordinate (rather than pressure) and expresses diabatic heating as $\dot{\theta}$ (rather than temperature tendency). The diabatic temperature tendency may be converted to $\dot{\theta}$ by multiplying by the factor $\left(\frac{p}{p_0}\right)^{-\kappa}$, just as in the conversion of temperature to potential temperature. This transformation of the vertical coordinate emphasizes the annual cycle of diabatic heating in the tropical LS. In addition to the annual cycle of the vertical profile of $\dot{\theta}$, **Figure A8.5** shows the mean annual evolution of the vertical location of the maximum diabatic heating within the tropical LS. ERA-Interim, MERRA-2, and CFSR all show that the location of this maximum shifts to lower potential temperatures during boreal summer, but the timing of this shift and the height of the maximum during boreal winter are quite different. By contrast, JRA-55 shows relatively little change in the vertical location of the maximum heating rate, and in fact places the maximum at a higher potential temperature during boreal summer than during boreal winter. These differences have implications for the magnitude and seasonal cycle of the rate at which trajectories ascend through the tropical LS when diabatic heating rates are used to drive vertical motion.

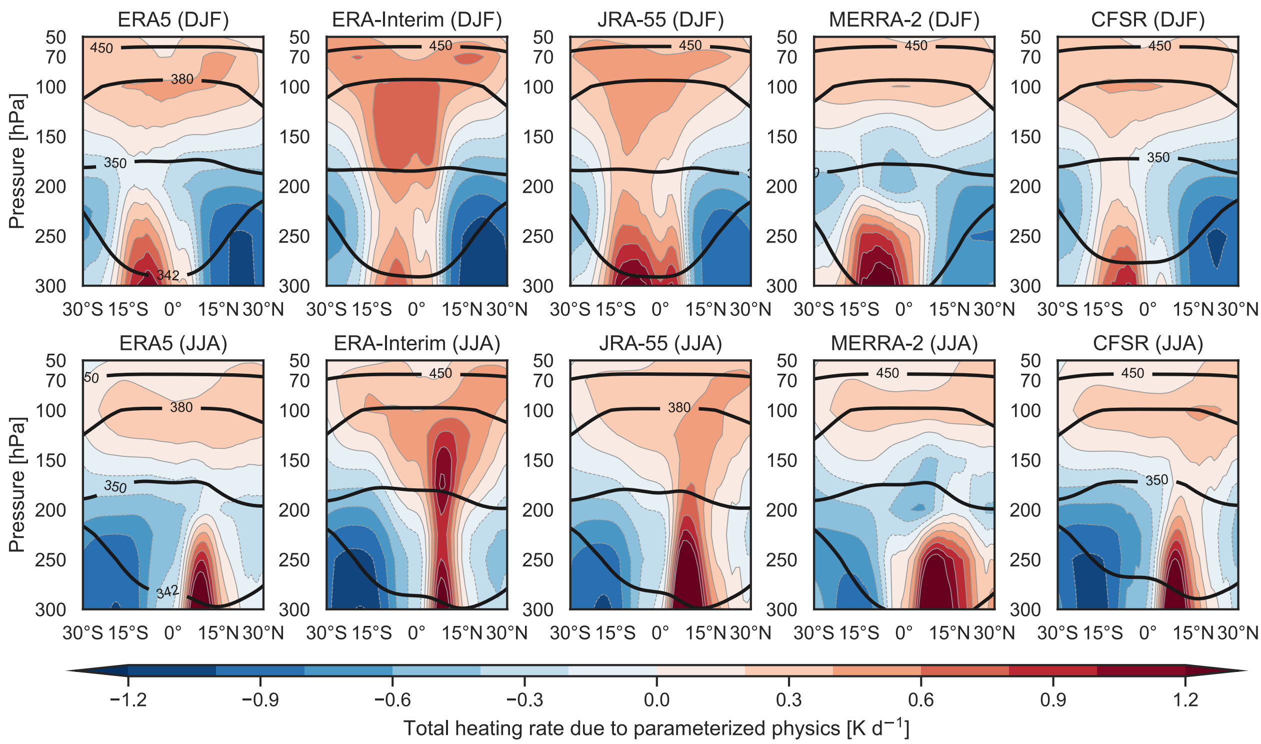


Figure A8.4: *As in **Fig. 8.20**, but for zonal mean total diabatic temperature tendencies (Q/c_p in K day^{-1}; shading and gray contours) and potential temperature (θ in K; black contours) averaged over 1980 - 2010 for the solstice seasons DJF (upper row) and JJA (lower row).*

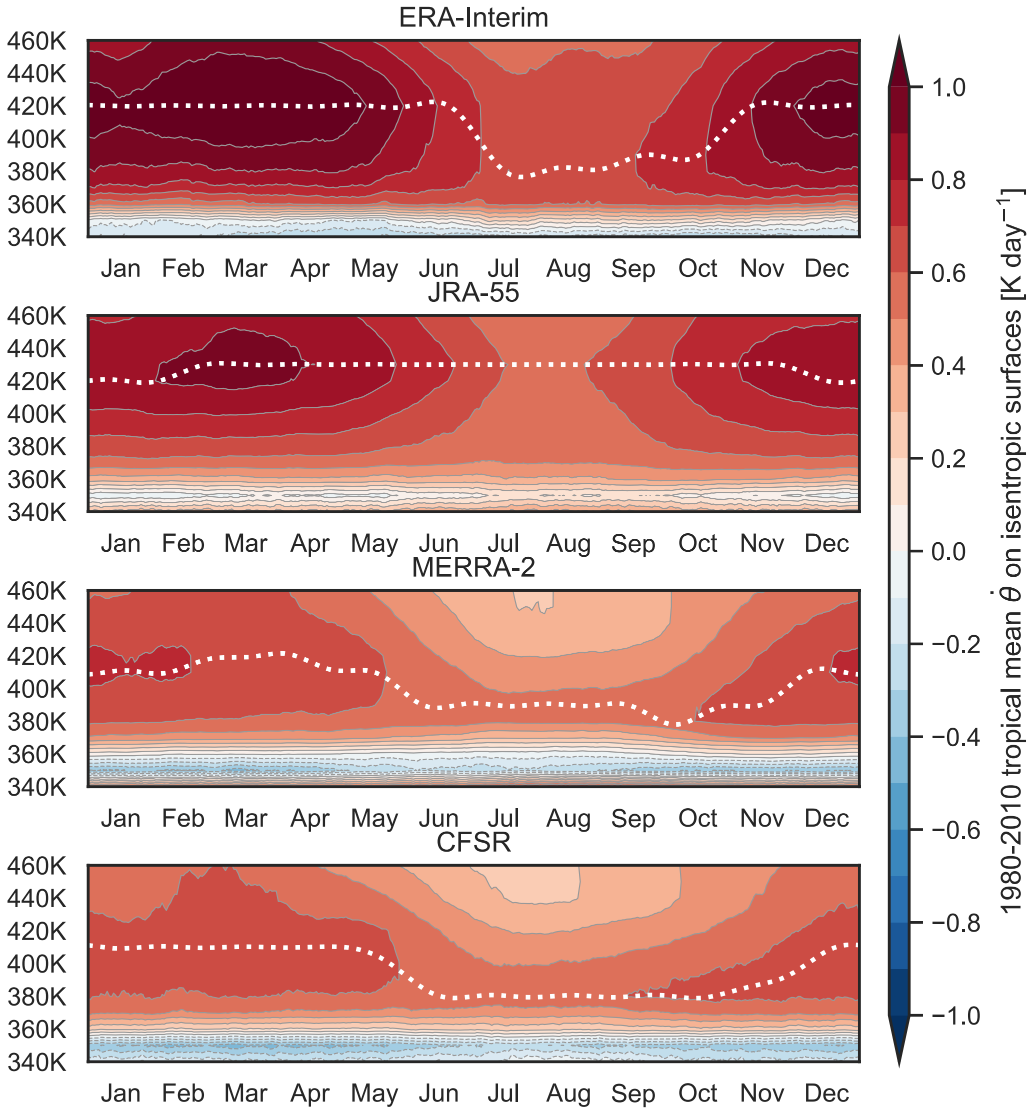


Figure A8.5: *Total diabatic potential temperature tendencies ($\dot{\theta}$ in K day^{-1}; shading and gray contours) averaged over the tropics (30 °S - 30 °N) during 1980 - 2010 for (from top) ERA-Interim, JRA-55, MERRA-2, and CFSR. The vertical location of the maximum in $\dot{\theta}$ is shown as a dotted white line. This figure uses a potential temperature vertical coordinate to better emphasize the annual cycle of diabatic heating within the tropical lower stratosphere.*

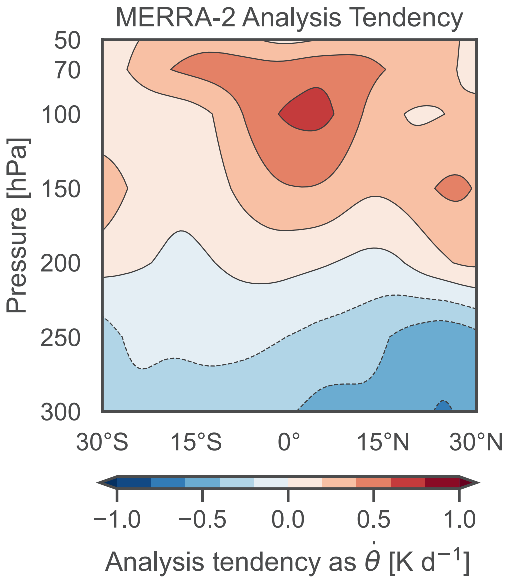


Figure A8.6: *MERRA-2 annual mean analysis tendency 1980 - 2010, as produced by the initial 3D-FGAT data assimilation and applied during the IAU corrector step (see Section 2.3 for details).*

A8.3 Supplementary Material for Section 8.8

Figures A8.7 through **A8.10** show component terms of diabatic heating within the SASM region during JJA 1980 - 2010. **Figures A8.7** and **A8.8** indicate the all-sky and clear-sky radiative heating terms that correspond to the total diabatic heating rates shown in **Figure 8.59** of the *Chapter 8* main text. **Figure A8.9** shows corresponding distributions, but for the sum of all non-radiative components of diabatic heating (see *Sect. 8.4*). **Figure A8.10** shows the spatial distribution of non-radiative components of diabatic heating on the 350 K and 380 K isentropic surfaces within the SASM region and surrounding areas, and corresponds to the total diabatic heating rates shown in **Figure 8.60** of the *Chapter 8* main text.

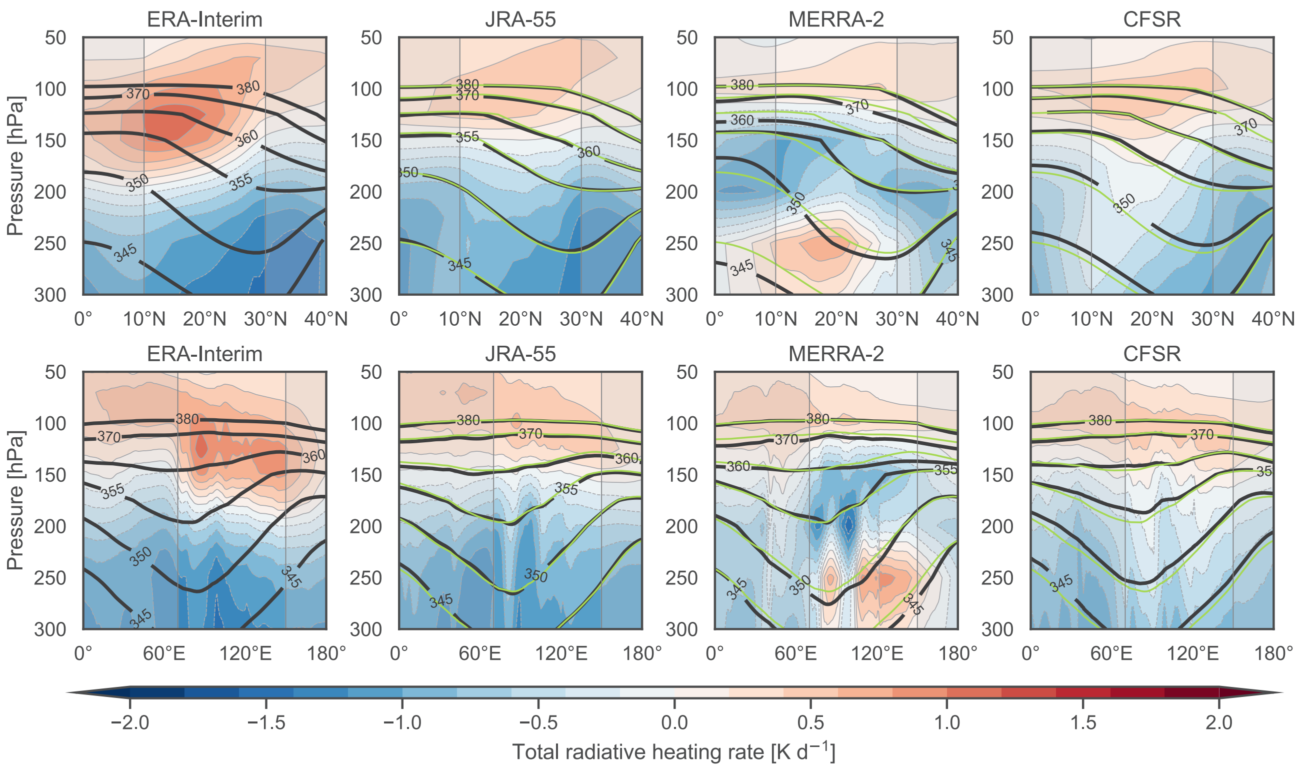


Figure A8.7: *As in* ***Fig. 8.59*** *but for all-sky radiative heating [K day^{-1}] averaged over JJA 1980 - 2010. Zonal means are calculated over 70 ° - 150 ° E (vertical lines in lower panels); area-weighted meridional means over 10 ° - 30 ° N (vertical lines in upper panels). ERA-Interim potential temperature contours are shown in light green on the JRA-55, MERRA-2, and CFSR panels for ease of comparison.*

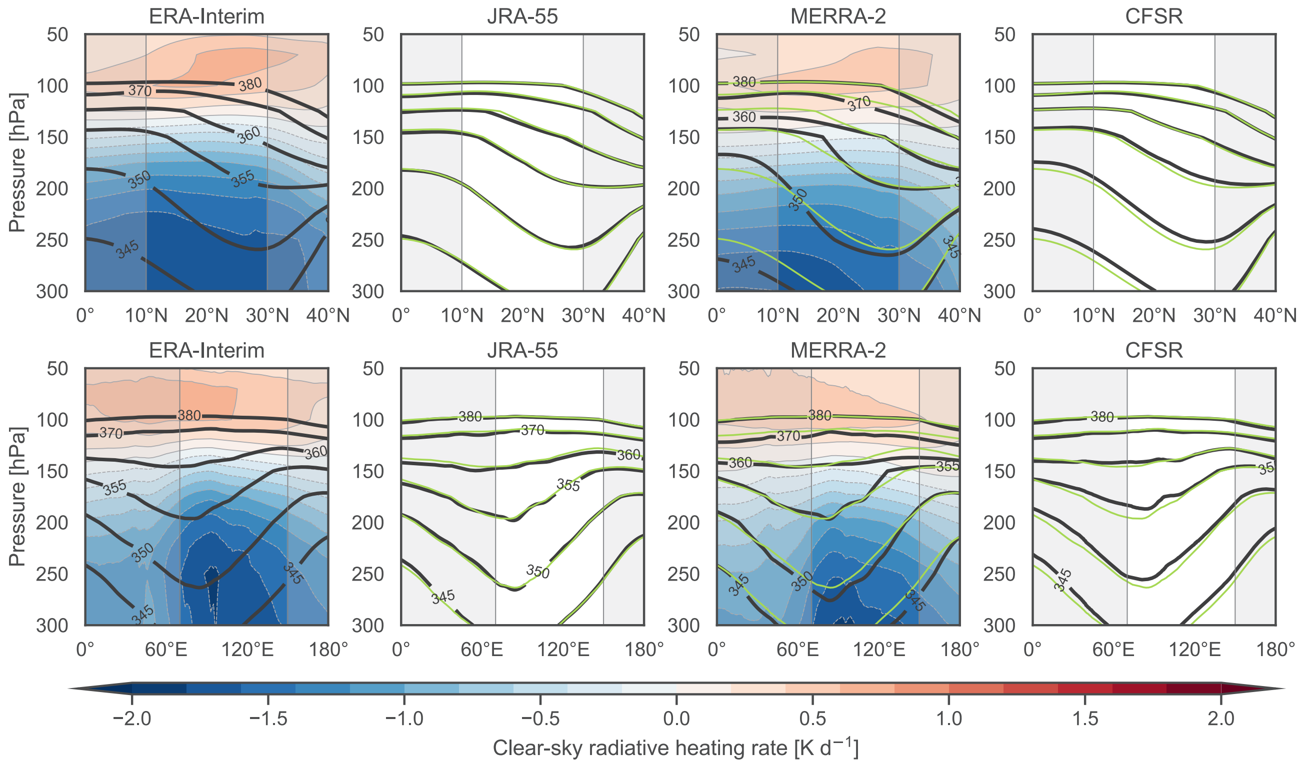


***Figure A8.8:** As in **Fig. 8.59**, but for clear-sky radiative heating [K day^{-1}] averaged over JJA 1980-2010. Zonal means are calculated over 70°-150°E (vertical lines in lower panels); area-weighted meridional means over 10°-30°N (vertical lines in upper panels). ERA-Interim potential temperature contours are shown in light green on the JRA-55, MERRA-2, and CFSR panels for ease of comparison. Clear-sky radiative heating rates are not provided by JRA-55 or CFSR.*

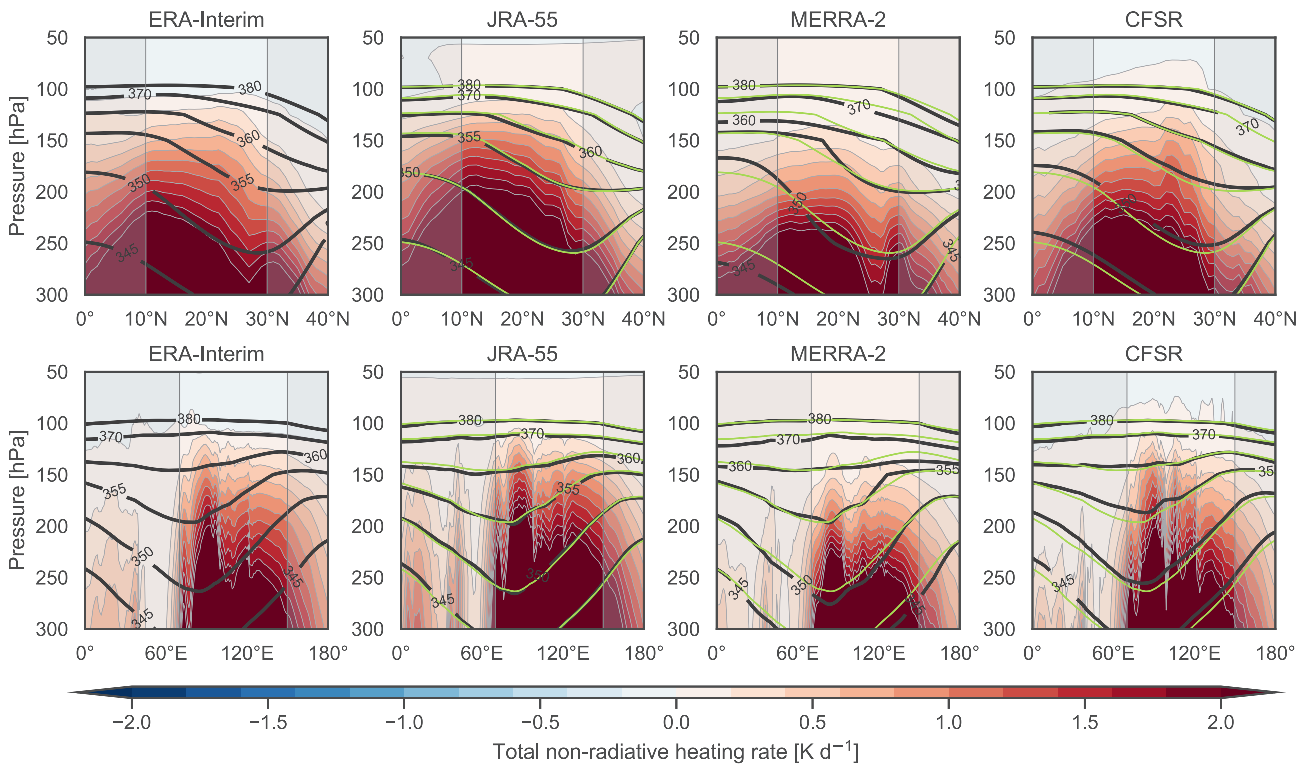


***Figure A8.9:** As in **Fig. 8.59**, but for non-radiative heating [K day^{-1}] averaged over JJA 1980-2010. Zonal means are calculated over 70°-150°E (vertical lines in lower panels); area-weighted meridional means over 10°-30°N (vertical lines in upper panels). ERA-Interim potential temperature contours are shown in light green on the JRA-55, MERRA-2, and CFSR panels for ease of comparison.*

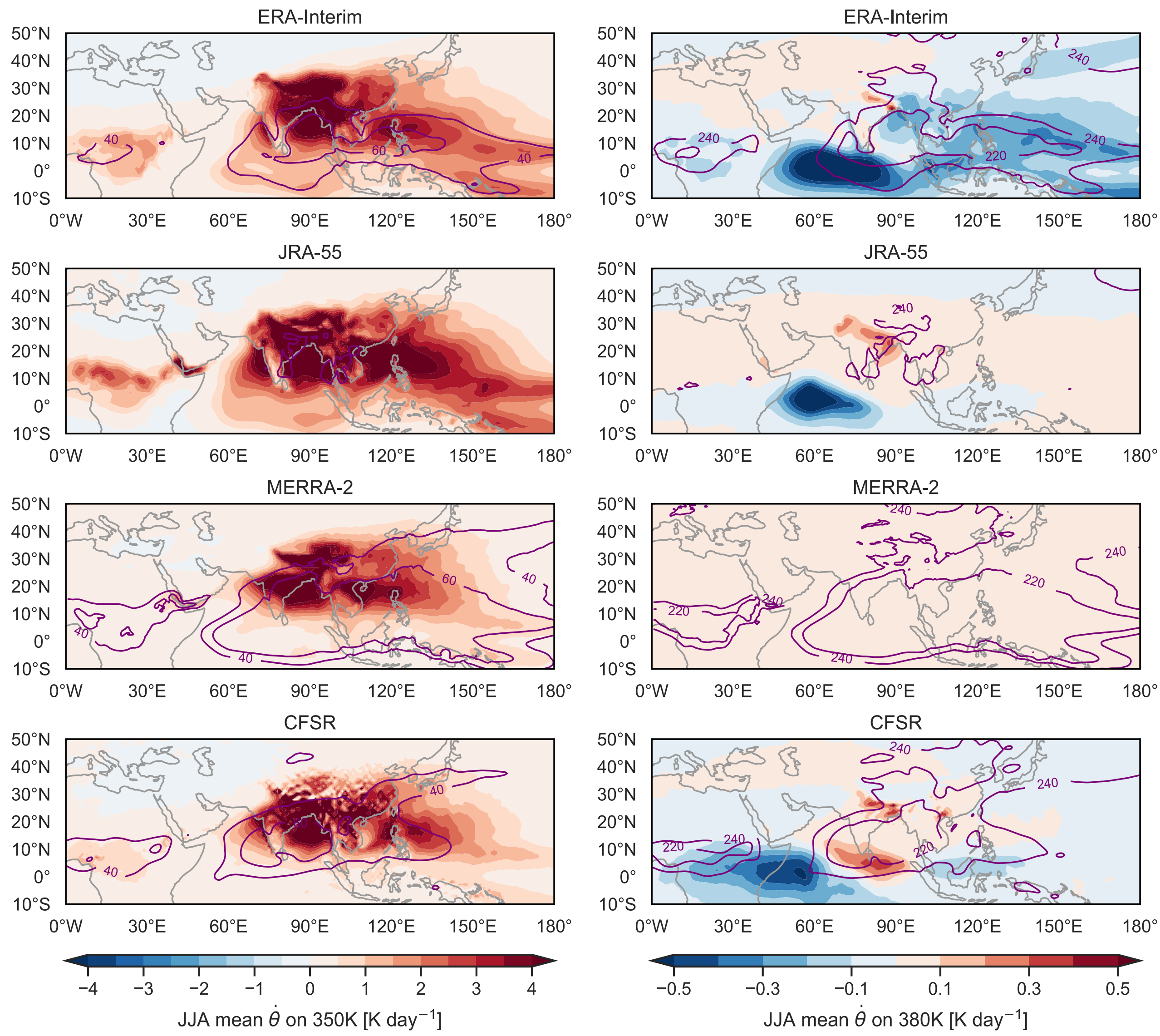


Figure A8.10: *As in **Fig. 8.60** but for non-radiative heating [K day^{-1}] averaged over JJA 1980 - 2010. Purple contours in the left column show LWCRE at the TOA (W m^{-2}, purple contours at 40 W m^{-2} and 60 W m^{-2}) over the same period; purple contours in the right column show OLR at the TOA (W m^{-2}, purple contours at 220 W m^{-2} and 240 W m^{-2}).*

Major abbreviations and terms

AIRS	Atmospheric Infrared Sounder
ANA	Analysed State
ASM	Assimilated State
ATOVS	Advanced TOVS
AUX	Auxiliary
BoB	Bay of Bengal
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO	CloudSat and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CDAAC	COSMIC Data Analysis and Archive Center
CERES	Clouds and the Earth's Radiant Energy System
CFMIP	Cloud Feedback Model Intercomparison Project
CFSR	Climate Forecast System Reanalysis
CHAMP	Challenging Minisatellite Payload
COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate
CP	Cold Point
CPT	Cold Point Tropopause
COSP	CFMIP Observations Simulator Package
CWC	Cloud Water Content
DJF	December, January, February
EBAF	Energy Balanced And Filled
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño Southern Oscillation
ERA5	the fifth major global reanalysis produced by ECMWF
ERA-Interim	ECMWF interim reanalysis
GCM	General Circulation Model
GEOS-4/ 5 DAS	Goddard Earth Observing System Data Assimilation System, version 4/5
GEWEX	Global Energy and Water Cycle Experiment
GHG	Greenhouse Gas
GMAO	Global Modeling and Assimilation Office
GNSS-RO	Global Navigation Satellite System - Radio Occultation
GOCCP	GCM-Oriented CALIPSO Cloud Product
GRACE	Gravity Recovery and Climate Experiment
HGM	High-resolution Global Monthly
FDP	Final Dehydration Point
FLXHR	Fluxes and Heating Rates
HadAT	Hadley Centre radiosonde temperature dataset
HSI-1	Horseshoe-Shaped Structure Index
HSI-K	Horseshoe-Shaped Structure Index Kelvin Response
HSI-R	Horseshoe-Shaped Structure Index Rossby Response
IAU	Incremental Analysis Updates
IFS	Integrated Forecasting System
IGRA	Integrated Global Radiosonde Archive

Indian Sub	Indian Subcontinent
IP	Iranian Plateau (IP)
IPSL	Institut Pierre-Simon Laplace
ISCCP	International Satellite Cloud Climatology Project
ITCZ	Inter Tropical Convection Zone
IWC	Ice Water Content
JETPAC	JEt and Tropopause Products for Analysis and Characterization
JJA	June, July, August
JMA	Japan Meteorological Agency
JRA-25/55	Japanese 25-year Reanalysis / Japanese 55-year Reanalysis
LRT	Lapse Rate Tropopause
Lvq	Latent Energy Component
LW	Long-Wave
LWC	Liquid Water Content
LWCRE	Long-Wave Cloud Radiative Effect
LZRH	Level of Zero net Radiative Heating
MAM	March, April, May
MERRA; MERRA-2	Modern Era Retrospective-Analysis for Research and Applications / Version 2
MJO	Madden Julian Oscillation
MODIS	Moderate Resolution Imaging Spectroradiometer
MRG	Mixed Rossby-Gravity
MSE	Moist Static Energy
MSF	Montgomery Streamfunction
MSU	Microwave Sounding Unit
NASA	National Aeronautics and Space Administration
NCEI	National Centers for Environmental Information
NCEP-DOE R2	Reanalysis 2 of the NCEP and DOE
NCEP-NCAR R1	Reanalysis 1 of the NCEP and NCAR
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
OISST	Optimum Interpolation Sea Surface Temperature
OLR	Outgoing Longwave Radiation
Pen	Indochina Peninsula
PL	Pressure Level
PSD	Power Spectral Density
QBO	Quasi Biennial Oscillation
RAOBCORE	RAdiosonde OBservation COrrection using REanalyses
RATPAC	Radiosonde Atmospheric Temperature Products for Assessing Climate
REM	Reanalyses Ensemble Mean
RH	Relative Humidity
SAC-C	Scientific Application Satellite-C
SASM	South Asian Summer Monsoon
SCSPhi	Sea of China and Philippine Sea
SH	Southern Hemisphere
SHADOZ	Southern Hemisphere ADditional OZonesondes
SON	September, October, November
SPCZ	South Pacific Convergence Zone
SRB	Surface Radiation Budget

SST	Sea Surface Temperature
STJ	Subtropical Jet
SW	Short-Wave
SWCRE	Short-Wave Cloud Radiative Effect
SYN1Deg	Synoptic Radiative Fluxes and Clouds at 1-Degree Resolution
S-RIP	SPARC Reanalysis Intercomparison Project
TerraSAR-X	Terra Synthetic Aperture Radar - X
TOA	Top Of Atmosphere
TOVS	Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder
TP	Tibetan Plateau
TPB	Tropopause Break
TqJoint	Temperature and water vapour (q) Joint data group (AIRS)
TTL	Tropical Tropopause Layer
UT	Upper Troposphere
UTLS	Upper Troposphere Lower Stratosphere
WMO	World meteorological Organization
ZM	Zonal Mean
2B-CWC-RVOD	2B - Cloud Water Content - Radar-Visible Optical Depth
2B-GEOPROF-LIDAR	2B - Geometrical Profile - Lidar
2C-ICE	2C - Ice
3D-FGAT	Three-dimensional First Guess at Appropriate Time
θ_e	Equivalent Potential Temperature

Chapter 9: Quasi-Biennial Oscillation

Chapter lead authors

James A. Anstey	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada	Canada
Lesley J. Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK

Co-authors

Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Ioana Ivanciu	GEOMAR Helmholtz-Zentrum für Oceanforschung Kiel	Germany
Yoshio Kawatani	Japan Agency for Marine-Earth Science and Technology	Japan
George Kiladis	Physical Sciences Division, NOAA/Earth System Research Laboratory	USA
Young-Ha Kim	Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt *previously at: Severe Storm Research Center, Ewha Womans University, South Korea*	Germany
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan
Verena Schenzinger	Institute for Meteorology and Geophysics, University of Vienna	Austria
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Corwin Wright	Centre for Space, Atmospheric and Oceanic Science, University of Bath	UK

Abstract. The diagnostics in this chapter include analysis of the tropical QBO in zonal wind and temperature, tropical waves and the QBO zonal momentum budget, and extra-tropical teleconnections of the QBO. Observations used for validation include operational and campaign radiosondes, and satellite observations from GNSS-RO, HIRDLS, SABER, COSMIC and AIRS. QBO zonal winds and temperatures agree well with observations except for older reanalyses (NCEP-NCAR and NCEP-DOE), where the QBO phase is usually correct but the amplitude is underestimated. Inter-reanalysis spread decreases over time, consistent with increased availability of observations. Most of the spread occurs during QBO phase transitions, especially westerly QBO onsets. Substantial spread in zonal wind strength and spatial structure is found in the tropical upper troposphere and tropopause region, which has implications for modelling of tropical wave filtering. There is good agreement on relative forcing contributions from the various tropical waves, with greatest spread coming from the Kelvin wave contribution during the descending westerly phase. There are also clear differences in wave characteristics when derived on model versus pressure surfaces, and model levels are recommended wherever possible. There is good agreement on the Holton-Tan QBO influence on the NH winter polar vortex, with a clear impact in early winter but an apparent late winter reversal seen in the 1979 - 2016 analysis is not robust over the 1958 - 2016 period. Using the longest available datasets is therefore recommended for these studies. A QBO impact on tropical upper tropospheric winds is seen in boreal winter, of opposite sign to the overlying QBO phase in the lower stratosphere, with an accompanying impact on the subtropical jet in the winter hemisphere. A QBO modulation of mean sea level pressure in NH winter and tropical precipitation over the extended 1958 - 2016 period is confirmed in two different reanalysis datasets.

Contents

9.1 Introduction 393
9.1.1 Issues in representing the QBO 393
9.1.2 Data and methods 395
9.2 Monthly-mean equatorial variability 396
9.2.1 Time evolution 396
9.2.2 QBO amplitude and phase transitions 403
9.2.3 Mean meridional circulation 411
9.2.4 Near-tropopause temperature 413
9.3 Tropical waves and QBO forcing 416
9.3.1 Characteristics of the equatorial waves 418
9.3.2 Momentum budget of the QBO 421
9.3.3 Direct comparison with satellite data 426
9.4 QBO teleconnections 428
9.4.1 Stratospheric teleconnections 429
9.4.2 Tropospheric teleconnections 438
9.4.3 QBO teleconnections in context 439
9.4.4 Surface teleconnections 441
9.5 Summary, key findings, and recommendations 444
9.5.1 Summary for monthly-mean equatorial variability 444
9.5.2 Summary for tropical waves and QBO forcing 445
9.5.3 Summary for QBO teleconnections 446
9.5.4 Recommendations 446
References 449
Appendix A: Additional Figures 453
A9.1 Supplemental 453
A9.2 Supplement A: QBO winds in all reanalyses 478
A9.3 Supplement B: QBO metrics for all reanalyses 483
Major abbreviations and terms 488

9.1 Introduction

The dominant mode of variability in the tropical lower-to-mid stratosphere is the quasi-biennial oscillation (QBO), which is approximately zonally symmetric and has a period of roughly 28 months (*Baldwin et al.*, 2001). An extensive review of the main characteristics of the observed QBO and its primary mechanisms is available elsewhere (*Baldwin et al.*, 2001) and is not repeated here. The tropical upper stratosphere, in contrast, is dominated by the annual cycle and semi-annual oscillation (SAO) although there is still a small component of QBO variability extending into the upper stratosphere, mesosphere and beyond. The QBO and SAO oscillations blend together in the mid-stratosphere, between altitudes of about 10 hPa (32 km) to 3 hPa (41 km). Superimposed on these zonal-mean oscillations are a wide variety of tropical waves, with zonal wavelengths ranging from the planetary scale down to tens of kilometers and periods ranging from days to tens of minutes. The zonal-mean wind oscillations act to filter the vertically propagating waves and in turn are driven by them. Co-existing with these atmospheric internal modes of variability, other signals arising from external influences are also discernible, including the El Niño-Southern Oscillation (ENSO), the 11-year solar cycle, and large tropical volcanic eruptions.

All of these phenomena are established features of the tropical stratosphere, yet their precise details vary among reanalyses, sometimes considerably. This chapter focuses on tropical stratospheric variability in the height region from the tropical tropopause (which is examined in *Chapter 8*) up to the mid-stratosphere, the region dominated by the QBO. It examines the representation of the QBO in the various reanalyses, details of the waves that drive it and also some of the remote impacts of the QBO, including coupling with the winter stratospheric polar vortices and impacts on the underlying troposphere. Representation of variability in the upper stratosphere / lower mesosphere (USLM) is examined in *Chapter 11*. We note that these two regions of the tropical stratosphere are dynamically linked because the QBO filters the waves that drive the SAO and, conversely, the SAO may affect the QBO's partial synchronization with the seasonal cycle.

Reanalyses are observationally constrained estimates of the true state of the atmosphere. Uncertainty in these estimates occurs for many reasons and can be difficult to attribute. *Randel et al.* (2004) documented very large differences in the representation of the QBO among different stratospheric analyses, including two reanalysis datasets. Since that study, reanalyses have improved and the differences in QBO representation among modern reanalyses are typically much smaller (*e.g.*, *Kawatani et al.*, 2016). It is of interest to try to understand why the reanalyses have converged toward improved QBO representation. Improvements in the underlying forecast models may be a factor, but the fact that almost none of the forecast models are able to self-generate a QBO [1] implies that model improvement cannot be the only (or even the prime) reason. Improvements in the assimilation methodology and in the type and quantity of assimilated data are very likely the main reason.

Notwithstanding these clear improvements over time, it is important to regularly assess how well the reanalyses capture the atmospheric variability, by comparing with independent observations where these are available, and to examine where there are still remaining differences between the reanalyses. This is particularly important for modellers who rely on the reanalyses and their diagnostics as a guide to the realism of their model simulations. Typical questions include: Is it appropriate to use just one of the reanalysis datasets to compare against, and if so which one is most appropriate? Are there any aspects of a particular reanalysis that should be borne in mind when using it? Is it appropriate to use reanalysis data from the pre-satellite era? Does it matter that the earlier reanalyses have a limited height domain? How much can we trust the QBO signal in fields such as vertical velocity and precipitation that are known to be difficult to represent in the reanalyses? How well do the reanalyses represent the individual wave-types that force the QBO?

In this chapter we attempt to address these questions as they relate to the QBO (*Section 9.2*), stratospheric equatorial waves (*Section 9.3*), and teleconnections of the QBO (*Section 9.4*). A summary of the chapter findings is then presented in *Section 9.5*. In the remainder of *Section 9.1*, *Subsection 9.1.1* provides an overview of aspects of data assimilation that are particularly relevant to the tropics, and discusses the accompanying uncertainties that are most likely to contribute to inter-reanalysis spread in the tropical stratosphere. *Subsection 9.1.2* describes the analysis methods used in *Sections 9.2 - 9.4*.

9.1.1 Issues in representing the QBO

Historically, the most important observations that constrain the QBO in reanalyses are wind profiles from tropical radiosonde stations. Unlike in the extra-tropics, thermal wind balance does not provide as strong a constraint on the wind distribution in the tropics (*Randel et al.*, 2004; *Pawson and Fiorino*, 1998). It is less obvious, therefore, that the global coverage provided by satellite radiance observations from the late 1970s and their substantial increase from around the late 1990s will have such a major impact in the tropics as it does in the extra-tropics. Comparison of pre-satellite and post-satellite data periods can help to identify the impact of assimilating satellite radiances, but this does not entirely preclude the effects of improved radiosonde coverage over time, or of differences in natural variability between the two periods, or secular trends.

[1] MERRA-2 is an exception, as its forecast model does produce a QBO (*Coy et al.*, 2016; *Molod et al.*, 2015)

A more effective approach is to compare reanalyses that use the same data assimilation system but assimilate different observations over the same time period. The JRA-55 and JRA-55C reanalyses are a publicly available pair of reanalysis datasets allowing this comparison: JRA-55 is a "full-input" reanalysis that assimilates all types of atmospheric observations including satellite data, while JRA-55C assimilates only "conventional" observations, which excludes satellite data (*Kobayashi et al.*, 2014). For a more detailed description of the differences between "full-input", "conventional-input" and "surface-input" reanalyses see *Chapter 2* or *Fujiwara et al.* (2017). In this chapter we make extensive use of the JRA-55 vs. JRA-55C comparison to gain insight into the impact of the assimilation of satellite data.

In terms of data assimilation methodology there are some clear differences between the older and more recent reanalyses datasets, such as the change from assimilating derived satellite temperatures to the direct assimilation of the observed radiances, and the transition to assimilation of observations at the actual times they were taken (as in 4-D VAR) rather than at fixed time intervals (see *Chapter 2* or *Fujiwara et al.*, 2017, for more information). However, there are also more subtle differences that can be relevant to the QBO representation. One example arises because of the sparsity of the available wind radiosonde observations. The QBO is generally believed to be zonally symmetric, so that Singapore monthly-mean radiosonde winds can be used as a proxy for the zonal-mean winds. It is difficult to verify this from observations, given the sparsity of tropical radiosonde stations, particularly the large gaps over the oceans (*Kawatani et al.*, 2016). These gaps can potentially be filled by the assimilation of satellite radiances (but note the caveats regarding the resulting induced vertical circulation, described below). However, there are various choices to make in the assimilation systems, such as the assigned "weighting" of the different observations. The weighting also includes an effective spatial area over which the assimilated observation has influence. Given the sparsity of equatorial radiosonde stations in the zonal direction, this radius of influence might usefully be set relatively large in the zonal direction but a correspondingly large radius of influence in the meridional direction would not be advantageous. Additionally, the fact that reanalyses are strongly anchored by Singapore observations (due to Singapore providing very frequent and high reaching observations – see **Figure 9.11**) and less strongly by the other radiosonde stations (that in general provide a lower quality of observations than Singapore) might adversely affect the longitudinal nature of the QBO in the different reanalyses. The treatment of all these issues are likely to differ between the reanalyses.

The influence of radiosonde winds versus satellite radiance also raises an interesting issue in terms of the QBO-induced mean meridional circulation. The underlying mechanism of the QBO is associated with momentum transfer from waves to the background flow. It is therefore more appropriate to force the modelled winds (effectively via the momentum equation) and allow the temperature distribution to respond, rather than to force the temperatures directly (via the thermodynamic equation). For example, it is well known that a descending westerly QBO phase has an associated warm temperature anomaly in the region below the westerly maximum, due to adiabatic heating by the QBO-induced circulation that acts to maintain thermal wind balance (*Plumb and Bell*, 1982). The induced circulation in this case is a region of anomalous descent over the equator (with anomalous ascent in the subtropics forming the return arm of the induced circulation). However, introducing this QBO temperature anomaly by assimilating radiance observations is effectively a diabatic process that will induce anomalous upward vertical motion above the temperature anomaly (*Politowicz and Hitchman*, 1997). The induced circulation in the latter approach therefore has the incorrect sign.

While many of the forecast models used for data assimilation are now run at sufficiently high resolution to resolve the larger scale tropical waves that contribute to the QBO (such as Kelvin waves and mixed Rossby-gravity waves) they are nevertheless unable to resolve small-scale wave contributions from gravity waves. Several of the reanalysis models – CFSv2, ERA-20C, MERRA, and MERRA-2 – include a non-orographic gravity wave drag (GWD) parametrization to emulate the impacts of these waves, but this does not guarantee the generation of a realistic QBO. The non-orographic GWD parametrization used in the CFSv2 reanalysis only represents waves with horizontal phase speeds of zero and hence cannot force a QBO. A QBO-like oscillation is produced in the ERA-20C reanalysis, but since ERA-20C is a surface-input reanalyses, no observations are assimilated in the tropical stratosphere and the oscillation is not in phase with the real QBO. The MERRA and MERRA-2 reanalyses both include parametrized GWD, but the GWD in MERRA was not tuned to produce a QBO. Thus the majority of reanalyses examined here – all of them except MERRA-2 – rely on the assimilation of the observations to generate a QBO in the reanalysis. Hence the role of data assimilation in these cases is not to simply correct the trajectory of the atmospheric state toward the trajectory of the observed atmosphere (that is, to keep the QBO in the reanalysis in phase with the real QBO), but rather to compensate for a significant model bias, the absence of a QBO in the model. Whether this really matters is an open question. It is likely to be important in the context of seasonal and decadal-scale forecasting, where the model is initialized to the observed state but then continues in its free-running state so that it will likely drift substantially over time. However, it is less clear whether the absence or otherwise of an internally generated QBO is important for reanalyses, since the analysis increments are small (6 hours or less) compared to the relatively long radiative timescales (of order months) in the lower-to-mid stratosphere. (We also note that the lack of an adequate internally-generated SAO/QBO is likely to be much more important in the upper stratosphere, where radiative timescales are much shorter; of order minutes).

Comparison of MERRA-2 with the other reanalysis datasets may shed some light on this (see *e.g.,* the discussion in *Coy et al*., 2016) although there are other simultaneous improvements to the model that preclude a definitive answer to this question.

Although the basic QBO mechanism is well understood, the precise balance of wave types that contribute to forcing the QBO is not yet clear. Reanalyses can provide some degree of benchmark for the broad spectrum of equatorial waves, to the extent that these waves are constrained by the data assimilation. For example, large-scale Kelvin waves have a temperature signal that should be well constrained by assimilated satellite radiance observations. However, it should be cautioned that the wave activity in reanalyses may also be significantly influenced by forecast model characteristics that are less constrained by observations, such as tropical precipitation which can vary substantially among reanalyses (*Kim et al*., 2014; *Kim and Alexander*, 2013; *Pfeifroth et al*., 2013). Gravity waves make a large contribution to forcing the QBO (*Alexander et al*., 2010; *Baldwin et al*., 2001) and it is not clear whether reanalyses can provide meaningful constraints on these waves since they might be largely determined by the forecast model. The "background state" of tropical upper tropospheric winds, through which upward-propagating waves travel en route to the QBO region, may also suffer from weak observational constraint due to the aforementioned sparseness of the tropical radiosonde network.

As well as differences in the underlying model dynamics and the choice of assimilation set-up, there are further differences that can potentially influence the representation of the QBO. JRA-55 is a particular case. As described in *Chapter 4*, the ozone field used in JRA-55 radiative calculations from 1979 onward is generated by a chemistry-climate model, MRI-CCM1, that assimilates total column ozone observations. The version of MRI-CCM1 used to generate the JRA-55 ozone is the QBO-resolving version of the model documented in *Shibata and Deushi* (2005). The winds in this MRI-CCM1 version are nudged to JRA-25 winds, so that the wind QBO will be approximately be in phase with the observed QBO. This is likely the reason why the JRA-55 ozone QBO compares favourably with observations, as shown in *Chapter 4* (**Figure 4.11**). Hence the radiative heating rates due to ozone in JRA-55 are likely to contain a realistic QBO signal and will therefore contribute to the characteristics of the QBO simulated by this reanalysis.

The following sections of the chapter (*9.2–9.4*) attempt to address many of these questions, and *Section 9.5* provides a summary of the results and recommendations.

9.1.2 Data and methods

This section describes or provides references for the various analysis methods used in *Sections 9.2 - 9.4*.

Zonal-mean data (monthly-mean and daily-mean) were obtained from the S-RIP common-gridded dataset at 2.5° resolution in latitude and longitude on a standard set of pressure levels as prepared by *Martineau* (2017).

While we note the presence of jumps in the reanalysis fields due to the introduction of additional satellite data and the use of parallel processing streams (*e.g., Long et al*., 2017, *Chapter 3*), these are primarily evident in the temperature fields and are much less evident in the zonal wind fields.

The QBO index used in the multiple linear regression (MLR) analysis was derived from radiosonde observations issued by Freie Universität Berlin (FUB; *Naujokat*, 1986) [2]. The FUB data are a combination from three different stations: Canton Island (3 °S/172 °W, January 1953 to August 1967), Gan / Maldive Islands (1 °S/73 °E, September 1967 to December 1975) and Singapore (1 °N/104 °E, since January 1976). The merged data are provided as monthly averages interpolated on the 70, 50, 40, 30, 20, 15 and 10 hPa levels (see also *Fujiwara et al*., 2020). The QBO index was calculated from the FUB data rather than using the reanalysis equatorial winds so that it characterised the observed QBO as closely as possible, thus avoiding any possible degradation by the data assimilation process (*Kawatani et al*., 2016). This also ensures that the same index was used across all reanalyses. Further information on the QBO index used for any given analysis (*e.g.*, the vertical level used) is provided in the relevant part of the chapter.

Wave spectra in *Section 9.3.1* are calculated following the method of *Kim et al.* (2019). Details of this method are also provided in *Section 8.6.2* of this report. Wave forcing of the QBO in *Section 9.3.2* was partitioned into different wave types using the method of *Kim and Chun* (2015). Details of the comparison between satellite observations and reanalyses in *Section 9.3.3* are given by *Wright and Hindley* (2018).

Precipitation fields from the Global Precipitation Climatology Centre (GPCC) dataset for the period 1979 - 2016 were also employed [3]. The monthly-averaged global precipitation observations at 1 ° latitude-longitude resolution (*Schneider et al*., 2014) were obtained from NOAA[4]. The composite difference analysis of the GPCC precipitation data was performed in a similar manner to that described in *Liess and Geller* (2012).

[2] Available at **https://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/index.html**

[3] Available at **https://www.dwd.de/EN/ourservices/gpcc/gpcc.html**

[4] Available at **https://psl.noaa.gov/data/gridded/data.gpcc.html**

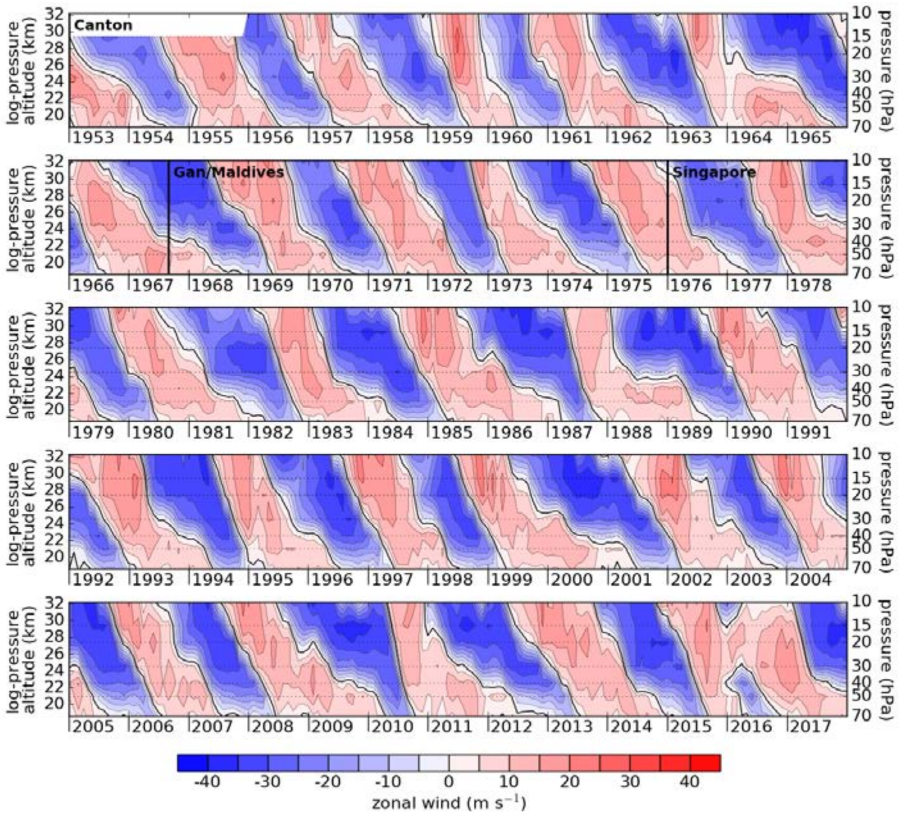


Figure 9.1: *Monthly-mean zonal winds (m/s) from the radiosonde observation dataset produced by Freie Universität Berlin (FUB), updated from Naujokat (1986). Text labels and vertical black lines indicate the station locations used for the different parts of the FUB record: Canton (2.8 °S, 171.7 °W), Gan/Maldives (0.7 °S, 73.2 °E), and Singapore (1.4 °N, 103.9 °E).*

Composites were compiled from those months where the equatorial wind at the selected pressure level (see *Section 9.4.4*) was at least 3 m s^{-1} above or below the corresponding monthly average.

The MLR analysis used in *Section 9.4* is as described in *Crooks and Gray* (2005) and *Gray et al.* (2013, 2016): the time-series at each grid-point is fitted using a number of indices (timeseries) that characterise the observed variability associated with (1) volcanic eruptions, (2) El Niño Southern Oscillation (ENSO); (3) solar radiative forcing, (4) the QBO and (5) a long-term trend. Details of the height and month(s) of the equatorial zonal winds used to characterise the QBO in the composite and MLR analyses are described in the relevant text where necessary. ENSO variability is characterised by a time-series of averaged sea surface temperatures from the Niño 3.4 region (120 °- 170 °W, 5 °N - 5 °S) using monthly averaged data on a 1 ° spatial grid from the Hadley Centre HadISST dataset (*Rayner et al.*, 2003)[5]. The volcanic eruption index used the updated GISS Sato Index (*Sato et al.*, 1993) extended to 2016[6]. The 11-yr solar cycle index was derived from an updated version of the NRLSSI time-series of total solar irradiance (*Wang et al.*, 2005)[7]. A simple linear trend is used for the long-term trend index. An autoregressive noise model (AR-1) was included and a Student's t-test was employed to determine the probability that the regression coefficients are significantly different from the noise (in all figures, light grey and white stippled regions denote statistical significance at the 95 %/99 % level). The regression coefficients have been re-scaled to show the typical maximum response e.g. between opposite QBO/ENSO phases or between periods of solar max-min conditions. For further details see *Mitchell et al.* (2015) and *Gray et al.* (2018).

Definitions of acronyms used in the chapter are collected for convenienceat the end of the chapter. Unless noted otherwise, "CFSR" refers to concatenation of the CFSR with CFSv2 from January 2011 onward. In some cases the NCEP-NCAR R1 and NCEP-DOE R2 reanalyses are alternately referred to as *R1* and *R2* (or NCEP R1 and NCEP R2), respectively, consistent with common usage.

9.2 Monthly-mean equatorial variability

9.2.1 Time evolution

The QBO is often characterized using the monthly-mean zonal wind radiosonde dataset issued by FUB (*Naujokat*, 1986) made up of consecutive records from three near-equatorial stations (Canton, Gan/Maldives, Singapore). The Singapore station is particularly valuable because it has been reliably observing since at least 1976 (**Figure 9.1**). It is often assumed that the FUB winds provide a good estimation of the zonal-mean zonal wind since performing the monthly mean effectively averages out any wave motions and because the QBO is generally understood to be a zonally symmetric phenomenon (*Lindzen and Holton* 1968). Verifying this assumption is difficult because radiosonde stations are so sparsely distributed in the tropics.

[5] Available at **https://psl.noaa.gov/gcos_wgsp/Timeseries/Nino34/**

[6] Available at **https://data.giss.nasa.gov/modelforce/strataer**

[7] Available at **http://solarisheppa.geomar.de/solarisheppa/cmip5**

Figure 9.2, reproduced from *Kawatani et al.* (2016), shows a comparison of the time series of monthly-mean zonal winds at different levels in the stratosphere from 9 full-input reanalysis datasets extracted at the location of Singapore, along with the observed FUB Singapore radiosonde winds for comparison. Overall there is broad agreement between the reanalyses and the FUB winds on the phase and amplitude of the QBO. This agreement is unsurprising because the reanalyses assimilate radiosonde winds, which provide a strong constraint in the tropics. Differences between the reanalyses and FUB winds tend to be larger at the lowest altitude shown, 70 hPa, consistent with the flow being less zonally symmetric at these altitudes and hence less well constrained by the sparse distribution of tropical radiosonde observations. Differences are also large at the highest altitude shown, 10 hPa. This is the highest altitude that radiosondes reach, but not all soundings achieve this level and in general the number of available sonde observations decreases with increasing altitude (*Kawatani et al.*, 2016). The MERRA-2 reanalysis is a clear outlier at the upper levels in the period before the mid-1990s, which may be associated with the downward propagation of very strong westerly SAO phases in this reanalysis (*Coy et al.*, 2016); see *Chapter 11* of this report for evaluation of reanalyses in the upper stratosphere. **Figure 9.2** also shows a clear reduction in the inter-reanalysis spread with time, consistent with the increasing number of available observations (*Kawatani et al.*, 2016).

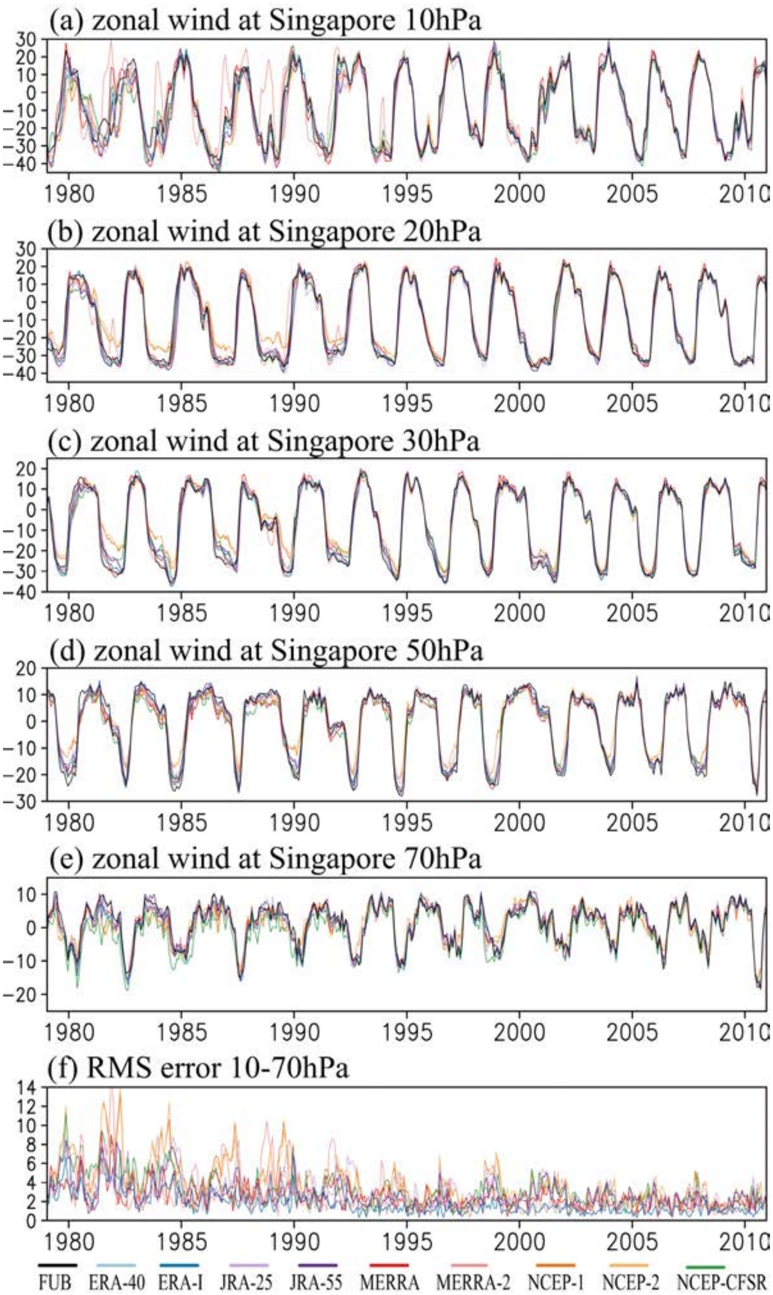


Figure 9.2: *Time variation of monthly-mean zonal wind (m s^{-1}) over Singapore from the reanalyses and from the FUB Singapore radiosonde observations, at the indicated altitudes. The RMS difference of each reanalysis from the FUB radiosonde zonal wind, averaged over the 70 - 10 hPa layer, are also shown (bottom panel). From Kawatani et al. (2016).*

Turning now to examination of the zonal-mean flow, **Figure 9.3** shows the reanalysis ensemble mean (REM) of monthly-mean, zonal-mean zonal wind averaged over 2 °S - 2 °N for the four modern full-input reanalyses: ERA-Interim, JRA-55, MERRA-2 and CFSR for the 1980 - 2016 period. Although there are some differences in the timing of QBO phase onsets between the reanalyses, discussed in more detail below, the good overall agreement between the reanalyses indicated by **Figure 9.2** and similarities between **Figure 9.3** and **Figure 9.1** suggests that this REM is a suitable "best estimate" of the zonally-averaged QBO state. These four reanalyses were selected because they are the most recent full-input reanalyses and because they are available up to approximately the present day. The 1980 - 2016 period is chosen because this is their common period (MERRA-2 begins in 1980) and it is also the period over which the S-RIP common-gridded zonal-mean dataset and diagnostics are available (*Martineau*, 2017).

Figure 9.3 illustrates the basic features of the zonally-averaged QBO. It is similar to **Figure 9.1** (FUB winds at Singapore), but with smoother variations because fluctuations have been averaged out by taking both the zonal mean and the mean over reanalyses. (*Appendix A9.2* shows the corresponding time series for the individual reanalyses over the whole time period spanned by each reanalysis.)

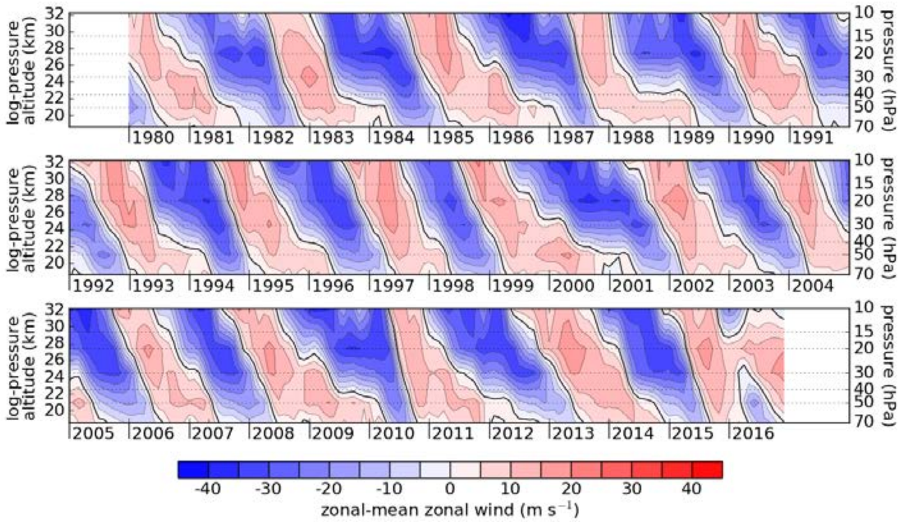


***Figure 9.3:** Time-series of monthly-mean equatorial (2°S-2°N average) zonal-mean zonal wind (m s^{-1}) for 1980-2016 from the reanalysis ensemble mean (REM) of the ERA-Interim, JRA-55, MERRA-2 and CFSR datasets.*

Westerly and easterly winds descend in succession, with the duration of the individual phases being somewhat variable. At the lowest levels the westerly QBO phases (QBO-W) tend to persist longer than the easterly phases (QBO-E) while the opposite is true at the higher levels. The QBO-W phases have roughly constant amplitude with height while QBO-E phases tend to strengthen with increasing height. These are all expected features of the observed QBO (*Baldwin et al.*, 2001). The 2015-2016 disruption of the QBO (*Coy et al.*, 2017; *Newman et al.*, 2016; *Osprey et al.*, 2016) is seen at the end of the record as the intrusion near 40 hPa and subsequent descent of an easterly anomaly within a QBO-W phase.

All of the reanalyses agree on these general features of the QBO, although the older NCEP-NCAR R1 and NCEP-DOE R2 have a somewhat poorer representation, with a much weaker QBO amplitude (*Appendix A9.2*: **Figures AA9.9** and **AA9.10**). Nevertheless, even those older reanalysis systems display an essentially correct qualitative representation of the QBO during recent decades, which is presumably due to the constraint provided by radiosonde wind observations as well as the fact that QBO phase transition times are short compared to the duration of QBO phase, as **Figure 9.2** shows.

Note, however, that NCEP-NCAR R1's representation becomes poorer further back in the record, *e.g.*, during the 1953-1978 period (**Figure AA9.9**; *cf.*, FUB winds in **Figure 9.1**). This degradation is presumably due to the fact that fewer radiosonde observations were available earlier in the record, which affects all of the reanalyses but is particularly marked in the case of the NCEP-NCAR R1 reanalysis (*Kalnay et al.*, 1996). The CFSR reanalysis also suffers some degradation in the early part of its record (**Figure AA9.8**; *Saha et al.*, 2010).

The degree of disagreement between the reanalyses is quantified in **Figure 9.4** by the inter-reanalysis standard deviation (SD; *Kawatani et al.*, 2016) for the four REM datasets used to produce **Figure 9.3**. Similar to the differences between reanalyses seen in the Singapore zonal wind (**Fig. 9.2**) the inter-reanalysis spread of zonal-mean zonal wind tends to be larger earlier in the record and at higher altitudes (and, to a lesser degree, there is some increased spread at lower altitudes present earlier in the record). **Figure 9.4** also shows that inter-reanalysis spread at all altitudes tends to be greatest during QBO phase transitions, more so for QBO-W onsets than for QBO-E onsets (as illustrated by the superimposed zero-wind line contours in **Figure 9.4**; refer to **Figure 9.3** to see which contours correspond to QBO-W and QBO-E onsets). The inter-reanalysis SD including the other reanalyses is shown in **Figure AS9.1** and displays the same basic

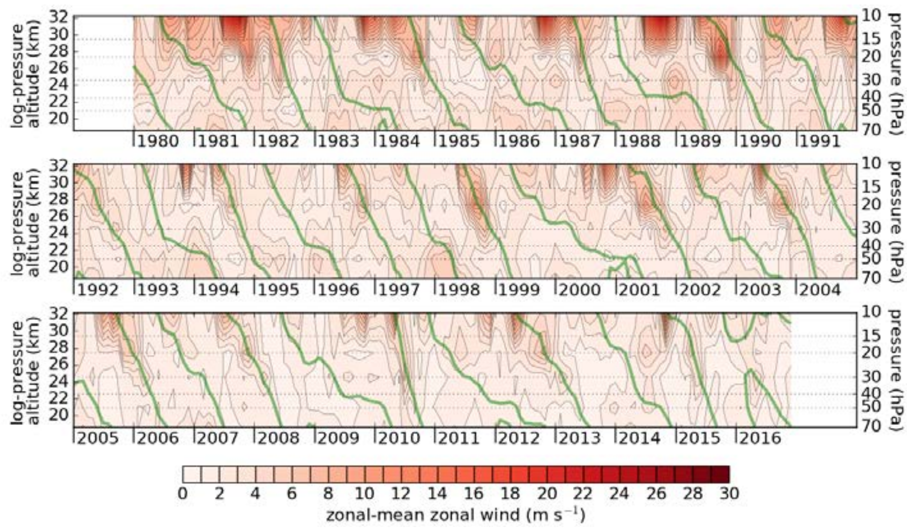


***Figure 9.4:** Inter-reanalysis standard deviation (SD) of monthly-mean 2°S–2°N zonal-mean zonal wind for ERA-Interim, JRA-55, MERRA-2 and CFSR. Thick green contours show the zero-wind line of the REM (**Figure 9.3**).*

features as seen in **Figure 9.4**. As noted earlier, reasonably good agreement between QBO winds in different reanalyses is expected due to the strong constraint provided by assimilation of radiosonde winds. Since the assimilation of satellite data has generally led to improvements in the quality of reanalyses (as documented elsewhere in this report) it is of interest to assess its impact on the QBO. **Figure 9.5** shows the inter-reanalysis SD for just two reanalyses, JRA-55 and JRA-55C, which as noted in *Section 9.1* are identical except that JRA-55C does not assimilate any satellite data. Small SD values in **Figure 9.5** (note that the contour spacing in **Figure 9.5** is half that in **Figure 9.4**) indicate that the removal of satellite data has little effect on the representation of the QBO in the JRA- 55/55C reanalysis system, consistent with the expectation that assimilation of radiosonde winds provides a strong constraint on the QBO.

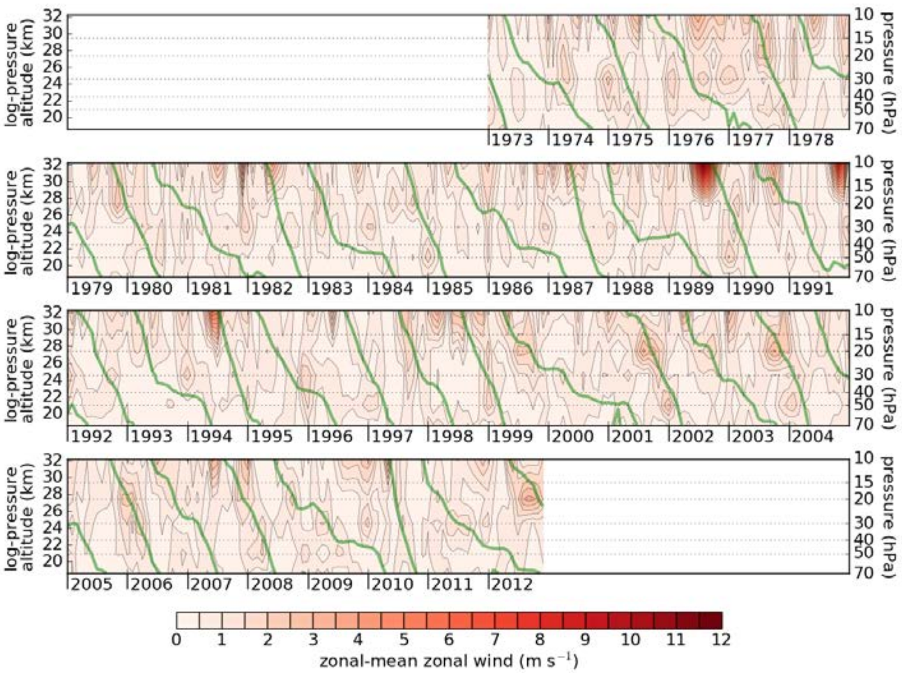


Figure 9.5: *Inter-reanalysis standard deviation (SD) of monthly-mean 2 °S - 2 °N zonal-mean zonal wind for the common year range 1973 - 2012 of JRA-55 and JRA-55C. Thick green contours show the zero-wind line of the REM for these two reanalyses. Note that the contour values are half as large as in **Figure 9.4**.*

Figures 9.6 and **9.7** show the corresponding analysis of the deseasonalized zonal-mean temperatures. In **Figure 9.6** the superimposed zero-wind line contours highlight, as expected, that temperatures are anomalously warm in descending QBO-W phases and anomalously cold in descending QBO-E phases, consistent with thermal wind balance and the QBO-induced mean-meridional circulation (*Plumb and Bell*, 1982). The temperature anomalies reach amplitudes of ≈ 2 - 4 K during each descending phase (*Baldwin et al.*, 2001). **Figure 9.7** shows that the inter-reanalysis SD of deseasonalized zonal-mean temperature tends to be larger at higher altitudes and earlier in the record, similar to the wind SD shown in **Figure 9.4**. A tendency for greater spread during or just below the QBO-W descents is also seen, and there is less indication of a corresponding increase during QBO-E descents. Even for some relatively recent QBO phases there is greater inter-reanalysis spread during some QBO-W descents (*e.g.*, during 2012 - 2013 and 2015).

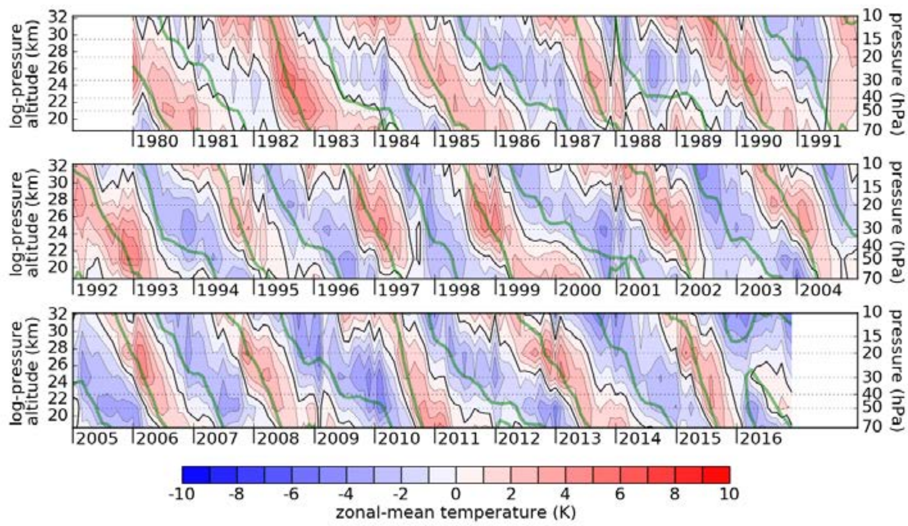


Figure 9.6: *REM as in **Figure 9.3**, but for deseasonalized monthly-mean 2 °S - 2 °N zonal-mean temperature. Thick green contours show the zero-wind line of the REM (**Figure 9.3**).*

It was noted in the Introduction that the inter-reanalysis spread is expected to be larger in the tropics than extra-tropics due to the weaker constraint provided by satellite-derived temperature observations. **Figure 9.8**, after Figure 3 of *Kawatani et al.* (2016), illustrates the variation in both latitude and altitude of the inter-reanalysis spread for various combinations of reanalyses and for different time periods (these distributions were derived by taking the time-mean of inter-reanalysis SD time series like that shown in **Figure 9.4** for the indicated combinations of reanalyses and time periods.)

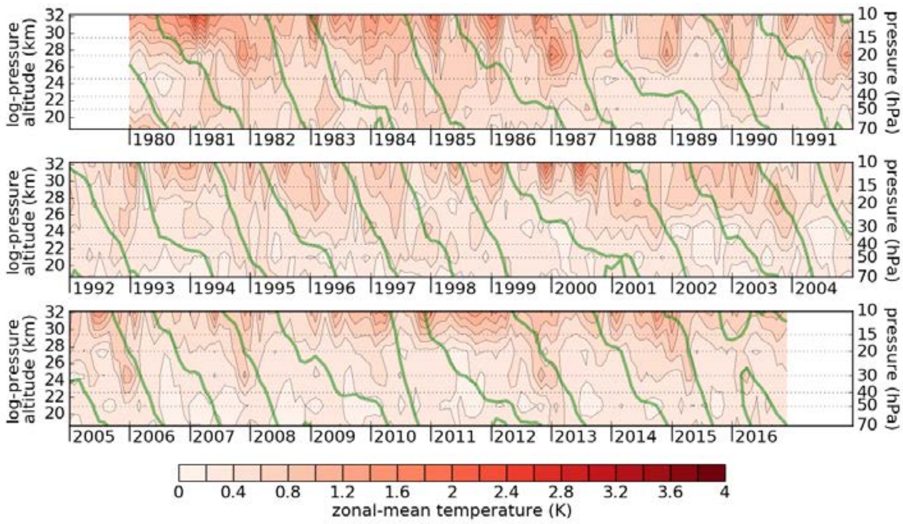


Figure 9.7: *Inter-reanalysis SD as in **Figure 9.4**, but for deseasonalized monthly-mean 2 °S - 2 °N zonal-mean temperature. Thick green contours show the zero-wind line of the zonal wind REM (**Figure 9.3**).*

higher altitudes is consistent with fewer radiosonde observations being available to constrain the model, despite the flow being more zonally uniform at these altitudes.

Figure 9.8 also shows that below the near-tropopause maximum, the inter-reanalysis spread in zonal-mean zonal wind is reduced. Below ≈ 70 hPa there is significant zonal asymmetry in the background tropical circulation, as shown in **Figure 9.9a**, and this asymmetry also weakens below the 150 hPa level. This corresponds to the zonal wind maximum of the upper branch of the Walker circulation which is located just below the near-tropopause maximum in **Figure 9.8**. **Figure 9.9b** shows the inter-reanalysis spread of the background zonally varying zonal wind and indicates that the spread in the background flow tends to be largest where there are fewest or no radiosonde observations, namely in the central Pacific (dateline to ≈ 60 °W), tropical Atlantic (≈ 40 °W - 10 °E), and Indian (≈ 50 °E - 90 °E) oceans, and that the spread is large in the vicinity of the Walker cell upper branch.

Two equatorial maxima are evident in all panels: a smaller one just below the tropical tropopause (100 - 150 hPa) and a much larger one at higher altitudes (≈ 10 hPa). The near-tropopause maximum is consistent with zonal asymmetries of the flow being larger in the lowermost stratosphere / upper troposphere than at higher altitudes, and consequently the sparse distribution of tropical radiosondes provides less adequate spatial sampling than at higher altitudes. The maximum at the

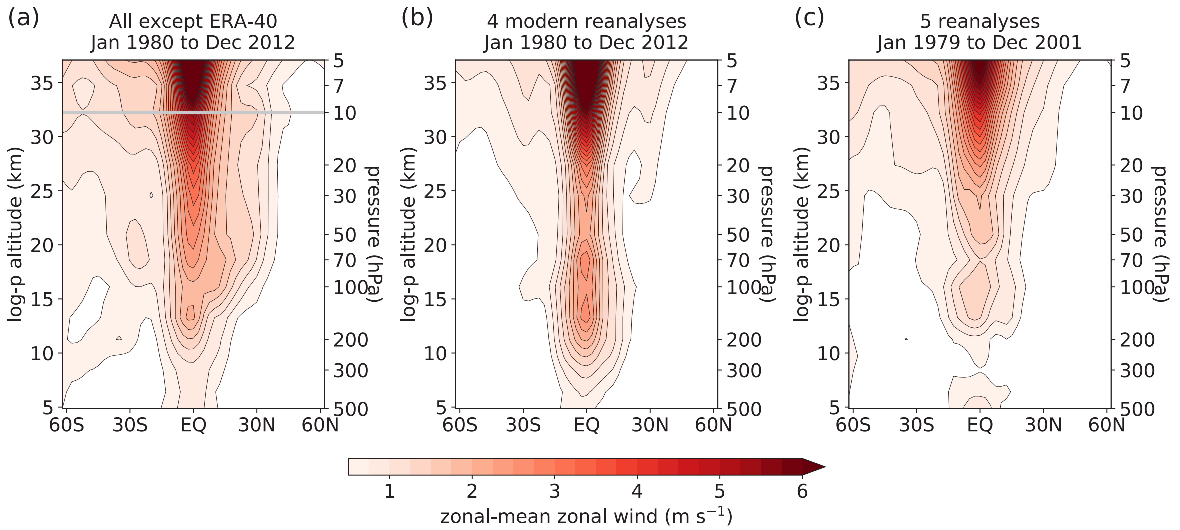


Figure 9.8: *Latitude-height cross sections of the inter-reanalysis standard deviation (SD) of zonal-mean monthly-mean zonal wind. The inter-reanalysis SD is calculated as a function of time and then averaged over the time period indicated in each panel title. (a) SD among all reanalyses (ERA-Interim, JRA-25, JRA-55, JRA-55C, MERRA, MERRA-2, CFSR, NCEP-NCAR, NCEP-DOE) except for ERA-40, which is excluded because it does not cover the whole 1980–2012 time period. Above 10 hPa (horizontal grey line) the NCEP-NCAR and NCEP-DOE reanalyses are excluded because they provide no data at these altitudes. (b) SD among the four modern reanalyses ERA-Interim, CFSR, JRA-55 and MERRA-2. (c) SD among the five reanalyses ERA-40, ERA-Interim, JRA-25, JRA-55, and MERRA, as shown in Figure 3a of Kawatani et al. (2016). These are the same five renalyses as used in **Figures 9.9, 9.10, 9.11**. Contour spacing is 0.25 m s^{-1}, starting at 0.5 m s^{-1}.*

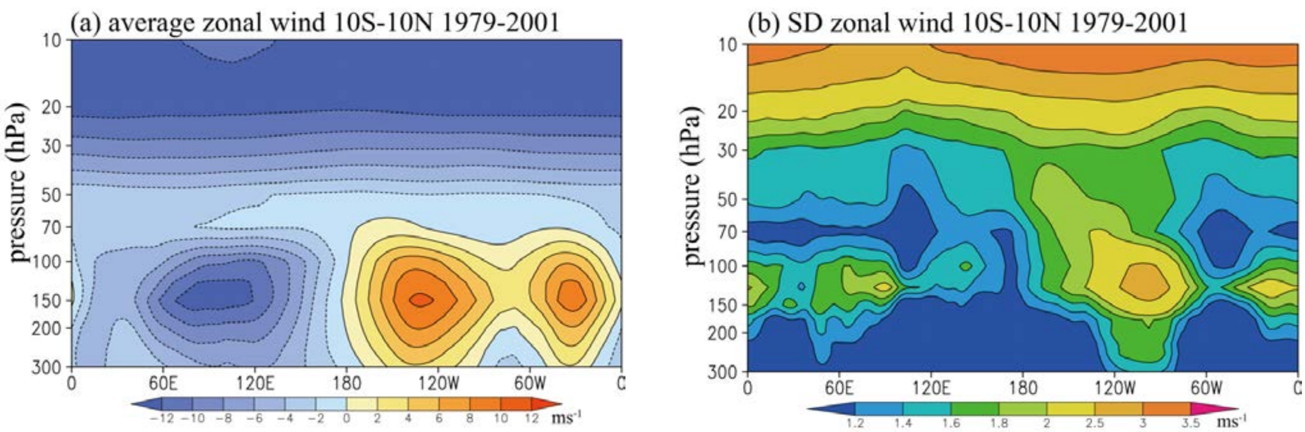


Figure 9.9: *Longitude-altitude cross section of (a) zonal wind averaged over five reanalyses (ERA-40, ERA-Interim, JRA-25, JRA-55, and MERRA), and (b) the inter-reanalysis standard deviation among these five reanalyses. Average over 10 °N - 10 °S, 1979 - 2001. The color intervals are (a) 2 m s^{-1} and (b) 0.2 m s^{-1} with values less than 2 m s^{-1} and 0.5 m s^{-1} with values more than 2 m s^{-1}. From Kawatani et al. (2016).*

This has implications for modelling the evolution of tropical waves as they propagate vertically and interact with the background zonal flow, because of this relatively large uncertainty in the background flow in this crucial region of the upper troposphere. In free-running atmospheric general circulation models (AGCMs), non-orographic GWD parametrizations are often tuned to improve model performance in the stratosphere and mesosphere, including the QBO. Such tuning likely compensates to some degree for model errors in upper tropospheric resolved winds.

A time-evolving view of the inter-reanalysis mean and spread of background zonal-mean zonal wind and deseasonalized temperature over an expanded altitude range from the surface to stratopause is shown in **Figures AS9.2 - AS9.5** (the equivalent of **Figures 9.3**, **9.4**, **9.6**, and **9.7**, but for the 1000 - 1 hPa altitude range). They confirm the vertical variations seen in **Figure 9.8**, and show in more detail how differences between the reanalyses have decreased over time. Prominent decreases are seen in both of the regions that have large inter-reanalysis spread: the near-tropopause (upper Walker cell) region, and the upper stratosphere. The spread increases rapidly above 10 hPa (≈ 32 km), as would be expected due to lack of radiosonde observations above ≈ 10 hPa, meaning that no direct wind observations are assimilated at these higher altitudes.

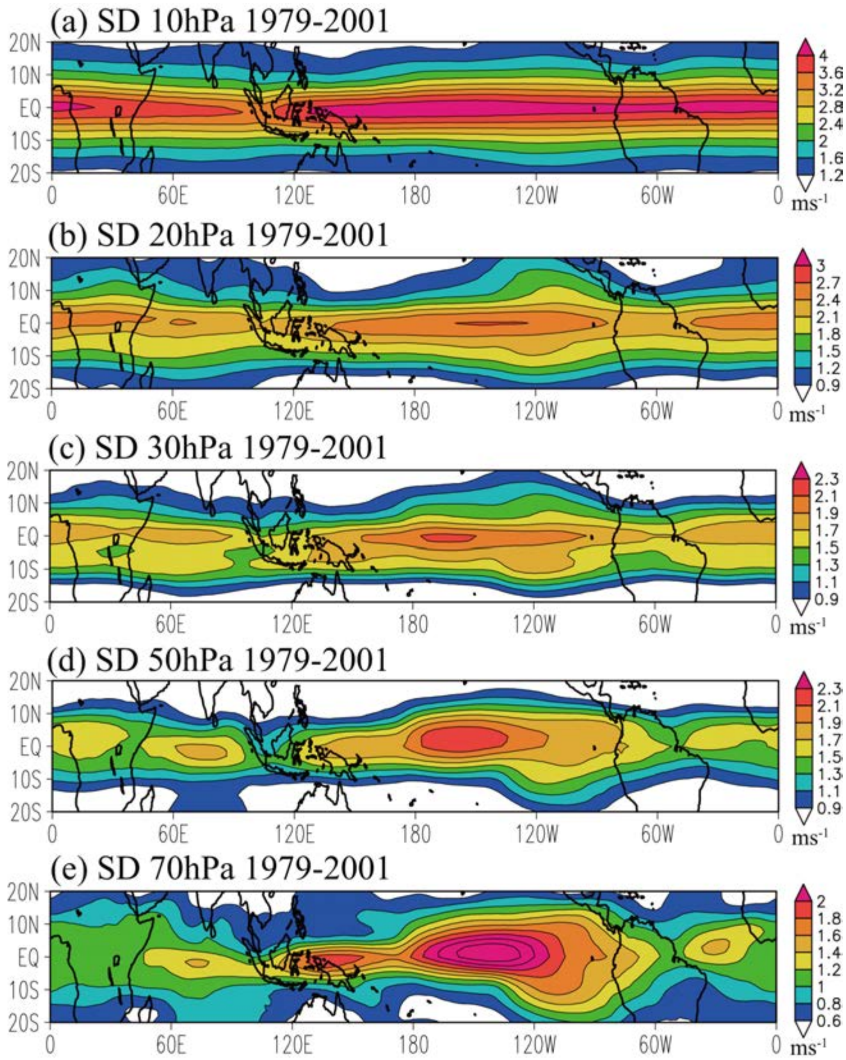


Figure 9.10: *Latitude-longitude distributions of the inter-reanalysis standard deviation (SD) at the indicated altitudes, averaged over the 1979 - 2001 period. The SD is for the reanalyes ERA-40, ERA-Interim, JRA-25, JRA-55, and MERRA. Note that the colour ranges are different among these heights, and the colour intervals are 0.4 m s^{-1} for (a), 0.3 m s^{-1} for (b), and 0.2 m s^{-1} for (c–e). From Kawatani et al. (2016).*

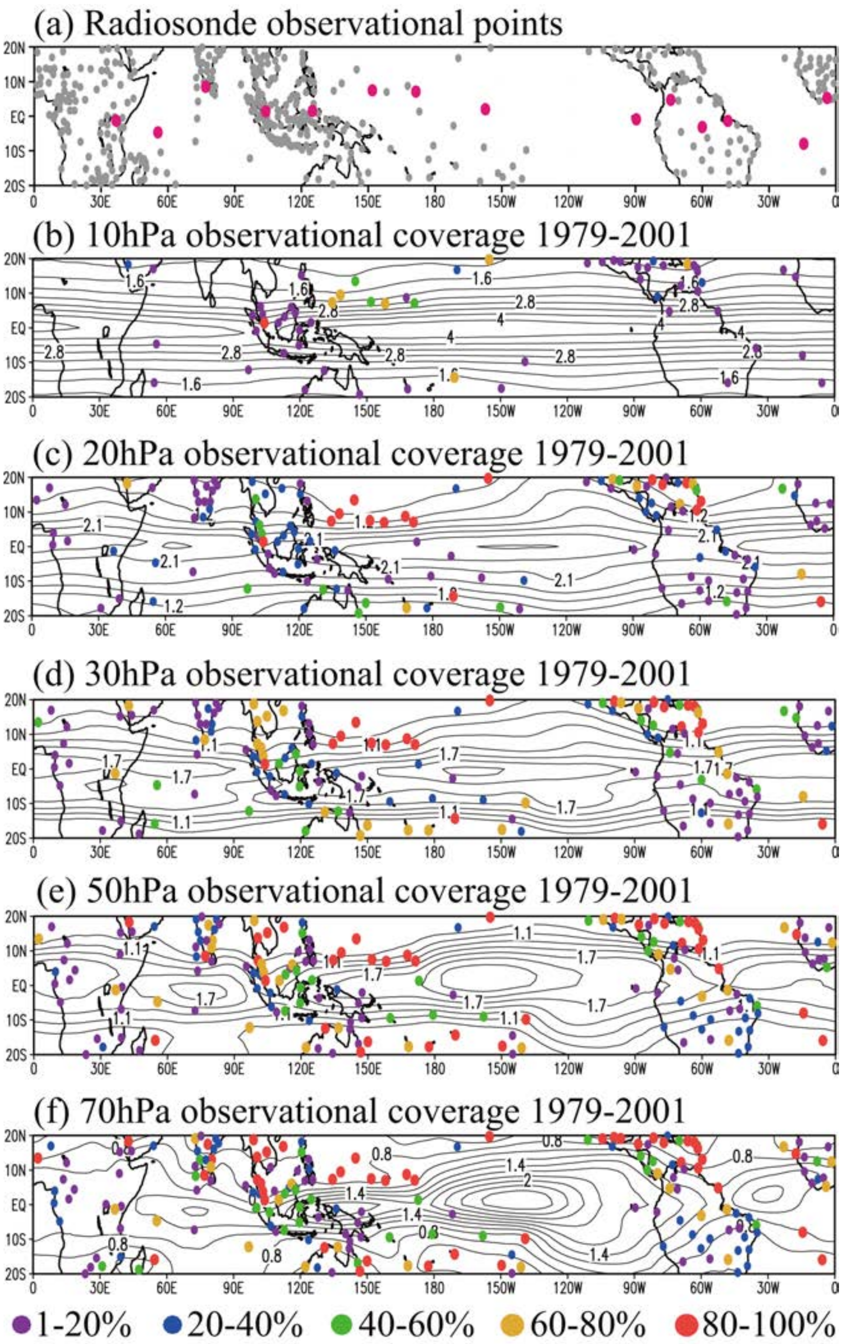


***Figure 9.11:** (a) Locations of all IGRA stations in the tropical region; magenta dots indicate the locations of the stations discussed in further detail in Kawatani et al. (2016). (b - f) IGRA stations with data coverage of (purple) 1 - 20%, (blue) 20 - 40 %, (green) 40 - 60 %, (yellow) 60 - 80 % and (red) 80 - 100 % at the indicated altitudes. Line contours show the standard deviation among reanalyses as shown in **Figure 9.10**. From Kawatani et al. (2016).*

The inter-reanalysis SD exhibits some semi-annual periodicity (see also *Chapter 11*), and is particularly large prior to and during the onset of the westerly SAO (SAO-W) phase, and during the downward descent of SAO westerlies through the 3 - 10 hPa (≈ 30 - 40 km) layer, including those instances when the SAO-W and QBO-W phases join to form a continuous descent of westerly winds throughout the entire tropical stratosphere. This is reminiscent of the behaviour seen in **Figure 9.4** for the QBO-W phase in the 10 - 70 hPa layer.

The zonal distribution of inter-reanalysis spread is further examined in **Figure 9.10**, which shows latitude-longitude maps of the inter-reanalysis SD in the zonal winds at the 10, 20, 30, 50 and 70 hPa levels, time-averaged over the 1979 - 2001 period, reproduced from *Kawatani et al.* (2016). At 70 hPa the largest SD is found over the oceans, and a major source of the near-tropopause spread in zonal-mean zonal wind (**Figure 9.8**) is seen to be in the central Pacific (as was also seen in **Figure 9.9b**). At higher altitudes the SD becomes more zonally symmetric, becoming almost zonally uniform at 10 hPa. The relationship of these patterns to the spatial distribution of radiosonde stations is shown in **Figure 9.11**. Here the station locations (dots) and the observational coverage at each level (percentage of times reporting, colour) are shown superimposed on the inter-reanalysis SD patterns from **Figure 9.10** (line contours). At the lower levels (50, 70 hPa) the large SD is clearly linked to regions of sparse or nonexistent coverage. At the upper levels (10, 20 hPa) the spatial pattern of SD bears little relation to the location of radiosonde stations. This same qualitative spatial pattern of SD is seen throughout the reanalysis record, but the magnitude of the SD decreases with time. **Figure 9.12** shows the time-averaged SD in the 50 - 70 hPa layer for the 1979 - 1989, 1990 - 2000 and 2001 - 2011 periods. While the spread remains largest over the tropical oceans, its magnitude in recent years is considerably reduced, consistent with the time evolution seen in **Figure 9.4**. Note that **Figure 9.11**, from *Kawatani et al.* (2016), shows the IGRA sondes, which are not necessarily the same as those assimilated by the reanalysis datasets; nevertheless, since both the reanalysis centres and the IGRA dataset compilers aim to maximise the number of sonde observations there may be good correspondence between the two.

In summary, to a large extent the noted differences between reanalyses have been shown to be consistent with spatial and temporal variations in tropical radiosonde observational coverage. The magnitude of this inter-reanalysis spread is relatively small in comparison with typical magnitudes of QBO winds, at least in the main QBO region of 10 - 70 hPa (**Figure 9.4**). As a result, most of the current reanalyses show reasonable representations of the QBO (*cf.*, **Figure 9.1** and **Figure 9.3**). There is nevertheless some spread in the timing of QBO phase onsets, particularly of the QBO-W phase (**Figure 9.4**).

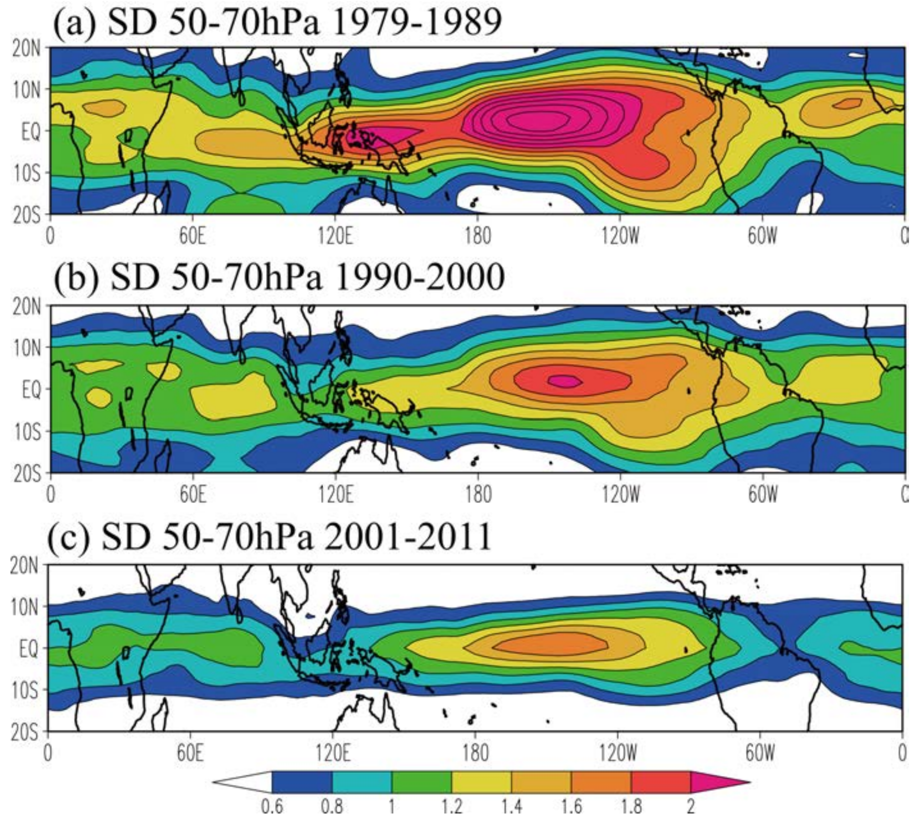


Figure 9.12: *Latitude-longitude distributions of the inter-reanalysis standard deviation (SD), as in* ***Figure 9.10****, but here showing time evolution of the SD of the ERA-Interim, JRA-25, JRA-55 and MERRA reanalyses. Latitude-longitude distributions of the inter-reanalysis SD averaged over the 50 - 70 hPa layer, averaged over the indicated time periods. From Kawatani et al. (2016).*

This uncertainty in timing of the QBO-W onset may be due to the fact that none of the underlying forecast models used in these full-input reanalyses, with the exception of MERRA-2, spontaneously generates a QBO. Equatorial westerlies require wave forcing to generate an equatorial angular momentum maximum (easterlies, in contrast, can be generated by wave forcing but also by cross-equatorial advection, as is believed to contribute strongly to the SAO-E phase). If the source of westerly wave driving is too small in a model, which is consistent with an inability to spontaneously generate a QBO, then the QBO-W onset is likely to be delayed. This can be checked by comparing the timing of the QBO-W onsets in the MERRA-2 reanalysis with the onsets in the reanalyses that are unable to spontaneously generate a QBO. **Figure 9.13** shows the descending zero-wind contours of MERRA-2 (black) and the other three reanalyses (green) that contributed to the 4-reanalysis REM (JRA-55, CFSR, ERA-I). The MERRA-2 westerly onsets tend to occur earlier than westerly onsets in the other three reanalyses. This is particularly clear at the upper levels in the earlier period, for example at 10 - 20 hPa in 1980/81 and 1989/90, but is also evident in the later period *e.g.*, 2012/13 and at lower levels *e.g.*, 1995. A similar effect is less evident in the QBO-E phase onsets although there are some occurrences, for example at upper levels in 1990/91. Comparing MERRA-2 to all nine of the other reanalyses shows the same result (**Figure 9.14**). A similar lag in QBO-W onset is also apparent when the reanalysis winds at Singapore are compared to the FUB winds (*Kawatani et al.*, 2016). This suggests that the relatively large inter-reanalysis spread seen during QBO-W descents (**Figure 9.4**) is likely due to a systematic error shared by almost all of the reanalysis forecast models. (As an aside, we note that if this is the case, then a similar underestimation of the wave forcing responsible for the SAO-W phases, which are also wave-driven, would result in a similar delay in onset and, as suggested in *Section 9.1*, this is likely to be more severe since the radiative timescales are much shorter at these higher altitudes so that the underlying bias in the forecast model is likely to show more quickly than the bias at the lower QBO levels.)

9.2.2 QBO amplitude and phase transitions

The analysis shown in **Figures 9.13** and **9.14** suggests that aspects of the representation of the QBO (and SAO) in reanalyses could be further improved if the representation of wave driving in the underlying forecast model is improved so that it is able to self-generate a realistic QBO (and SAO). There is currently a large degree of uncertainty about what is required to achieve a QBO in freerunning (*i.e.*, run without data assimilation) atmospheric general circulation models (*Butchart et al.*, 2018), and these models exhibit much larger quantitative and qualitative differences in their representation of the QBO than are seen in the different reanalyses (*Bushell et al.*, 2020; *Schenzinger et al.*, 2017).

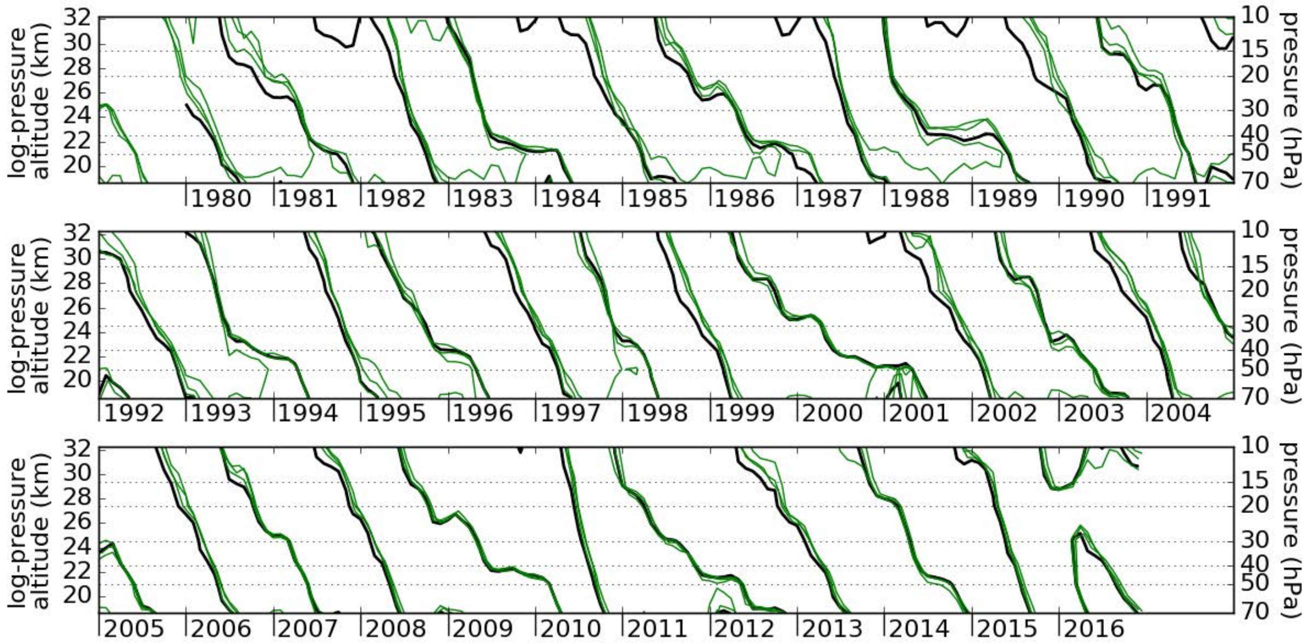


Figure 9.13: *Descent of zero wind lines of the monthly-mean 2 ° S - 2 ° N zonal-mean zonal wind. Thick black: MERRA-2, thin green: ERA-Interim, JRA-55, CFSR.*

Characterisation of the QBO using reanalyses, as performed in this chapter, therefore contributes to the development of the underlying forecast models (including climate models) by providing an observation-based description of major aspects of the QBO for model evaluation. As a benchmark for such efforts, we present here selected key metrics of the QBO based on those defined in *Schenzinger et al.* (2017).

Figure 9.15 shows a summary of these QBO metrics for the 1980 - 2012 period derived by taking the average of the four most recent full-input reanalyses (ERA-Interim, MERRA-2, JRA-55 and CFSR) that comprise the reanalysis mean. Note that the 1980 - 2012 period excludes the QBO disruption of 2015/16 during which the tropical wind state was very unlike the typical structure of the QBO (*Coy et al.*, 2017; *Newman et al.*, 2016; *Osprey et al.*, 2016). The top two rows (panels **a** - **f**) show results from a spectral analysis of the zonal-mean zonal wind and zonal-mean temperature. The diagnostics are calculated separately for each reanalysis and then averaged together for display in **Figure 9.15**; *Appendix A9.3* shows these metrics separately for each of 10 different reanalyses.

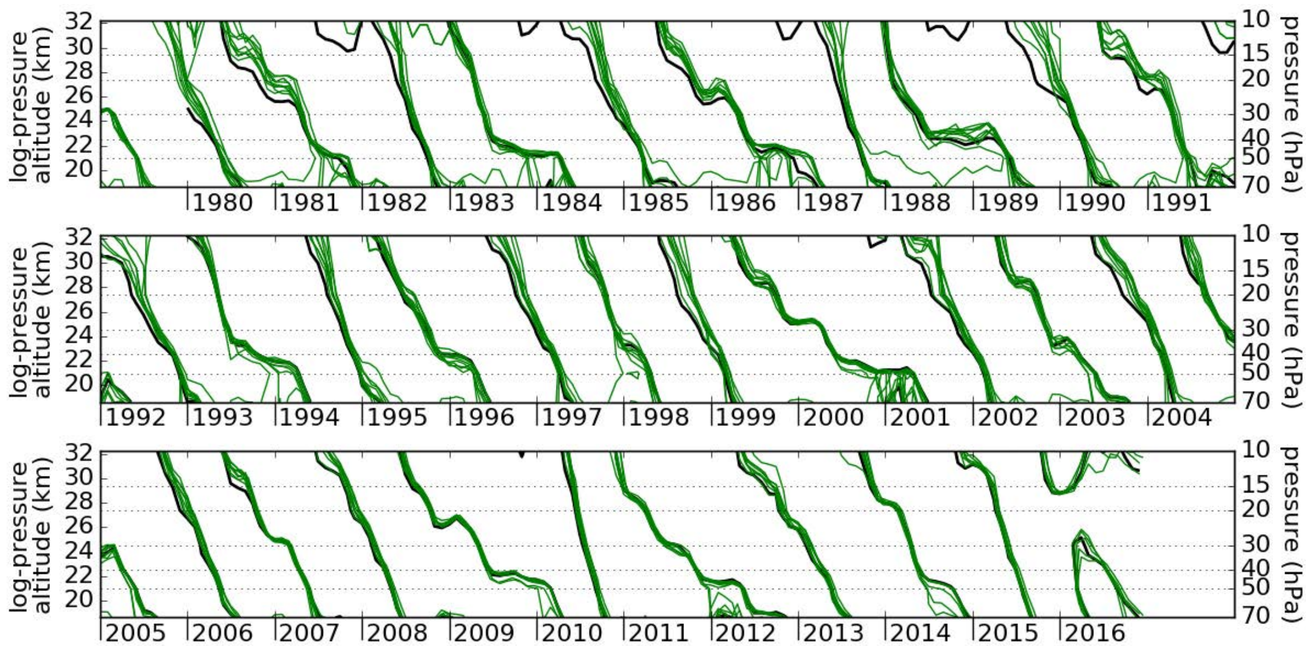


Figure 9.14: *Descent of zero wind lines of the monthly-mean 2 ° S - 2 ° N zonal-mean zonal wind. Thick black: MERRA-2, thin green: all others (ERA-40, ERA-Interim, MERRA, JRA-25, JRA-55, JRA-55C, CFSR, NCEP-NCAR, NCEP-DOE).*

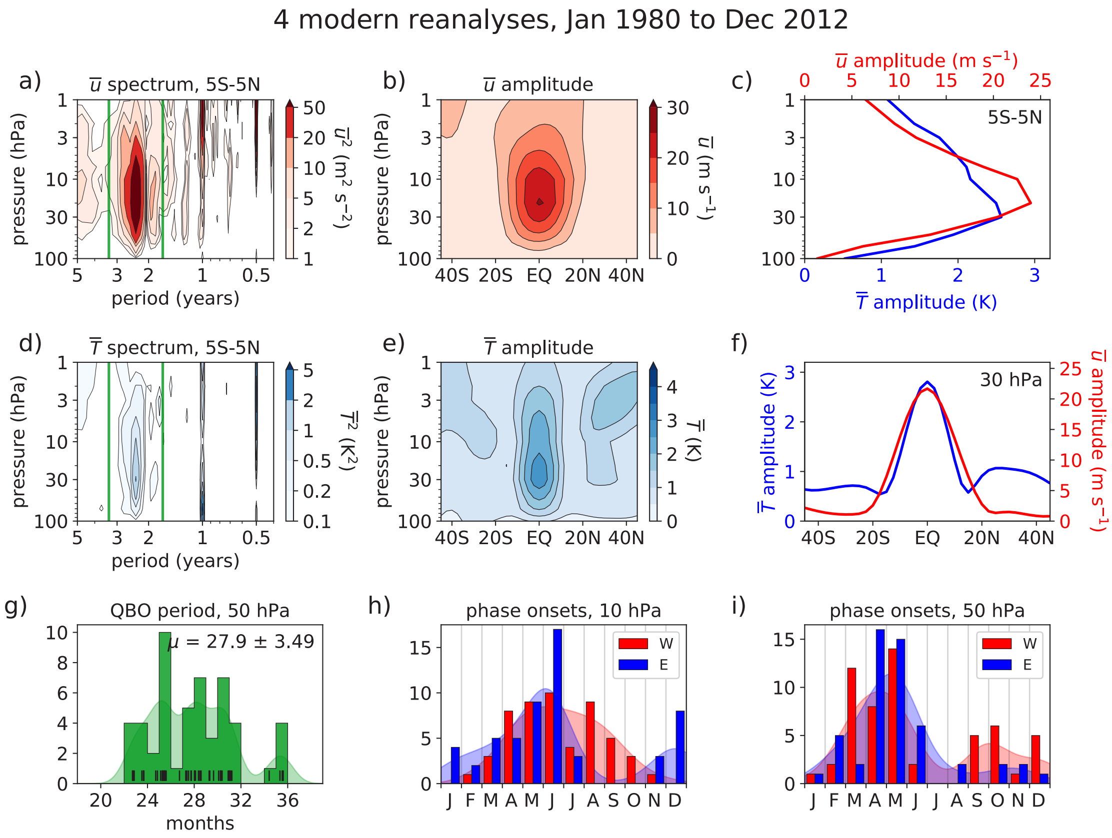


Figure 9.15: *QBO metrics based on Schenzinger et al. (2017), for the four modern reanalyses ERA-Interim, CFSR, JRA-55 and MERRA-2, for the 1980 - 2012 period, using zonal-mean zonal wind, u, and zonal-mean temperature, T. (a - f): Metrics based on Fourier decomposition, computed separately for each reanalysis and then averaged. (g - i): Metrics based on QBO phase transitions, aggregated for the four reanalyses. Fourier components computed from u, T averaged over 5 °S - 5 °N are shown in (a) and (d), respectively, with vertical green lines indicating 20 and 40 month periods. QBO spectral amplitudes are shown (b,c,e,f), defined as in* ***Figure 9.19*** *by averaging the Fourier components with periods between 20 and 40 months. (b,e) show amplitudes computed separately at each gridpoint. (c) shows amplitudes computed from u, T averaged over 5 °S - 5 °N. (f) shows amplitudes computed using u, T at 30 hPa. In (g - i) each distribution is shown both as a histogram (bars) and Gaussian kernel estimate (smooth curve in background). In (g) the exact cycle durations for each complete QBO phase (defined here as the time between successive 50 hPa westerly phase onsets) are shown as short black vertical bars at bottom, and the mean plus/minus one standard deviation (in months) of the QBO period distribution are indicated at top right. In (h,i) the red and blue bars indicate QBO westerly (W) and easterly (E) phase onsets, respectively, at the altitudes given in the panel titles. Appendix A9.3: plots in this format for each reanalysis separately.*

Figures 9.15(a) and **(d)** show temporal spectra of the wind and temperature, respectively, as a function of altitude. The QBO, being a somewhat irregular oscillation, straddles a range of periods between two and three years. Annual and semi-annual harmonics are visible in the SAO region above ≈ 5 hPa for wind, and also at lower altitudes for temperature. The QBO clearly dominates the wind variability in the 10 - 70 hPa layer, while for temperature the semi-annual variability extends deeper into the mid-stratosphere and there is significant annual temperature variability above the tropical tropopause. (The small annual component of wind variability near these altitudes can be seen in **Figure 9.19**.)

The vertical green lines in panels **(a,d)** of **Figure 9.15** indicate the range of periods over which the spectral amplitudes are averaged to derive the QBO amplitudes shown in panels **(b,c,e,f)**. The zonal wind QBO peaks in the mid-stratosphere, while the temperature peak occurs slightly below the wind peak. This is consistent with the temperature peak being associated with the vertical shear of the zonal wind below the descending QBO phase. In panels **(e)** and **(f)** the subtropical lobes in the temperature amplitude are indicative of the QBO mean meridional circulation, discussed below (*Section 9.2.3*).

Figure 9.15(g) shows the distribution of QBO periods (in months) at 50 hPa aggregated for the four modern reanalyses. Durations of QBO cycles span a range from just under 2 years up to 3 years, with a mean period close to 28 months and standard deviation of about 3.5 months. The seasonal distribution of the QBO-W and QBO-E phase onsets at 10 hPa and 50 hPa are shown in **Figure 9.15(h)** and **(i)** respectively.

The periods and phase onsets were evaluated separately in each of the four reanalyses and then combined together to form these distributions. The distributions show that QBO phase onsets can occur at any time of year but there is a seasonal preference. The tendency of the QBO-E onsets at 50 hPa to occur during Northern Hemisphere (NH) spring may be associated with "stalling" of QBO-E descents during NH winter, possibly due to increased tropical upwelling of the Brewer-Dobson circulation at that time (*e.g.*, *Hampson and Haynes*, 2004) although seasonal variations in wave forcing may also play a role (*Maruyama*, 1991). These same factors could also affect 10 hPa and 50 hPa QBO-W onsets, although QBO-W descent is typically faster and more regular than QBO-E descent (*e.g.*, **Figure 9.1**). Given the systematic delay in the descent of the QBO-W phase seen in the reanalyses (**Figures 9.13** and **9.14**) it is likely that the true location of the peak likelihood of QBO-W onsets is shifted slightly left, toward earlier in NH spring.

A comparison of the REM QBO amplitudes is shown in **Figure 9.16** for various combinations of the reanalyses, along with their inter-reanalysis spread. The spatial pattern of inter-reanalysis zonal-mean zonal wind SD when all of the reanalyses are included (panel **a**) qualitatively resembles that shown in **Figure 9.8**, but with the whole pattern shifted upward in altitude. Note that **Figure 9.8** showed the total inter-reanalysis SD whereas **Figure 9.16** shows the SD of the spectrally filtered wind that has retained only the prescribed range of QBO periodicities (20 - 40 months). A similar pattern is evident when only the four most recent full-input reanalyses are included (panel **b**), and the magnitude of the SD is reduced. The inter-reanalysis wind SD is further reduced at the lower altitudes (below 20 hPa) when CFSR is removed from the ensemble-average (panel **c**), leaving only ERA-Interim, JRA-55 and MERRA-2 in the ensemble. The temperature amplitude (panels d - f) behaves similarly (although removing CFSR has little impact as lower altitudes). However, appreciable inter-reanalysis differences remain at upper altitudes (above 20 hPa) even for the group of three modern reanalyses (panels **c,f**). This is expected for zonal wind since the upper altitude limit of radiosonde observations is ≈ 10 hPa, but **Figure 9.16f** shows it is also true for the QBO temperature amplitude.

Given the prominent inter-reanalysis differences in QBO amplitude at higher altitudes shown in **Figure 9.16**, characteristics of the QBO should be regarded as increasingly uncertain as altitude increases (as was also indicated by **Figure 9.8**). The seasonal timing of QBO phase onsets (*Dunkerton*, 1990) for altitudes from 5 hPa to 70 hPa is shown in **Figures 9.17** and **9.18** for the actual and deseasonalized winds, respectively, for the 1980 - 2012 period, using the four most recent full-input reanalyses (aggregated in the same manner as for the histograms of **Figure 9.15**).

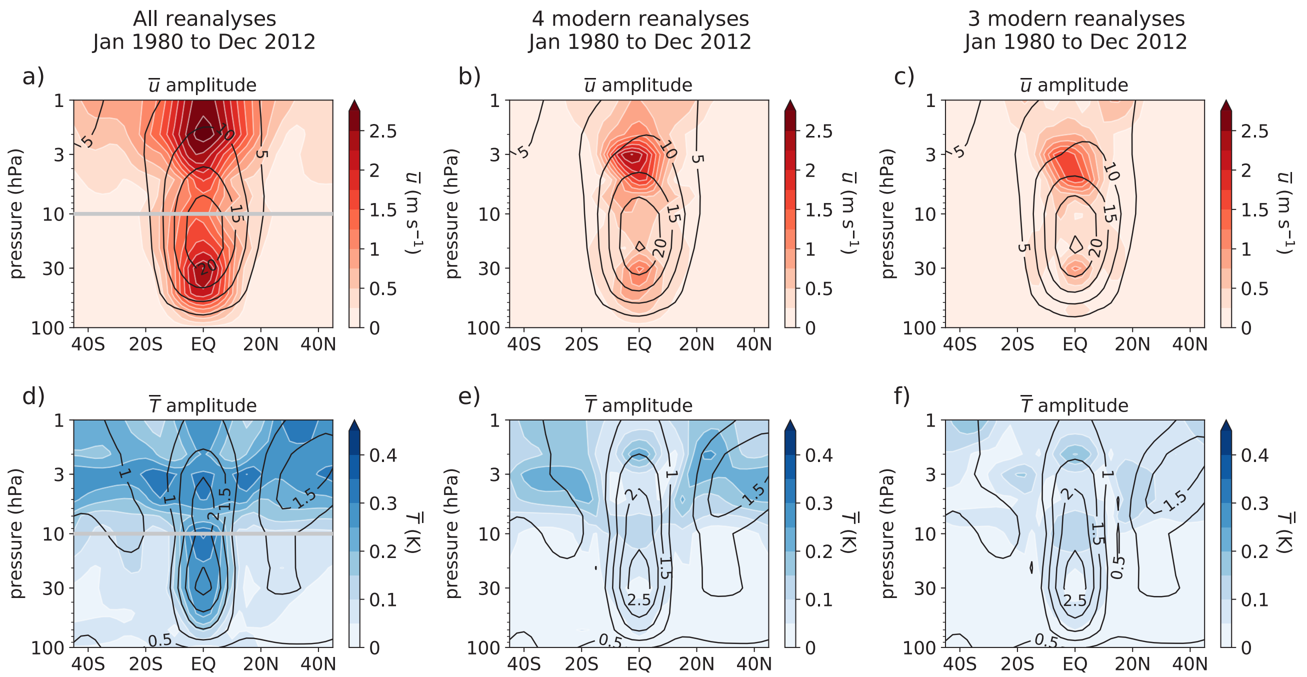


Figure 9.16: *Latitude-altitude distribution of reanalysis-mean QBO spectral amplitude defined as in* ***Figure 9.15*** *(line contours) and its inter-reanalysis standard deviation (coloured contours), for the 1980 - 2012 period. As in* ***Figure 9.15****, spectral amplitudes are first computed separately for each reanalysis, and their mean and standard deviation across reanalyses are then computed. (a - c): Amplitude for zonal-mean zonal wind, u. (d - f): Amplitude for zonal-mean temperature, T. (a,d): All reanalyses. ERA-40 is included although it only extends to August 2002; excluding it gives similar results. Above 10 hPa (horizontal grey line) the NCEP-NCAR and NCEP-DOE reanalyses are excluded because they provide no data at these altitudes. (b,e): ERA-Interim, CFSR, JRA-55 and MERRA-2. (c,f): As (b,e) but excluding CFSR.*

Onset times shift to the right with descending altitude, due to the descent of QBO phases. Semi-annual periodicity at the upper levels is readily apparent for transitions of the actual wind (**Figure 9.17**) and is reduced but not removed entirely in transitions of the deseasonalized wind (**Figure 9.18**). The wide spread of onset times at the upper levels is partly due to inter-reanalysis disagreements at these levels (see **Figures AS9.6 - AS9.13** for the corresponding plots for the reanalyses individually). This combination of four reanalyses is useful to show a coherent pattern of seasonally varying phase descent. Similar patterns are seen for each individual reanalysis but are noisier due to variations in the timing of QBO onsets that result in the transitions being grouped into different months in the different reanalyses (**Figures AS9.6 - AS9.13**). Hence the reanalysis ensemble (**Figures 9.17** and **9.18**) should characterize this behaviour more reliably than any single reanalysis, although with the caveats that 1) systematic errors in the QBO-W onset timing likely influences the results; 2) the results are more uncertain at higher altitudes; and 3) the shortness of the observed record (37 years in **Figures 9.17** and **9.18**) implies that sampling uncertainty may be appreciable.

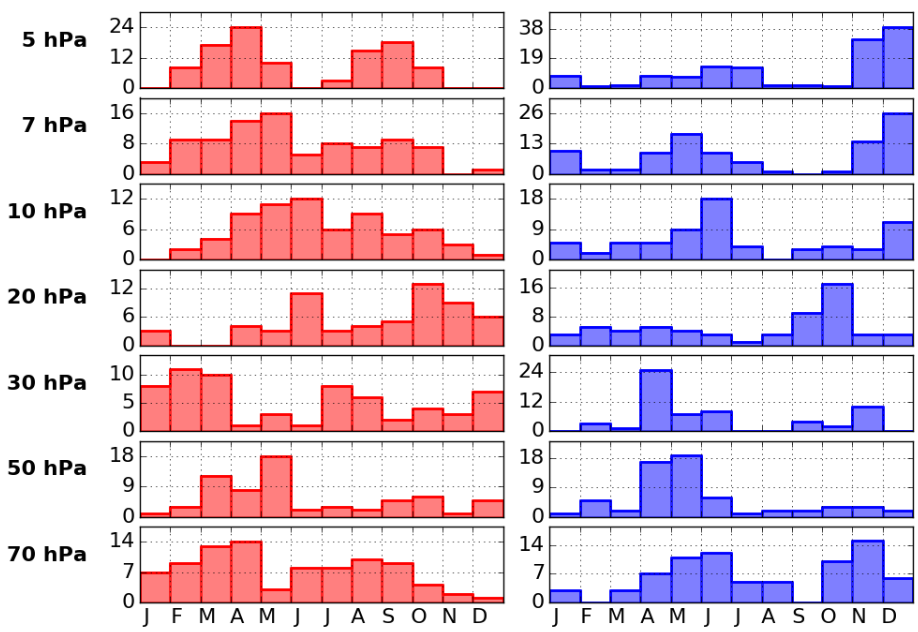


Figure 9.17: *Seasonal distribution of QBO phase onsets during the 1980 - 2016 period, combining the onsets in ERA-Interim, MERRA-2, JRA-55 and CFSR into single histograms. Timing of onsets is diagnosed from the monthly-mean 5 °S - 5 °N zonal-mean zonal wind, interpolated to the time of the zero crossing. Red bars (left column) indicate QBO-W onsets and blue bars (right column) indicate QBO-E onsets.*

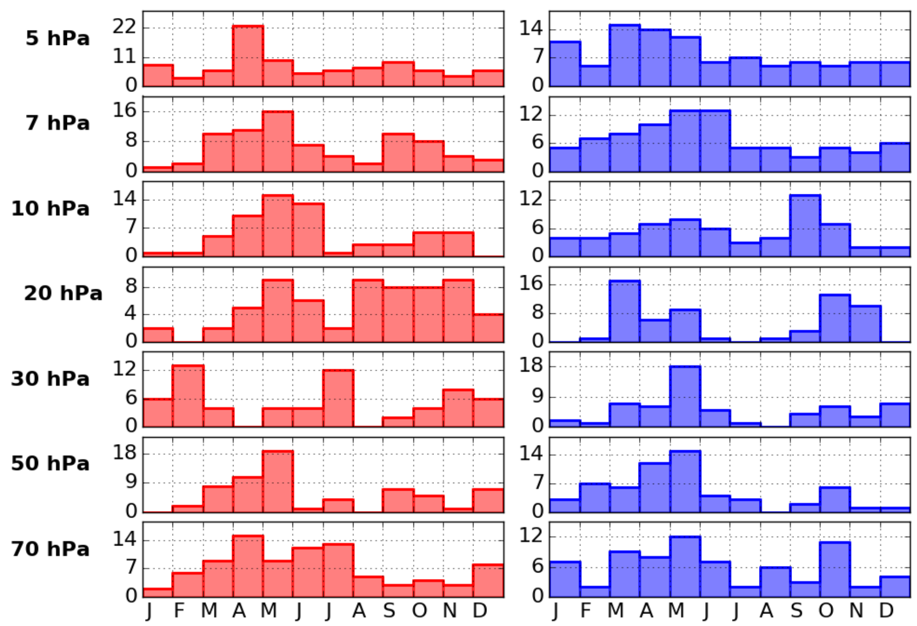


Figure 9.18: *As Figure 9.17, but onsets are defined using deseasonalized monthly-mean 5 °S - 5 °N zonal-mean zonal wind.*

To more closely examine inter-reanalysis differences in the vertical structure of tropical stratospheric variability, **Figure 9.19** compares the vertical profiles of QBO, annual cycle, and SAO zonal-mean zonal wind amplitude in all reanalyses. These are defined by averaging the Fourier components with periods between 20 and 40 months for the QBO (in **Figures 9.15** and **9.16**), and by the 12- and 6-month components for the annual cycle and SAO. The 20 - 40 month window encompasses the complete range of QBO periods shown in **Figure 9.15(g)**, but the overall vertical structure of QBO amplitude is not sensitive to this choice. Another commonly used measure of QBO wind amplitude (*e.g.*, *Kawatani and Hamilton*, 2013) is the temporal standard deviation of deseasonalized wind multiplied by $\sqrt{2}$ (*Dunkerton and Delisi*, 1985). **Figure AS9.14** compares this amplitude to the spectral amplitude. In the 10 - 70 hPa layer the two are virtually identical except that the spectral amplitude is about 10 - 15 % smaller.

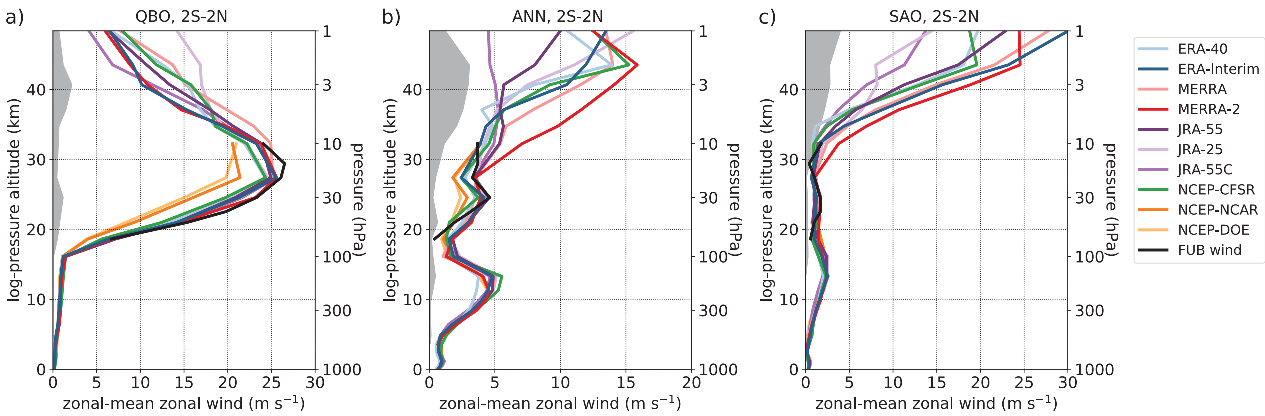


Figure 9.19: *Spectral amplitude of monthly-mean 2 °S - 2 °N zonal-mean zonal wind for the 1980 - 2012 period in all reanalyses, and in monthly-mean FUB zonal wind, for the (a) QBO, (b) annual, and (c) semi-annual periodicities. (For ERA-40, the 1980 - 2002 period is used.) QBO periodicity is defined by a 20 - 40 month window. Grey shading shows the inter-reanalysis standard deviation of the four most recent full-input reanalyses (ERA-Interim, MERRA-2, JRA-55 and CFSR). Grey shading cuts off at ≈ 4 km, below which MERRA-2 zonal means are not available. (Note the different x-axis ranges.)*

This slight reduction is due to the 20 - 40 month window used to define the spectral amplitude; broadening the window to encompass variability at all frequencies makes the two amplitudes identical (not shown). Note that the definition of spectral amplitude used here also includes the $\sqrt{2}$ factor, making the spectral and Dunkerton-Delisi amplitudes exactly comparable[8]. Interestingly, although FUB winds have larger amplitude than the reanalyses by either measure (**Figure AS9.14a,b**) the ratio of amplitudes for FUB is generally smaller (**Figure AS9.14c**), indicating that variability outside the 20 - 40 month window is stronger in FUB than the reanalyses. Above and below the 10 - 70 hPa layer, the fraction of variability associated with the 20 - 40 month range of QBO periodicities decreases, as expected based on **Figure 9.19**[9]. All reanalyses agree on this overall structure.

Figure 9.19 indicates that in all reanalyses the QBO dominates tropical zonal-mean zonal wind variability in the lower stratosphere and the SAO dominates the upper stratosphere; the annual cycle is mostly small compared to both, except in the troposphere and upper stratosphere. Inter-reanalysis disagreement increases with altitude. Amplitudes for monthly-mean FUB zonal winds are also shown in **Figure 9.19**. These are not exactly comparable to the reanalysis zonal-mean amplitudes because FUB winds are sampled at one location, Singapore (1.4 °N, 103.9 °E), during the 1980 - 2012 period (**Figure 9.1**). Bearing this important caveat in mind, we assume here that monthly means at a single longitude approximate the true zonal mean, due to the expected zonal symmetry of the QBO (*i.e.*, to the extent that the monthly mean adequately removes wave signatures from the radiosonde winds). The FUB amplitude peaks at 15 hPa, but as this pressure level is not included in the standard set of pressure levels for the reanalysis data used here, the apparent large disagreement at 15 hPa mainly reflects the absence of the 15 hPa level in the reanalysis data. Nevertheless, at all other levels in the 10 - 70 hPa layer, **Figure 9.19** shows that the QBO zonal-mean amplitude in all reanalyses is weaker than the FUB amplitude[10]. This suggests that in reanalyses the QBO amplitude in the 10 - 70 hPa layer might be generally too weak throughout the tropical belt, although examination of a wider range of radiosonde stations would be useful to confirm this. The weak amplitude would be consistent with the results of *Das et al.* (2016) who compared the QBO amplitude in radiosonde observations at Thumba, India (8.5 °N, 76.9 °E) with that in the ERA-40, ERA-Interim, MERRA and NCEP-NCAR R1 reanalyses sampled near the location of Thumba.

Figure 9.20(a) shows the differences between reanalysis and FUB wind amplitude, which have magnitudes of less than 4 m s^{-1} at all altitudes except in the two older NCEP reanalyses. In most reanalyses, including the older NCEP ones, the largest underestimates occur near 30 hPa. In relative terms (normalized by the FUB amplitude), **Figure 9.20(b)** shows that these differences tend to be largest at 50 hPa, where the reanalysis amplitudes range from about 5 % (MERRA-2, JRA- 55C) to almost 50 % (NCEP-NCAR R1, NCEP-DOE R2) smaller than FUB.

[8] The usefulness of the $\sqrt{2}$ factor is to make the defined amplitude representative of the magnitude of the peak of the oscillation. For a sinusoidal oscillation $A\sin(t)$, the peak value is A and the variance is ½ A^2. Hence $\sigma = A/\sqrt{2}$, *i.e.*, the peak value is $\sqrt{2}\sigma$ and the peak-to-peak variation is twice this value.

[9] Notably the 20 - 40 month variability extends weakly into the troposphere, with all reanalyses in good agreement. This diagnostic does not indicate whether or not the tropospheric variability at these periods is coherent with the stratospheric QBO.

[10] Visual comparison with the ERA-Interim amplitude shown in **Figure 10** of *Bushell et al.* (2020), which does include the 15 hPa level, indicates that ERA-Interim is smaller than FUB at this level as well.

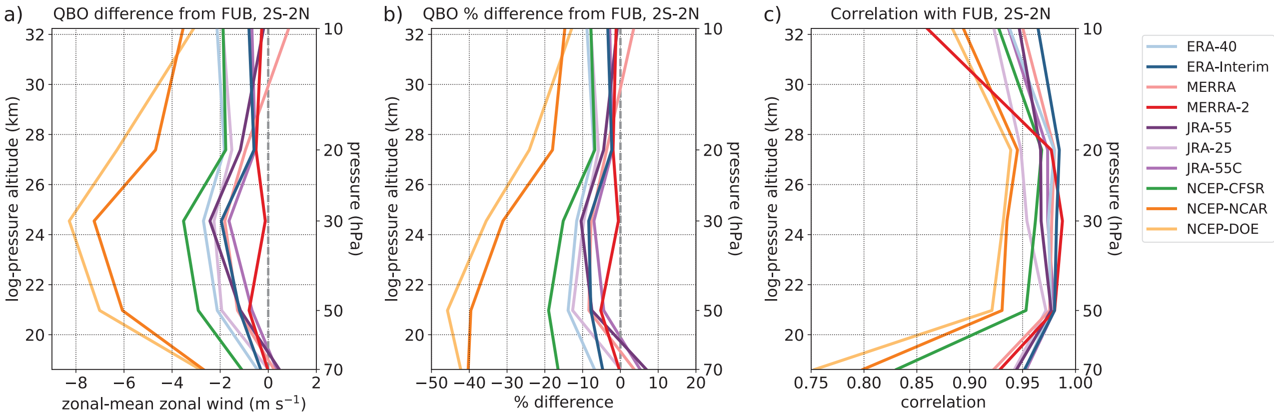


Figure 9.20: *(a) Difference of QBO spectral amplitude (**Figure 9.19**) between reanalyses and FUB wind (reanalysis minus FUB) at the 5 vertical levels common between FUB and the reanalysis pressure levels data (10, 20, 30, 50, and 70 hPa). (b) As (a), but in terms in percentage difference (the denominator is the FUB amplitude). (c) Linear correlation coefficient between deseasonalized reanalysis zonal-mean zonal wind and deseasonalized FUB zonal wind.*

Weak QBO amplitude near 50 hPa is a common problem in QBO-resolving AGCMs (*Bushell et al.*, 2020). **Figure 9.20(b)** suggests that all of the reanalysis forecast models share this problem, and that data assimilation has ameliorated but not entirely removed it. Interestingly, JRA-55C shows a slightly larger amplitude than JRA-55 at 20 - 50 hPa, possibly indicating a small deleterious impact of satellite radiance assimilation on the QBO amplitude; such an effect was found by *Pawson and Fiorino* (1998) in early reanalyses that assimilated retrieved temperature profiles rather than radiances. (Recomputing the spectral amplitudes for a variety of different subperiods, including before and after the 1998 AMSU transition, gives similar results.) At the lowest altitude, 70 hPa, reanalyses show both signs of difference with FUB, but at this altitude the tropical circulation has more zonal asymmetry and the FUB winds are not as good a proxy for the zonal mean (**Figure 9.10**; *Kawatani et al.*, 2016). At 10 hPa most reanalyses underestimate the QBO amplitude, whereas free-running AGCMs may underestimate or overestimate the QBO amplitude at this altitude (*e.g.*, **Figure 10b** of *Bushell et al.*, 2020), which is likely due to AGCMs commonly relying on parameterized non-orographic GWD to provide much of the QBO wave forcing. Of the reanalyses shown here, only MERRA-2 has a GWD scheme that is tuned to yield a realistic QBO in the forecast model (*Coy et al.*, 2016). This model also shows the best agreement with FUB, suggesting that in this case the data assimilation is mainly acting to nudge the reanalysis QBO toward the phase of the real QBO, rather than correcting for a significant model bias, *i.e.*, the lack of a spontaneous QBO in the forecast model. At 30 hPa, the altitude where most reanalyses show the largest wind difference with FUB, MERRA-2 shows extremely close agreement.

Figure 9.20(c) shows the correlation of reanalysis monthly-mean deseasonalized zonal-mean zonal wind with FUB wind. For every reanalysis, correlations are highest in the 20 - 50 hPa layer, with MERRA-2 again showing the best agreement at 30 hPa. All reanalyses have correlations higher than 0.90 in this layer, and above 0.95 for the most recent full-input reanalyses. Poor correlations at 70 hPa are likely due to the FUB wind at 70 hPa being a poor proxy for the zonal mean, while the poor correlation of MERRA-2 at 10 hPa reflects its unrealistic features at this level (**Figure 9.2**; *Kawatani et al.*, 2016). Since the correlation coefficient is insensitive to amplitude differences, the fact that all correlations are less than 1 indicates differences in the timing of QBO phase onsets between FUB and the reanalyses. As noted earlier (**Figures 9.13** and **9.14**) phase onsets in reanalyses are often slightly delayed, especially for the westerly QBO phase. *Pawson and Fiorino* (1998) examined two early reanalyses, NCEP-NCAR R1 and ERA (a precursor to ERA- 40), and found that NCEP at all levels in the 10 - 70 hPa layer correlated better with ERA than with the observations. The correlation matrix for all reanalyses and FUB winds is shown in **Figure 9.21(a)**. All correlations are high (over 0.90), but a number of the weaker correlations occur between FUB and reanalyses as well as between MERRA-2 and other reanalyses. This behaviour is associated with the representation of QBO transitions in the reanalyses. In **Figure 9.21(b)**, each correlation is computed using only times during which both time series have a magnitude less than 0.5σ and hence are close to QBO phase onset times, which retains roughly 10 % to 20 % of the data (depending on the reanalysis). The correlations of FUB winds with reanalyses (rightmost column) are mostly lower than the correlations between reanalyses (all other columns), similar to the result of *Pawson and Fiorino* (1998). The exception is MERRA-2, which correlates well with FUB but poorly with many of the reanalyses. This suggests the forecast model improvements in MERRA- 2 have significantly improved the representation of QBO phase transitions. The FUB winds are not a perfect proxy for the zonal mean but should be best suited to this purpose in the middle of the 10 - 70 hPa layer (at 20, 30, or 50 hPa) since zonal asymmetries are larger at 70 hPa and data quality is poorer at 10 hPa. The correlation matrix for 20 hPa (not shown) looks very similar to 30 hPa.

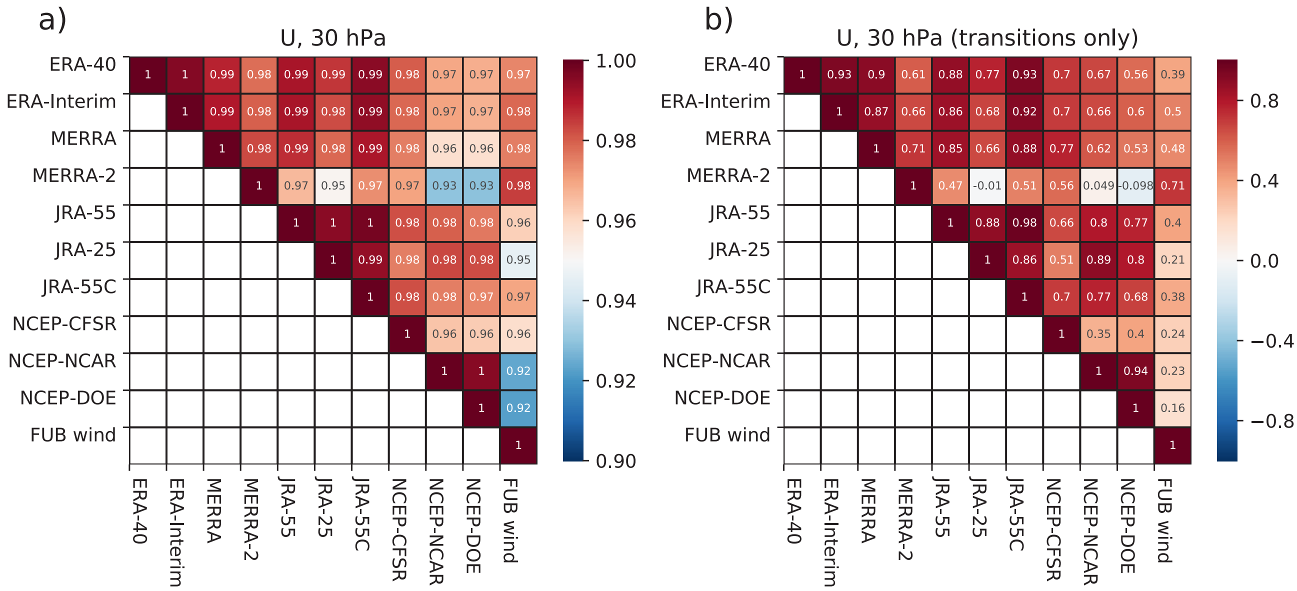


Figure 9.21: *(a) Correlations at 30 hPa between deseasonalized zonal-mean zonal wind in all reanalyses and deseasonalized FUB zonal wind. (b) As (a), but correlations are computed only for times during which both time series have a magnitude less than 0.5σ. Note the different colour scales used in (a) and (b).*

The matrix for 50 hPa (not shown) shows similar overall behaviour but MERRA-2 does not stand out so clearly from the other reanalyses, which perhaps could be due to non-orographic GWD having a larger impact at higher altitudes than near the bottom of the QBO (50, 70 hPa). Finally, an interesting feature of **Figure 9.21** is that high correlations appear between families of reanalyses: all three JRA products, the two ERA products, and the two older NCEP reanalyses. This strongly suggests that errors in QBO phase transition timing are not random but instead are caused by systematic features of the reanalysis systems that persist across different generations of reanalysis products.

Vertical profiles similar to **Figure 9.19** but for QBO amplitude in zonal-mean temperature are shown in **Figure 9.22**. Similarly to the zonal wind, inter-reanalysis spread increases with altitude, but it is notable that all reanalyses except the two earlier NCEP ones agree well at the altitude of the peak QBO temperature amplitude, 30 hPa, and the spread is larger above and below this. The annual cycle is large in the tropopause region and also above it, up to ≈ 30 hPa. The temperature QBO peaks at 30 hPa, slightly lower than the peak wind amplitude, consistent with thermal wind balance. This temperature anomaly is balanced by a mean-meridional circulation, which extends into the subtropics (*Plumb and Bell*, 1982) and is discussed further in the next section. **Figure 9.23** shows latitudinal profiles of the QBO amplitude in zonal-mean zonal wind (panels **a–c**) and zonal-mean temperature (panels **d–f**) at the 10, 50 and 100 hPa levels. Subtropical lobes are clear in the temperature amplitude at 10 and 100 hPa, but much less clear at 50 hPa. At 100 hPa, near the tropical tropopause, the temperature amplitude (panel **f**) looks similar in all reanalyses except for three of the older ones (NCEP-NCAR R1, NCEP-DOE R2 and ERA-40). Further analysis of the 100 hPa QBO temperature amplitude is given in *Section 9.2.4*.

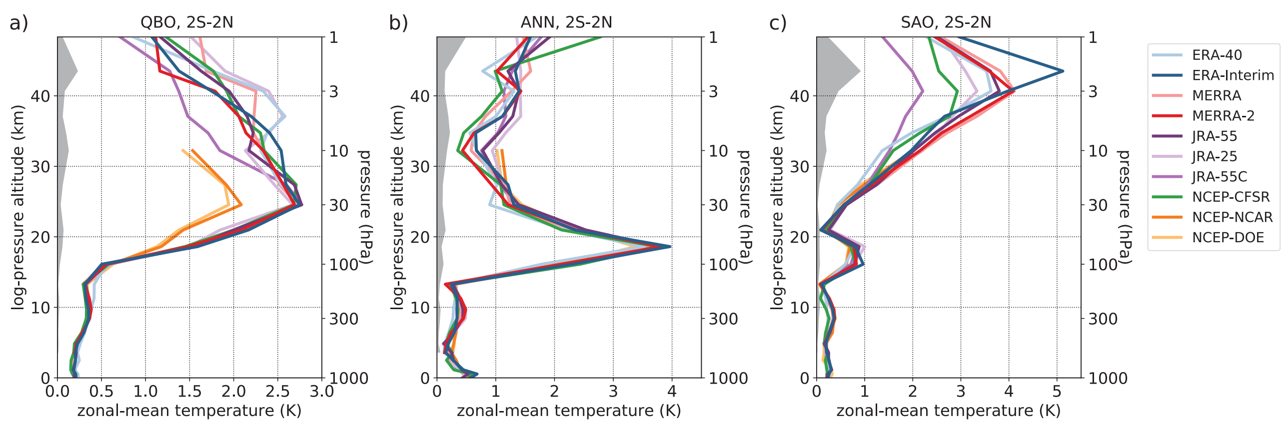


Figure 9.22: *As **Figure 9.19**, but for zonal-mean temperature. Note the different x-axis ranges.*

The zonal wind QBO amplitude at 100 hPa (panel **c**) has a different latitudinal structure than the wind amplitude at higher altitudes (panels **a,b**), with subtropical peaks centred near 20 °S and 20 °N in all reanalyses. These might be related to the "horseshoe" structure of subtropical wind anomalies extending downward from the QBO that has been noted in the literature (*e.g.*, **Figure 4** of *Anstey and Shepherd*, 2014). At the equator the reanalyses appear to segregate into two groups, with the lower amplitude group consisting of the two ERA products, the two older NCEP products, and JRA-55C. This bimodality is sensitive to slight changes in the definition of the QBO period window (*e.g.*, using 25 - 33 months instead of 20 - 40 months) but for any reasonable choice of QBO period window the inter-reanalyis spread in 100 hPa equatorial QBO wind amplitude is appreciable, even among the four most recent full-input reanalyses (shading in **Figure 9.23c**).

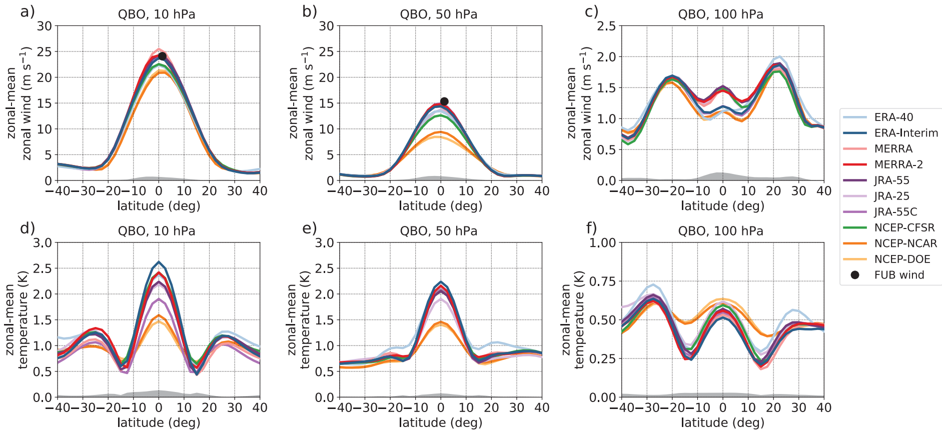


Figure 9.23: *Latitudinal profiles of QBO spectral amplitude for zonal-mean zonal wind and FUB wind (a–c) and zonal-mean temperature (d–f) at (a,d) 10 hPa, (b,e) 50 hPa, and (c,f) 100 hPa. As in* ***Figures 9.19*** *and* ***9.22****, grey shading shows the inter-reanalysis standard deviation of the four most recent full-input reanalyses (ERA-Interim, MERRA-2, JRA-55 and CFSR). (Note the different y-axis ranges used for the 100 hPa panels).*

9.2.3 Mean meridional circulation

As noted already, the QBO in zonal-mean zonal wind is in thermal wind balance with the zonal-mean temperature (*Andrews et al.*, 1987), and this balance is maintained by the mean meridional circulation associated with the QBO (*Plumb and Bell*, 1982). Since vertical velocities in the stratosphere are too small to be directly observed, reanalyses provide a way to examine this mean meridional circulation. This was first done by *Huesmann and Hitchman* (2001) using the NCEP-NCAR R1 reanalysis, and the advent since then of newer reanalyses with improved representations of the QBO makes it useful to update their results. However, it is well known that vertical velocity in reanalyses remains subject to considerable uncertainty due to the fact that it is not directly constrained by any assimilated observations (*e.g.*, *Polavarapu et al.*, 2005).

Figure 9.24 shows climatological tropical upwelling (TEM vertical residual velocity) in six modern reanalyses. Results are shown for both model levels and pressure levels data, but from the four reanalyses for which both types of levels are available it is seen that the two types of levels give virtually identical results: little information is lost to vertical interpolation (from model to pressure levels) in this case, although it will be shown in *Section 9.3.1* that this is not the case for wave spectra. All reanalyses shown a qualitatively similar shape of vertical profile, but with substantial quantitative differences at altitudes below 10 hPa. Curiously, all of the reanalyses with the exception of CFSR converge to similar values above 10 hPa. **Figure 9.24** also shows the standard deviation (in time) of the vertical shear of the zonal-mean zonal wind in the same six reanalyses. (Note that the shear can be positive or negative, associated with alternating QBO-W and -E phases, and so its climatological mean is not useful to estimate typical QBO vertical shears).

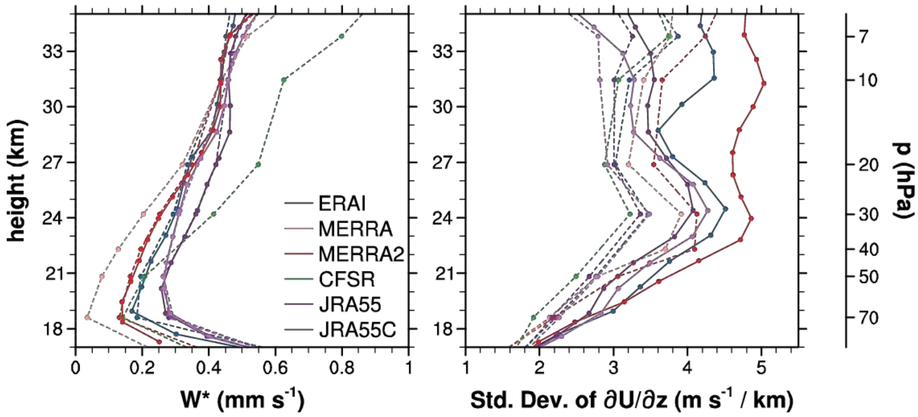


Figure 9.24: *(Left panel) Vertical profiles of climatological TEM residual vertical velocity in six modern reanalyses, averaged over 10 °S - 10 °N. Solid lines are results using model levels data and dashed lines are results using data on the standard pressure levels provided by the reanalysis centres (dots in both cases mark the level locations). (Right panel) Temporal standard deviation of vertical shear of zonal-mean zonal wind, in the same six reanalyses. Solid/dashed lines as in left panel. Updates Figure 5 of Kim and Chun (2015).*

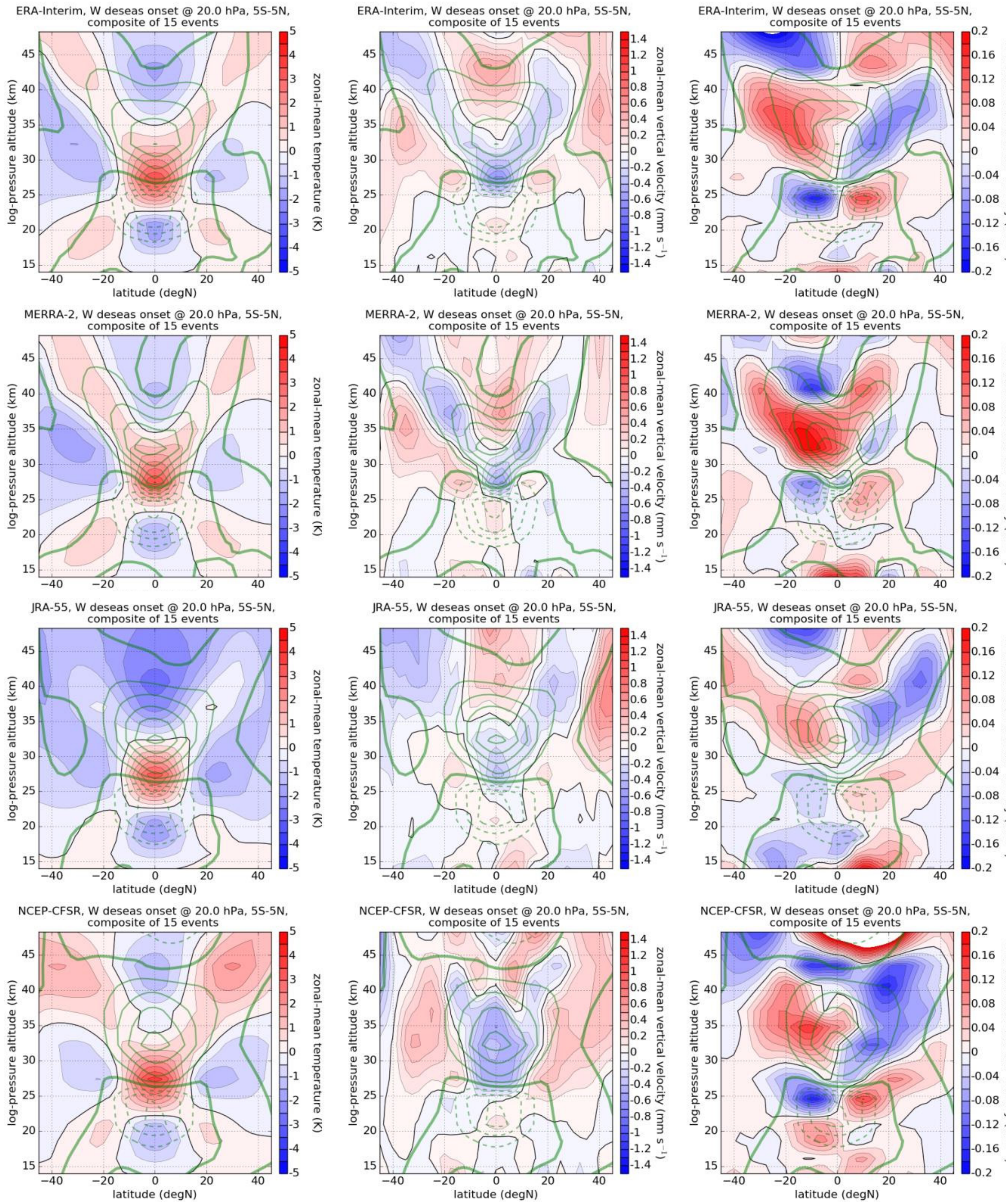


Figure 9.25: *QBO-composited temperature and mean-meridional circulation, based on QBO-W deseasonalized onsets at 20 hPa, for the 1980 - 2016 period of ERA-Interim (1st row), MERRA-2 (2nd row), JRA-55 (3rd row), CFSR (4th row). Left column: zonal-mean temperature, middle column: zonal-mean vertical velocity, right column: zonal-mean meridional velocity. Green lines show the corresponding zonal-mean zonal wind composites, 5 m s*$^{-1}$ *contours (solid: positive, dashed: negative, thick solid: zero). All fields shown are deseasonalized.*

Larger values are estimated using model levels, which have higher vertical resolution than the standard pressure levels. Significant inter-reanalyses variations are present, and the spread increases with altitude. The descent of QBO shear zones can be aided or inhibited by vertical advection, and the two panels show that inter-reanalysis spread in both the upwelling and the shear itself will contribute to inter-reanalysis spread in the magnitude of vertical advection. Since vertical advection affects the QBO descent rate, an estimate of this quantity from reanalyes is useful information for modellers attempting to improve the representation of the QBO. The QBO momentum budget will be discussed in more detail in *Section 9.3.2*.

The anomalous temperature and mean meridional circulation associated with one stage in the life cycle of the QBO – westerly onsets at 20 hPa – is shown in **Figure 9.25** for the four recent full-input reanalyses. The qualitative sense of the circulation is as expected: each of the four reanalyses (rows) shows an equatorial warm anomaly (left column) below the descending QBO-W winds (green contours, identical within each row) accompanied by equatorial downwelling (middle column), and meridional convergence above this downwelling and divergence below it (right column). Despite this general qualitative agreement, the reanalyses show appreciable quantitative differences in both vertical and meridional velocities. Equatorial downwelling below the peak QBO-W winds ranges from 0.3 mm s^{-1} to 0.6 mm s^{-1}, comparable to the climatological values shown in **Figure 9.24**. This indicates that in some cases actual downwelling (rather than just weakened upwelling) may occur during QBO-W descents.

9.2.4 Near-tropopause temperature

The temperature signals shown in **Figure 9.25** indicate a QBO temperature signal near the tropical tropopause. A possible mechanism for the apparent QBO influence on tropical convection is that the QBO affects deep convection by modulating tropical tropopause temperatures (*Tegtmeier et al.*, 2020; *Gray et al.*, 2018; ; *Son et al.*, 2017; *Nie and Sobel*, 2015; *Liess and Geller*, 2012; *Collimore et al.*, 2003). Such modulation could depend on the spatial structure of QBO-induced temperature anomalies, which change sign in latitude with a node at roughly 15° latitude (as indicated in **Figure 9.25**; see also *Collimore et al.*, 2003), and also might exhibit zonal asymmetries associated with the tropical circulation near 100 hPa (such as variations due to the Walker Circulation). To examine QBO-induced temperature near the tropopause and just above it, multiple linear regression (MLR) is used to extract the QBO components of variability from reanalysis, radiosonde, and GNSS-RO temperature data.

Figure 9.26 shows the time series of the QBO component of 10 °S–10 °N zonal-mean temperature variability at 70 hPa and 100 hPa in seven reanalyses and IGRA radiosonde data. Most of the reanalyses agree well with IGRA, with some small differences between them. The older *R1* (NCEP-NCAR R1) reanalysis does not agree well with IGRA, having too weak amplitude and, especially at 100 hPa, large errors in the timing of phase onsets. Excluding *R1*, the reanalyses and IGRA generally agree that the amplitude of QBO-induced temperature variations is, roughly, slightly over 1 K at 70 hPa and slightly under 0.5 K at 100 hPa (*i.e.,* peak-to-peak variations of ≈ 2 K and ≈ 1 K, respectively). The amplitude of the annual cycle in equatorial temperature is between 3.5 K and 4 K at 70 hPa and between 2 K and 2.5 K at 100 hPa depending on the reanalysis (**Figure 9.22b**, which shows 2°S-2°N amplitude but the corresponding values for 10°S-10°N are very similar). Hence the QBO variations are roughly 25-30 % and 15-20 % the size of the annual variation at 70 hPa and 100 hPa, respectively. Note that these regression-based QBO variations are fairly consistent with the 70 hPa and 100 hPa QBO amplitudes seen in **Figure 9.22(a)**, but an exception is NCEP-NCAR R1, which has the weakest 100 hPa amplitude in the regression analysis (**Figure 9.26**) but one of the largest for the spectral amplitude (**Figure 9.22a**, and also **Figure 9.23f**). One possible reason is that the NCEP-NCAR R1 QBO spectral amplitude is contaminated by other sources of variability. The MLR may provide a more reliable estimate of QBO variations since it removes other sources of variability at that project onto QBO timescales, such as ENSO.

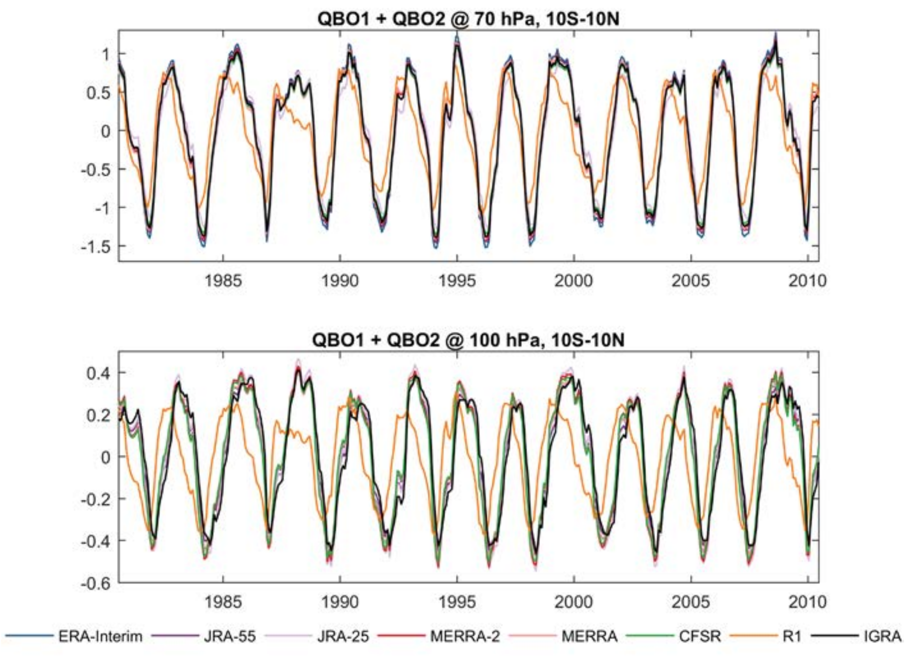


Figure 9.26: *Time series of QBO component of 70 hPa and 100 hPa 10°S-10°N zonal-mean temperature in reanalyses and IGRA radiosondes. (Note, R1 is the NCEP-NCAR reanalysis).*

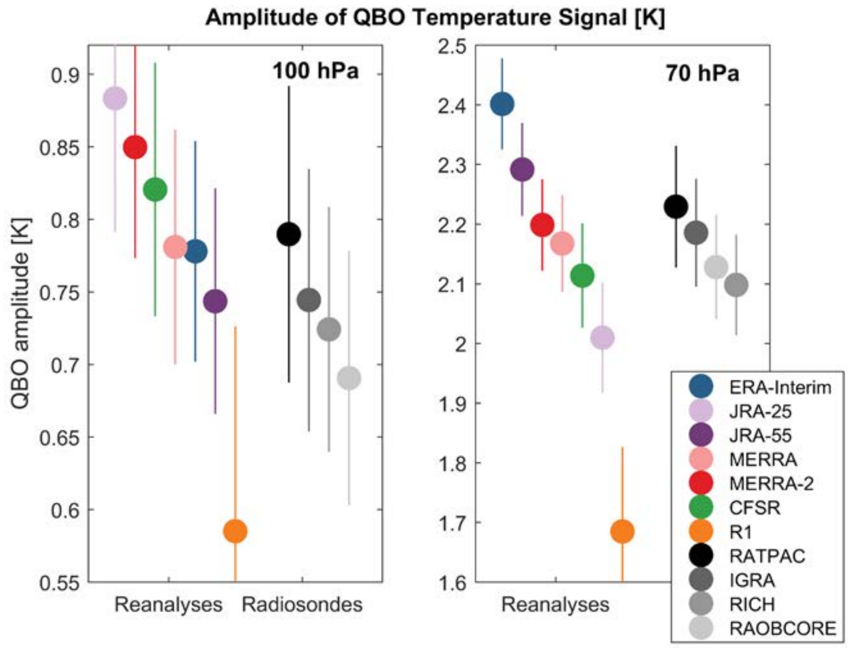


***Figure 9.27:** Regression-based QBO peak-to-peak amplitudes for zonal-mean 10°S-10°N temperature in reanalyses and different radiosonde datasets.*

The peak-to-peak amplitude of the QBO zonal-mean temperature variability at 100 hPa and 70 hPa from the MLR analysis is shown in **Figure 9.27**. At both 100 hPa and 70 hPa, the different sonde datasets agree reasonably well, with substantial overlap of their error bars. Since the different sonde datasets contain many of the same soundings, this suggests that different data processing choices made in the preparation of these datasets do not have a major impact on the QBO component of variability. At 100 hPa, the reanalyses mostly tend to overestimate the sonde-derived amplitudes (again, excluding *R1*). However, the overestimate is about 0.1 K, which is a small fraction (≈ 10 - 15 %) of the overall signal size. To within the uncertainty, as indicated by the large error bars on 100 hPa amplitudes, the sondes and reanalyses agree well. At 70 hPa the reanalyses are more centred on the sonde values, although two (ERA-Interim and JRA-55) still overestimate compared to the sondes. These two are at the lower end of the distribution at 100 hPa, suggesting that the reanalysis spread at 100 hPa is not explained simply as downward propagation of the spread from 70 hPa. Since the reanalyses shown here all assimilate radiosonde data, it is perhaps not surprising that they agree reasonably well with the sondes. The fact that some reanalyses show larger values than the sondes, particularly at 100 hPa, might be due to the assimilation of satellite data and the vertical depth of the weighting functions associated with different channels of nadir-sounding instruments (see *Chapter 2* for further information on satellite weighting functions).

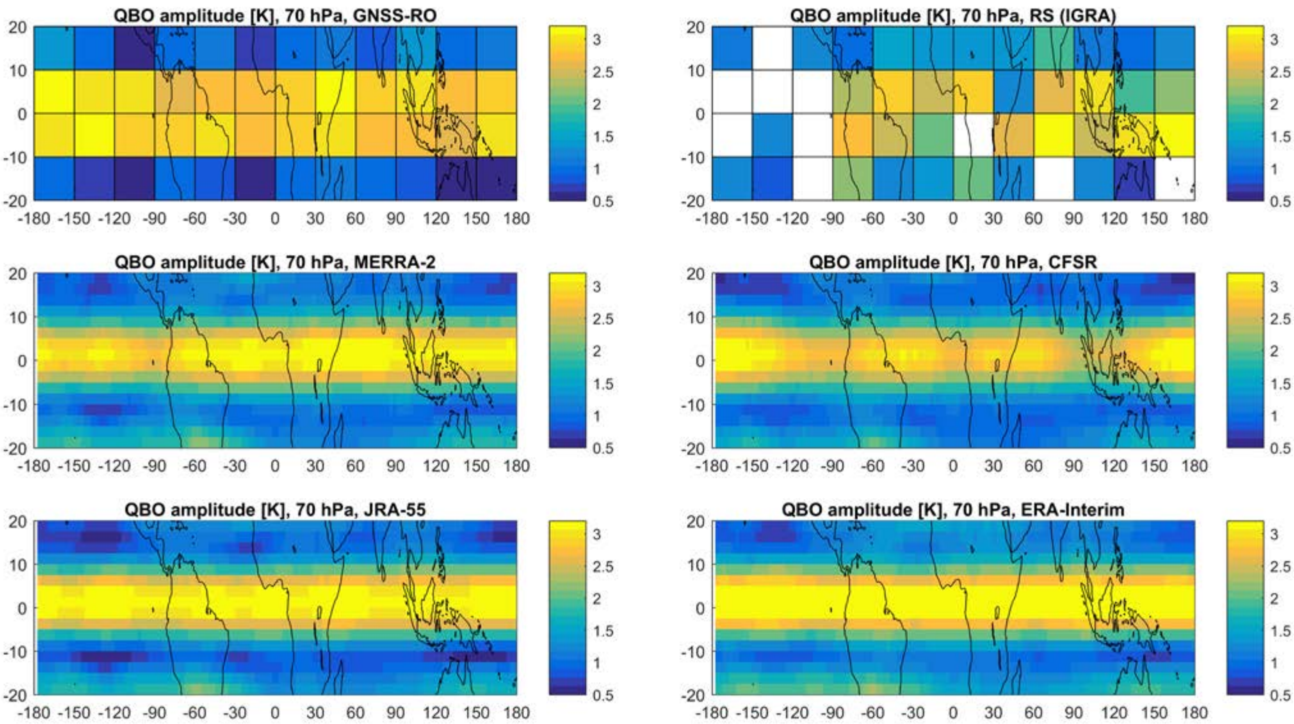


***Figure 9.28:** Latitude-longitude distribution at 70 hPa of regression-based QBO peak-to-peak amplitude for GNSS-RO, IGRA radiosondes, and the four most recent full-input reanalyses. White boxes have no radiosonde stations available.*

Note also that at both levels the inter-reanalysis spread is significantly smaller if the two older reanalyses shown, NCEP-NCAR R1 and JRA-25, are disregarded.

To examine possible zonal variations in QBO-induced temperature anomalies, as in *Tegtmeier et al.* (2020), **Figure 9.28** applies the same MLR method to data binned by its location in latitude and longitude. For the IGRA sondes and GNSS-RO, data are binned into 10 ° latitude × 30 ° longitude boxes. Because of the highly inhomogeneous distribution of tropical radiosondes (**Figure 9.11**), the number of stations in each latitude-longitude box varies widely. Boxes in which no stations are found are coloured white. For GNSS-RO, the relatively homogeneous satellite coverage implies similar numbers of observations in each box, although the exact locations of the individual observations varies. For the reanalyses, the MLR is applied at each spatial gridpoint. As in **Figures 9.26** and **9.27**, the QBO response is characterized by its overall amplitude. This metric obscures changes in sign, which should occur roughly at 15 °S, 15 °N (*Collimore et al.*, 2003, and **Figure 9.25**).

For the IGRA sondes, **Figure 9.28** suggests that at 70 hPa the QBO signal is strongest near the equator (10 °S - 10 °N) and roughly zonally symmetric, although since the results are quite noisy and it is hard to be confident about these features. The corresponding GNSS-RO result, although for a different time period (2002 - 2013, rather than the 1981 - 2010 period used for the sondes), seems to confirm this interpretation. In particular, zonal uniformity extends over the oceans where there are large data gaps in the sonde network. The reanalyses show a similar distribution of QBO temperature amplitude as GNSS-RO and sondes, but with varying amounts of zonal variation, with ERA-Interim being the most zonally symmetric and CFSR the least.

Figure 9.29 shows the corresponding results at 100 hPa, where the QBO temperature amplitude has much stronger zonal variation than seen at 70 hPa. IGRA sondes show the largest signal over Indonesia, Malaysia and the Indian Ocean. Occurrence of the largest signal nearest to the location of the most reliable observations (Singapore) may be cause for concern, but the GNSS-RO results show a similar pattern, albeit with the Indian Ocean response not as prominent and a strong response over West Africa. Taken together, the IGRA and GNSS-RO results suggest that a local maximum in QBO temperature variation may be a robust feature over Indonesia, since this feature appears in both datasets, covering different time periods. The QBO temperature amplitude at this location is ≈ 1 - 1.2 K, roughly 20 - 30 % larger than the ≈ 0.7 - 0.8 K inferred from the zonal-mean analysis of the sonde datasets (**Figure 9.27**). A similar feature over Indonesia is also seen in the four reanalyses shown. Since sonde observations at Singapore are likely to have a large impact on reanalysis QBOs (due to their frequency and high quality), any feature localized near Singapore naturally raises the suspicion that it may be an artefact of the data assimilation. The occurrence of this feature in the IGRA and GNSS-RO lends some confidence that it may be real. Note that the reanalyses assimilate both sonde and GNSS-RO data, which should contribute to the occurrence of this feature.

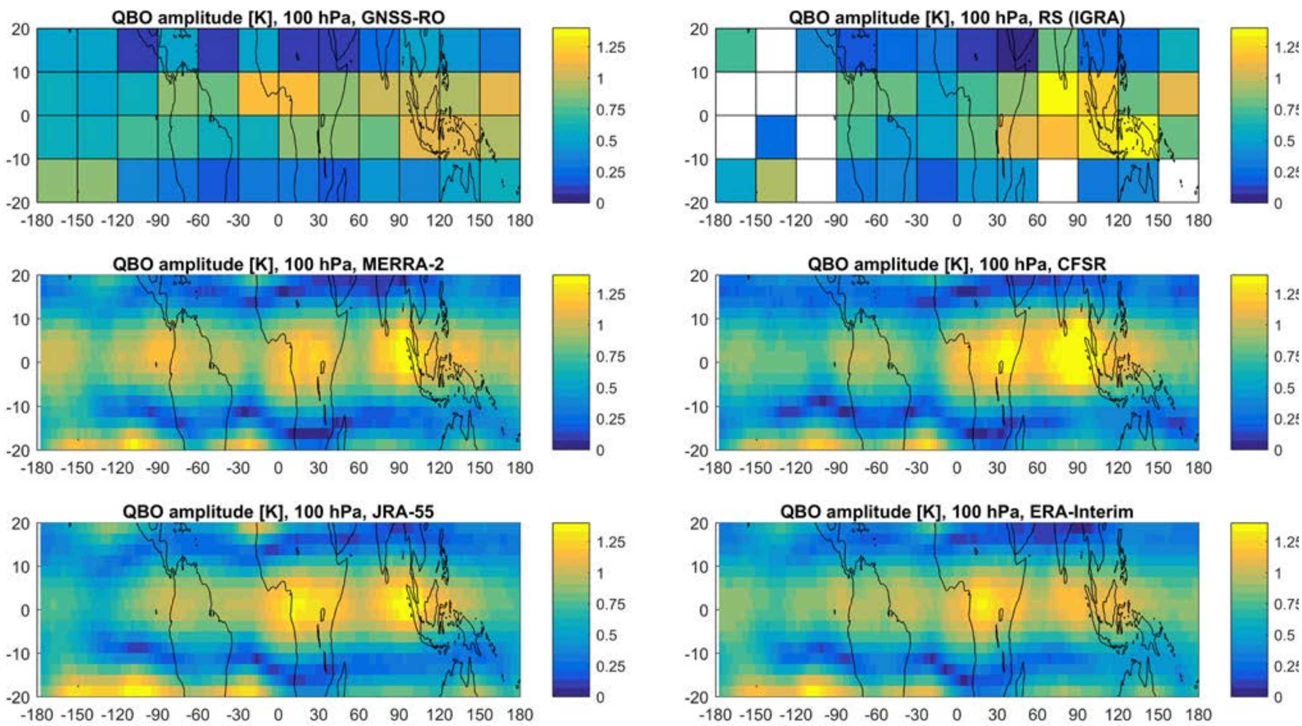


***Figure 9.29:** As **Figure 9.28**, but for 100 hPa QBO peak-to-peak amplitude.*

9.3 Tropical waves and QBO forcing

The QBO is forced by a wide spectrum of waves (*Baldwin et al.*, 2001). This section focuses on the tropical stratospheric waves. The tropical waves include equatorially trapped Kelvin and mixed Rossby-gravity (MRG) waves and Rossby waves, of which the time scales are generally longer than ≈1 day and zonal scales are larger than ≈3000 km (*Kiladis et al.*, 2009; *Matsuno*, 1966), and gravity waves on a wide range of scales (from mesoscales to planetary scales). These waves are known to be generated primarily by tropical convection and, when they propagate into the stratosphere, they interact with the QBO (*e.g.*, *Yang et al.*, 2011, 2012). In particular, the tropical waves provide the mean flow with the momentum required to drive the QBO where the waves dissipate.

General circulation models have not converged in simulating the tropical waves in terms of their spectral characteristics and amplitudes (*Lott et al.*, 2014; *Horinouchi et al.*, 2003). This might also lead to the diversity of simulated characteristics of the QBO among models (*e.g.*, *Schenzinger et al.*, 2017). Therefore, it must be worth investigating quantitatively the characteristics of tropical waves represented in various reanalyses. This section includes the analysis of the spectral characteristics of two prominent equatorial wave modes, Kelvin and MRG waves (*Section 9.3.1*). The momentum budget of the QBO including the forcing by the tropical waves represented in reanalyses is estimated, and its range and spread among the reanalyses are investigated (*Section 9.3.2*).

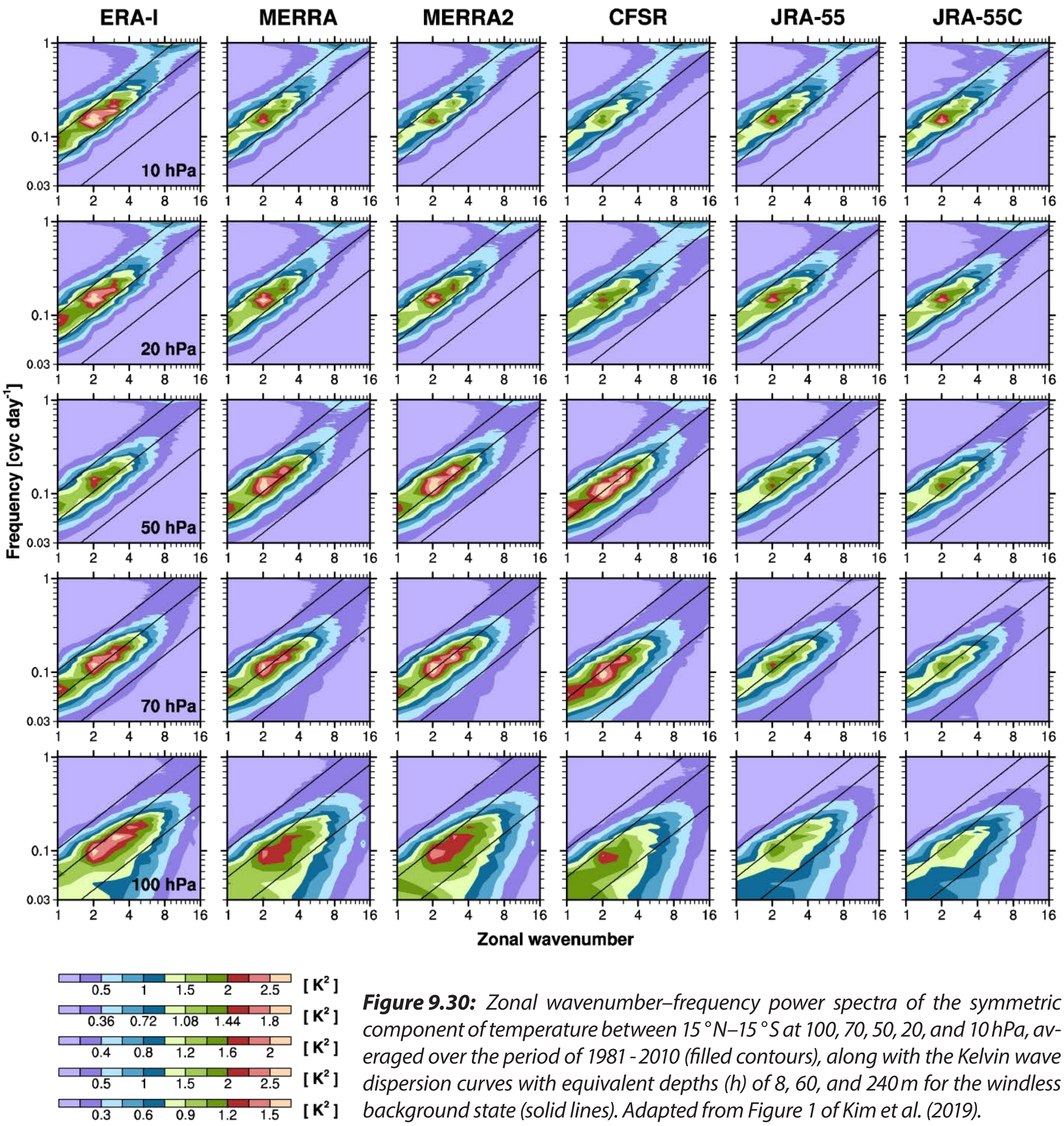


Figure 9.30: *Zonal wavenumber–frequency power spectra of the symmetric component of temperature between 15 °N–15 °S at 100, 70, 50, 20, and 10 hPa, averaged over the period of 1981 - 2010 (filled contours), along with the Kelvin wave dispersion curves with equivalent depths (h) of 8, 60, and 240 m for the windless background state (solid lines). Adapted from Figure 1 of Kim et al. (2019).*

The reanalyses have horizontal resolutions that are capable of resolving the large-scale waves (*i.e.*, Kelvin, MRG, and Rossby waves). Previous studies reported that the large-scale waves in assimilated fields showed a qualitatively good agreement with those observed by satellite measurements (*e.g.*, *Ern et al.*, 2008). On the other hand, representation of smaller-scale gravity waves in reanalyses might be rather challenging. Due to the spatial and temporal scales sampled by the satellite and conventional data that constrain the reanalyses, it is possible that smaller scale waves are unrealistically represented (see *Chapter 2* for model resolutions). Features near the smallest scales in atmospheric models may be affected by diffusion imposed for reasons of numerical stability, and it is also common practice in data assimilation to filter out gravity wave activity so as to remove spurious waves that are excited by insertion of the assimilation increments (*Polavarapu et al.*, 2005). It may be the case that some of the gravity wave activity removed is not spurious. The gravity wave spectrum also likely has some dependence on the forcing mechanisms of the waves, which might not be realistically represented (such as convective heating in the tropics, which is not directly constrained by data assimilation). In any case, even the highest resolution reanalysis forecast models are not expected to fully resolve the gravity wave spectrum (*Alexander et al.*, 2016). The last part of this section includes comparison of gravity waves in the reanalyses and satellites (*Section 9.3.3*).

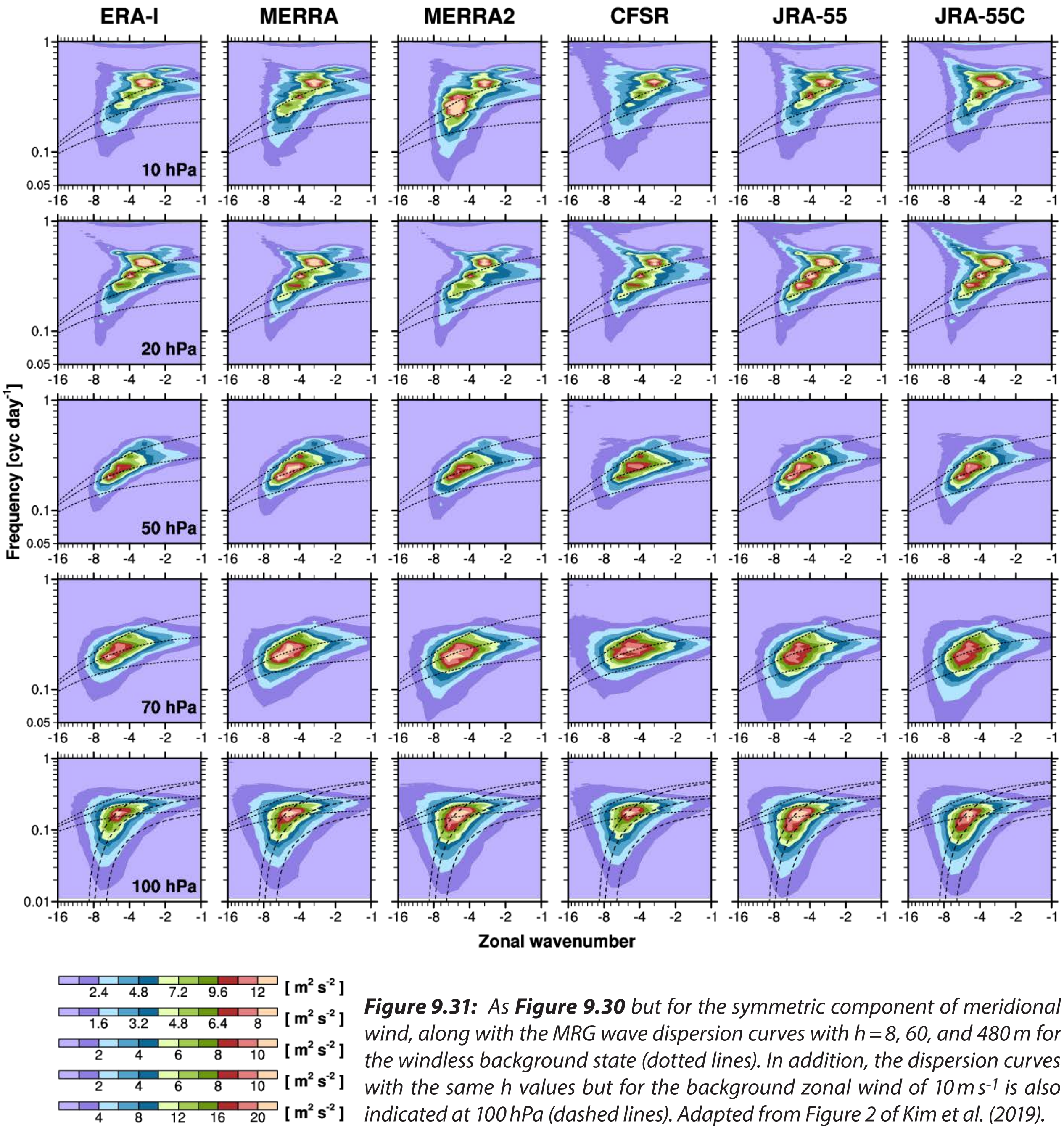


Figure 9.31: *As **Figure 9.30** but for the symmetric component of meridional wind, along with the MRG wave dispersion curves with h = 8, 60, and 480 m for the windless background state (dotted lines). In addition, the dispersion curves with the same h values but for the background zonal wind of 10 m s*$^{-1}$ *is also indicated at 100 hPa (dashed lines). Adapted from Figure 2 of Kim et al. (2019).*

9.3.1 Characteristics of the equatorial waves

The spectra of the equatorial waves are compared using six recent reanalyses (ERA-Interim, MERRA, MERRA-2, CFSR, JRA-55, and JRA-55C) for the period of 1981 - 2010 in *Kim et al.* (2019). In that study, the spectra are calculated at each latitude for each month using a 90-day window centered on the target month, then averaged over the latitude band of 15 °N - 15 °S. Details of the method can also be found in *Section 8.6.2* of this report, where the same method is used to analyse the equatorial waves in the TTL. The spectra are presented in log-log form, which tends to accentuate features at lower zonal wavenumbers and lower frequencies (*i.e.*, larger spatial scales and longer periods). **Figure 9.30** shows zonal wavenumber–frequency (k - ω) power spectra of the eastward-propagating ($k > 0$) latitudinally symmetric component of temperature at 100, 70, 50, 20, and 10 hPa, averaged over the period of 1981 - 2010. (Note that the 100 hPa spectra were also shown in *Chapter 8* and are included here for convenience of comparison with the other levels.) The symmetric temperature shows a good agreement in its spectral shapes among the reanalyses: major portions of the spectral powers are located between the dispersion curves of Kelvin waves with equivalent depths (h) of 8 m and 240 m (zonal phase speeds of ≈ 9 m s^{-1} and 48 m s^{-1}; **Fig 9.32**) at all altitudes, with peaks at $k = 2$ - 3. In all reanalyses the spectra shift to higher h (larger zonal phase speed, c) with increasing altitudes. A difference is found in the detailed shapes of the spectra: the spectral peaks at 70 hPa and 50 hPa in CFSR occur at slightly lower h than those in the others. In general CFSR at lower altitudes shows larger low-frequency power than the other reanalyses, most especially at 100 hPa. In the spectral magnitudes a notable difference exists between JRA-55 (and JRA-55C) and the other four reanalyses in that JRA reanalyses have smaller spectral powers at all altitudes below 20 hPa. The differences in the spectral shapes and magnitudes between JRA- 55 and JRA-55C seem small, compared to the inter-reanalysis differences among the others. The similarity between the two JRA reanalyses suggests that satellite observations are not essential for reanalyses to represent tropical stratospheric waves, at least for the JRA-55 reanalysis system, at these altitudes, and for symmetric temperature spectra. However, further quantitative comparison of JRA-55 and JRA-55C, below, will indicate where some differences do emerge as a result of satellite data assimilation.

Figure 9.31 shows k - ω spectra of the westward-propagating ($k < 0$) latitudinally symmetric component of meridional wind. The spectral shapes seem similar in general among the reanalyses. They exhibit spectral peaks at the same ranges of k and ω on the MRG wave dispersion curves with $h \approx 60$ m at 70 - 50 hPa and $h > 60$ m at 20 - 10 hPa. A slight difference is found at 70 hPa between CFSR and the others: the 70-hPa symmetric meridional wind spectrum in CFSR has a larger power than in the others outside the spectral region of the MRG wave dispersion curves, around $k \approx -10$, $\omega \approx 0.3$ cycle day^{-1}. This spectral component has substantial powers at 100 hPa in all reanalyses, but it tends to be filtered out above 100 hPa so that at 50 hPa most of the spectra are located within the region surrounded by the MRG wave dispersion curves with $h \approx 8$ - 480 m. At 20 - 10 hPa, a spectrum distinct from that of the MRG waves observed at lower altitudes is found in all reanalyses, which has a large power at $k \approx -4$, ≈ 0.5 cycle day^{-1} and extends to $k \approx -8$, $\omega \approx 0.7$ cycle day^{-1}. *Kim et al.* (2019) showed that this spectrum appears only when the monthly and zonal mean wind is easterly with substantial magnitudes (*e.g.*, faster than about -20 m s^{-1} at 20 hPa) and thus when the upward propagating MRG waves may not reach these altitudes due to wave dissipation by the easterly QBO wind below. Understanding of these high-frequency waves in the symmetric meridional wind spectrum may require future studies, and therefore in this section the inter-reanalysis comparison of the MRG waves will be continued only for the waves with $\omega < 0.33$ cycle day^{-1}.

In the wind spectra shown in **Figure 9.31** the JRA reanalyses have similar power to the other reanalyses, but the corresponding antisymmetric temperature spectra (not shown) indicate lower power in the JRA reanalyses, as was seen in **Figure 9.30**. Evidently the balance between wind and temperature perturbations differs between JRA and the other reanalyses, possibly indicating issues in JRA-55 with the assimilation of temperature.

Figure 9.32 presents zonal phase-speed (c) power spectra of the symmetric component of temperature filtered for $k = 1$ - 10 and periods (τ) of 2 - 20 days in 15 °N - 15 °S. The dashed and solid lines present the results calculated from standard pressure-surface datasets and model-level datasets, respectively. The power spectra obtained on model levels are interpolated to the standard pressure levels for comparison. Due to data availability, model-level results of MERRA are not included. The symmetric temperature spectra have a peak at $c = 12$ - 14 m s^{-1} at 100 hPa, and the peak shifts to higher c with increasing altitudes. While the spectral powers at relatively high phase speeds increase with height due to decrease of density, those at low phase speeds decrease primarily by the radiative dissipation below the westerly jet of the QBO (see **Figure 9.35**). These features in the spectra are found commonly in all reanalyses, although CFSR exhibits remarkably larger spectral powers at low phase speeds compared to the other reanalyses above 100 hPa. This is consistent with weaker filtering of slow phase speeds due to the weaker QBO amplitude in CFSR (**Figure 9.19**). The inter-reanalysis spread of the peak magnitudes is, for the results from the model-levels datasets, 22 % at 100 hPa and between 25 % and 37 % in the lower stratosphere with the largest (smallest) spread at 10 hPa (20 hPa). Here, the spread is defined as the difference between the largest and smallest peak values among the reanalyses except JRA-55C, relative to the ensemble average of the peak values.

The magnitudes of spectra obtained using the standard pressure-surface datasets are smaller than those using the model-level datasets by about 10 - 35%, except at 100 hPa in MERRA-2 (**Figure 9.32**). The degree of the underestimation in the standard pressure-surface results differs depending on the altitudes, phase speeds, and reanalyses (for more details, see *Kim et al.*, 2019). The underestimation is due to vertical interpolation of fields from the native model levels to the standard pressure surfaces. The significant amount of amplitude reduction by the interpolation implies that the number of model levels per vertical wavelengths of the assimilated equatorial waves is too small to accurately capture vertical gradients associated with the waves. The exception for MERRA-2 at 100 hPa, for which the reduction in the spectral power is less than 5%, is due to the fact that MERRA-2 has a model level that is very close to 100 hPa (*Appendix A2.6 of Chapter 2*), which minimizes the effect of interpolation to 100 hPa.

As found also in **Figure 9.30**, JRA-55 and JRA-55C exhibit relatively smaller temperature amplitudes than the others throughout the lower stratosphere below 20 hPa (**Figure 9.32**). Such systematically smaller amplitudes in these reanalyses are also found in the anti-symmetric temperature spectra but not in the symmetric/antisymmetric wind spectra (not shown). The distinct shape of the phase-speed spectrum in CFSR with larger powers at low phase speeds shown in **Figure 9.32**, compared to that in the other reanalyses, is conspicuous until the late 1990s. Afterward, the spectrum at low phase speeds becomes less emphasized than before, although it is still larger than that in the other reanalyses (not shown). This is consistent with the QBO amplitude in CFSR gradually strengthening over the 1981 - 2010 period.

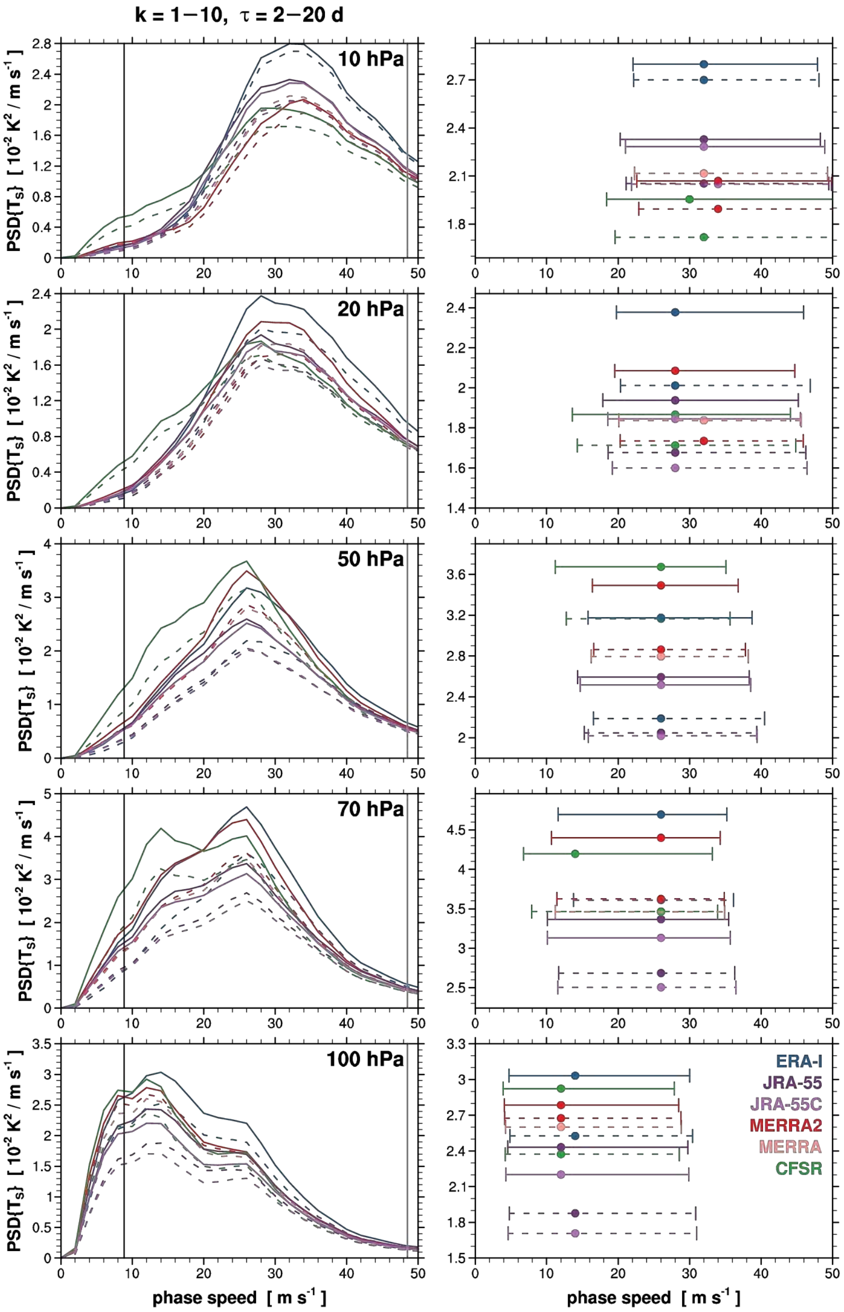


Figure 9.32: *(left) Zonal phase-speed power spectra of the symmetric component of temperature in 15°N - 15°S filtered for zonal wavenumbers up to 10 and periods of 2 - 20 days at 100, 70, 50, 20, and 10 hPa and averaged over the period of 1981 - 2010. (right) The peak phase speeds and magnitudes (dots) and the ranges between the half maxima of the spectra (horizontal lines). Dashed and solid lines indicate the results from the standard pressure-surface datasets and model-level datasets, respectively. The black and grey vertical lines in the left panel indicate the zonal phase speeds corresponding to h = 8 and 240 m, respectively.*

Figure 9.33 presents zonal phase-speed spectra of the symmetric component of meridional wind filtered for the spectral domain of k from -1 to -10, $\tau > 3$ days, and $h > 8$ m. The criterion for τ is applied to exclude the high-frequency waves at 20 - 10 hPa shown in **Figure 9.31**. The spectral shapes of the MRG waves are generally similar among the reanalyses below 20 hPa. The spectra tend to be broader with increasing altitudes up to 20 hPa, although the peak phase speed seems not to shift significantly with height, as shown in the right panels of **Figure 9.33**. The spectral magnitudes tend to be large in CFSR and MERRA-2 at all altitudes, compared to those in the others. At 10 hPa, MERRA-2 exhibits exceptionally large spectral powers at $-30 < c < -10$ m s^{-1}. These large powers are found only before 1998 (see **Figure 9.34**).

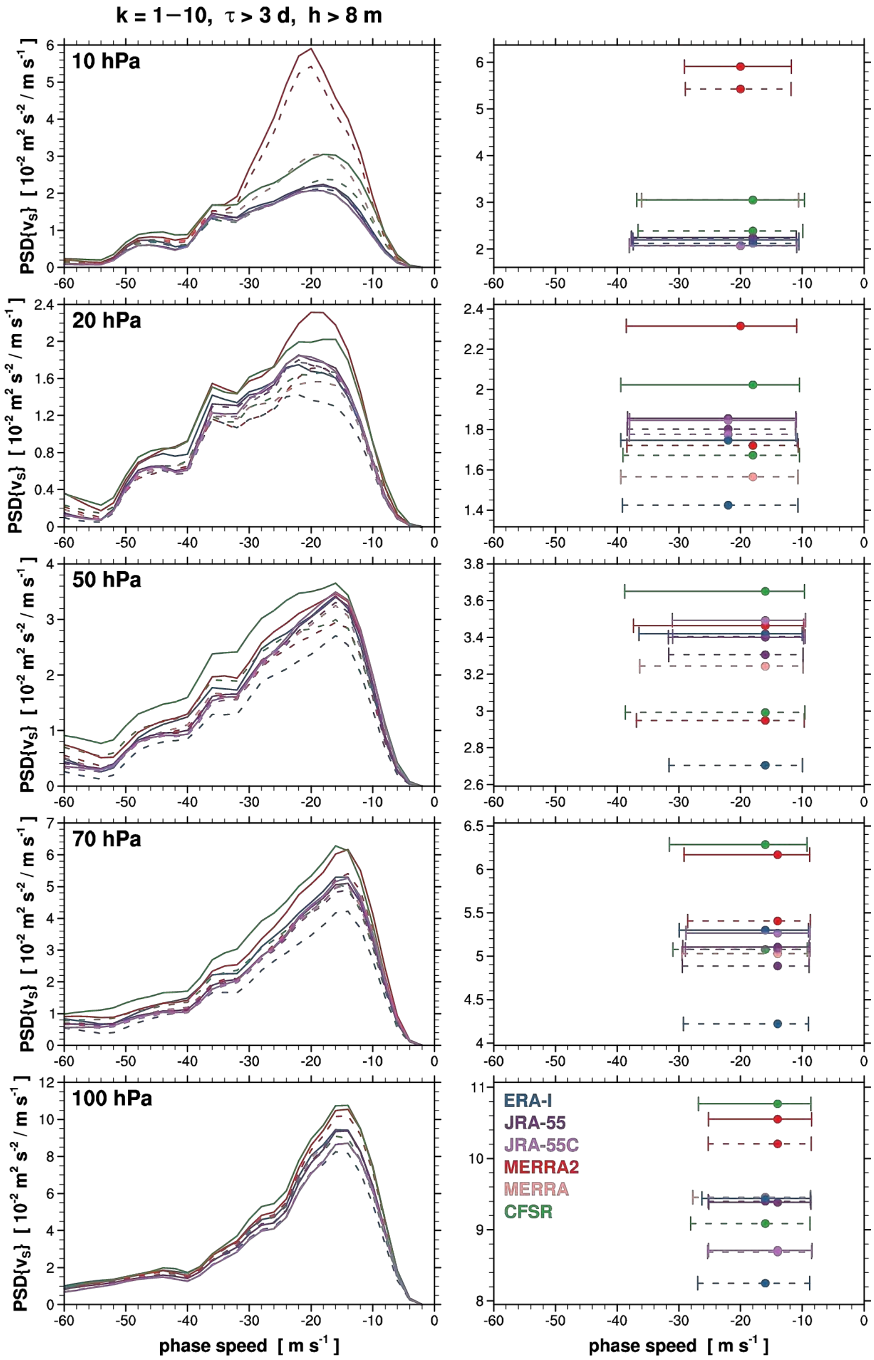


***Figure 9.33:** As **Figure 9.32**, but for the symmetric component of meridional wind filtered for zonal wavenumbers up to 10, periods larger than 3 days, and h > 8 m.*

As mentioned above, the two anomalous features shown in **Figures 9.32** and **9.33** – the emphasized amplitudes of the symmetric temperature at low phase speeds in CFSR, and the exceptionally large amplitudes of the 10 hPa symmetric meridional wind in MERRA-2 – are prominent until the late 1990s and become suppressed or disappear afterward. In addition to these, *Kim et al.* (2019) identified a systematic change around 1998 in the Kelvin wave amplitudes in JRA-55 when compared to JRA-55C. Given that there was a major transition of satellite instruments in 1998 from the TOVS to ATOVS suites (see *Chapter 2*), these changes might be a reflection of impacts of the different satellite data on the assimilated equatorial wave fields.

For the two periods before and after 1998, vertical profiles of the temperature (meridional wind) variances by the Kelvin (MRG) waves are shown in **Figure 9.34**. Here, the Kelvin waves are defined as the modes with $h = 8$ - 240 m among the spectral components shown in **Figure 9.32** (refer to the black and grey vertical lines in the left panels of **Figure 9.32** for the phase speeds corresponding to $h = 8$ m and 240 m, respectively), and the MRG waves as the same spectral components as in **Figure 9.33**. The Kelvin wave variances show a large difference between the two periods, in particular in the five reanalyses except JRA-55C. The variances increase by 16 - 19 % at 100 hPa and more than 25 % in the middle stratosphere in most reanalyses. The inter-reanalysis spread also increases significantly around 1998. On the other hand, JRA-55C exhibits relatively small changes in the Kelvin wave variances at 100 hPa and above 10 hPa, although the changes at ≈ 30 hPa reach 20 %. From comparison between JRA-55 and JRA-55C, it may be concluded that a large portion of the change in the variance at 100 hPa and above 10 hPa in JRA-55 around 1998 comes from the assimilation of the different satellites in the two periods. It might be also possible for the Kelvin wave variances in the other reanalyses to be affected by the satellite transition.

In addition, in both periods the underestimation of the Kelvin wave variances from the standard pressure-surface fields (dashed), compared to those using the model-level fields (solid), is generally large in the TTL and the lowermost stratosphere (100 - 50 hPa). In this layer, abrupt vertical changes exist in the static stability as well as in the amplitudes and vertical scales of Kelvin waves (*Randel and Wu*, 2005). These may suggest that use of finer vertical resolutions around the TTL and the lowermost stratosphere in forecast models might benefit the representation of equatorial waves in future reanalyses.

The MRG wave variances in the two periods (**Figure 9.34**) show a large change in ERA-Interim (more than 30 % at most altitudes), whereas the changes are small in JRA-55 and JRA-55C. As mentioned above, MERRA-2 exhibits exceptionally large MRG wave variances at 10 hPa and above before 1998, which become comparable to those in the other reanalyses afterward. In addition, CFSR exhibits a vertical fluctuation in the MRG wave variances for the model-level results before 1998 (**Figure 9.34**). This feature disappears from 1999. These results may indicate that assimilation of the ATOVS suite in the latter period helps to better constrain the wave fields in MERRA-2 and CFSR.

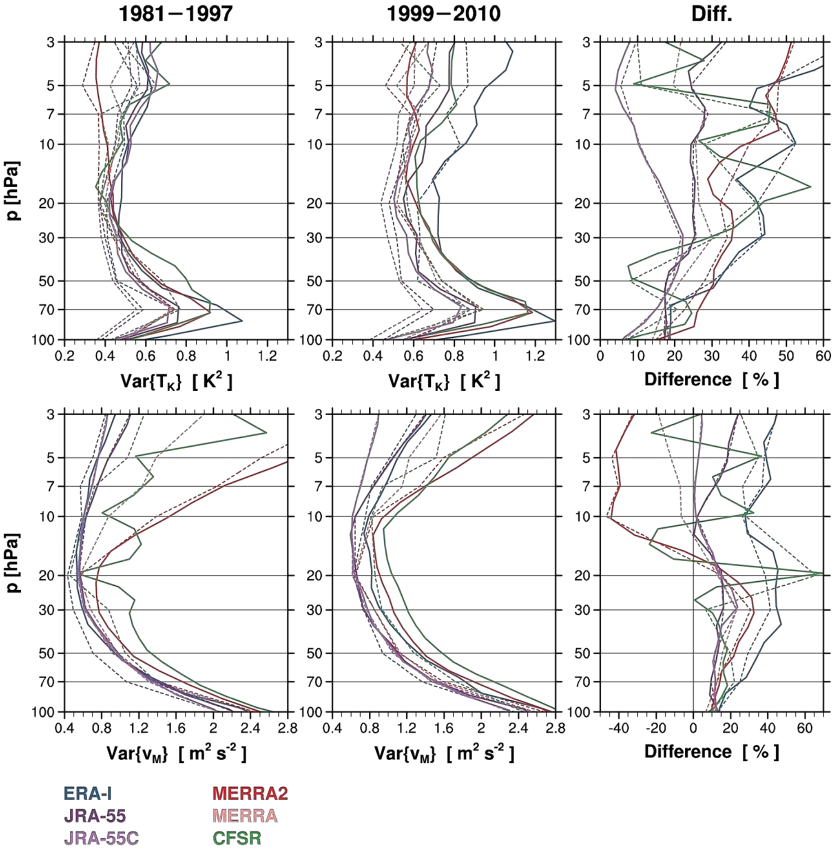


***Figure 9.34:** Vertical profiles of variances of (top) temperature filtered for the Kelvin waves and (bottom) meridional wind filtered for the mixed Rossby-gravity (MRG) waves, averaged over the periods of (left) 1981 - 1997 and (center) 1999 - 2010, and (right) their differences. The datasets used are the same as in **Figures 9.32** and **9.33**. Adapted from Figure 8 of Kim et al. (2019).*

The reanalysis representation of Kelvin and MRG wave interactions with the QBO is investigated through the spectra of EP flux and its divergence (EPD). The k - ω spectra of the EP flux are calculated for the symmetric and anti-symmetric wave modes in 15 °N - 15 °S in a similar way with **Figures 9.30** and **9.31**. The k - ω spectra of the vertical component of Eliassen-Palm flux for these wave modes, averaged over the period of 1981 - 2010, are included in *Appendix A9.1* (**Figures AS9.15** and **AS9.16**). The spectral shapes of the EP flux for the symmetric and anti-symmetric modes are broadly similar to those of the symmetric temperature and meridional wind, respectively.

Figure 9.35 shows vertical profiles of phase-speed spectra of the vertical EP flux and EPD by the Kelvin and MRG waves in ERA-Interim, MERRA-2, JRA-55 and JRA-55C, composited for four selected phases of the QBO during 1981 - 2010 (*Kim et al.*, 2019). The two wave modes are defined as before, and only the model-level results are presented. CFSR and MERRA are not included since one (vertical velocity) or more fields required for the EP flux calculation are not available on the model levels. In the QBO phase of westerly shear at 20 hPa (**Figure 9.35**, first row), the largest EPD in the shear layer by the Kelvin waves is found at similar phase-speed ranges but at slightly different altitudes among the reanalyses (*e.g.*, at 18 hPa and 12 hPa in MERRA-2 and JRA-55, respectively). At 15 hPa, a major portion of the EPD occurs at c = 17 - 28 m s^{-1} in the reanalyses. These phase speeds are roughly 10 - 20 m s^{-1} larger than the mean zonal wind speed at 15 hPa, which is consistent with radiative dissipation of Kelvin waves (*Ern and Preusse*, 2009). In the phase of westerly shear at 40 - 50 hPa (**Figure 9.35**, second row), the Kelvin wave dissipation at 40 hPa occurs at similar phase speeds to those at 15 hPa in the 20-hPa shear phase. The magnitudes of EPD by the Kelvin waves are largest in ERA-Interim. The EP flux and EPD spectra and their vertical evolution in JRA-55C are overall similar to those in JRA-55 with slightly smaller magnitudes. In the lowermost stratosphere, similar to the Kelvin waves, the MRG waves dissipate mostly at the phase speeds 10 - 25 m s^{-1} larger than the mean wind (**Figure 9.35**, the third and last rows), while at higher altitudes the waves appear to encounter critical levels. Above 70 hPa, the magnitudes of the EP flux and EPD by the MRG waves are similar among the reanalyses. The overall forcing by MRG waves is weaker than the Kelvin wave forcing by roughly a factor of five (note the different contour scales used for the two wave types).

9.3.2 Momentum budget of the QBO

While the method to identify the equatorial wave modes used in *Section 9.3.1*, which assigns the ranges of k, ω, and h to each wave mode, is simple and useful to investigate characteristics of the waves, it is less well suited to assessing the momentum budget by the waves. In relatively low frequency ranges or with time-varying background flows (*e.g.*, when a QBO phase is changing), it is ambiguous to separate the Kelvin and MRG waves from Rossby waves using this method since they can share some parts of the spectral components. *Kim and Chun* (2015) used another method to decompose the momentum budget contribution from each of the wave types as represented in four reanalyses.

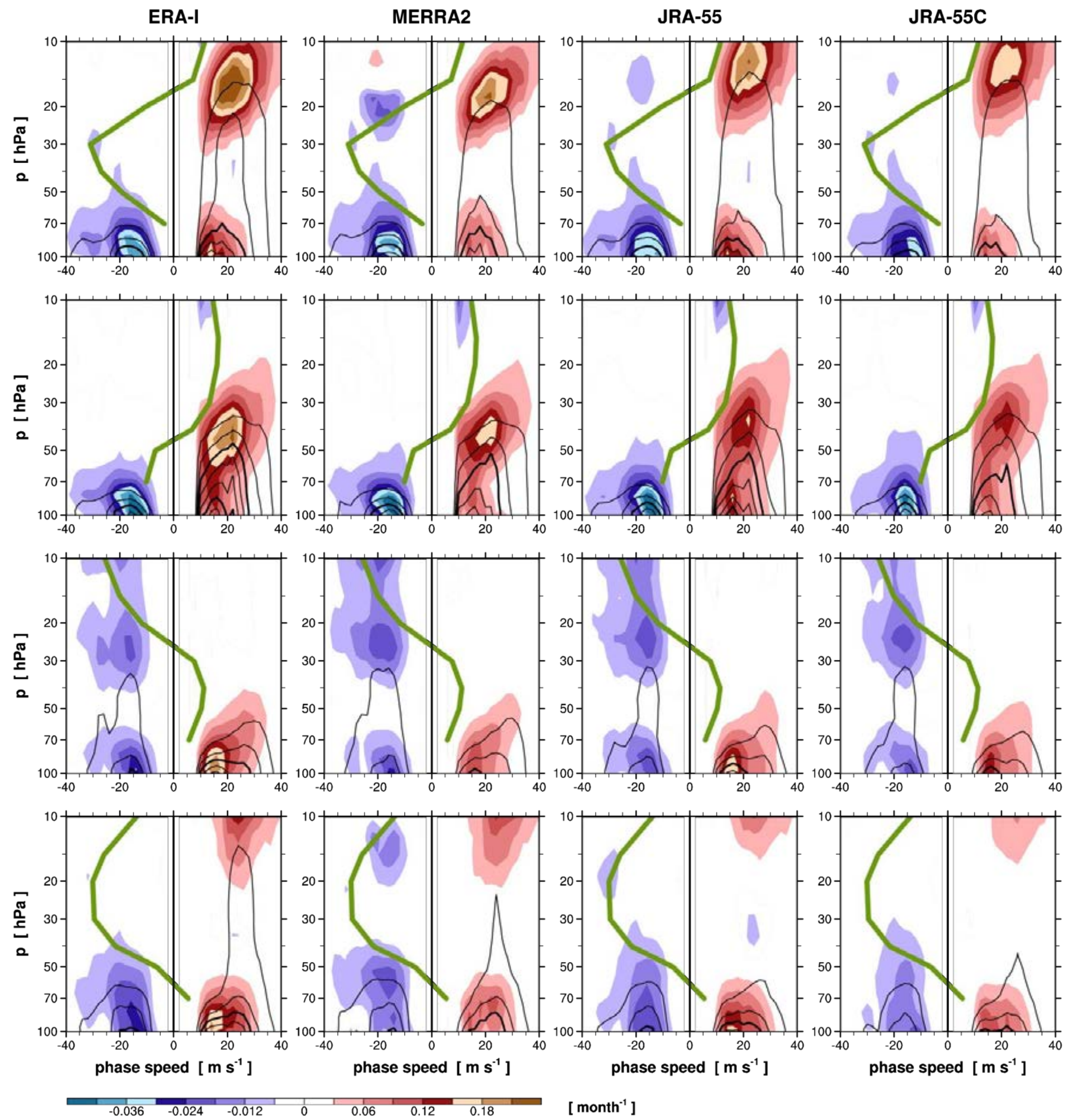


***Figure 9.35:** Vertical profiles of zonal phase-speed (c) spectra of the EP flux divergence (shading) and vertical component of EP flux (black contour) for Kelvin waves at c > 0 and MRG waves at c < 0, composited for the four QBO phases. The four QBO phases selected are maximum westerly tendency at 20 hPa and 50 hPa (first and second rows, respectively) and maximum easterly tendency at 20 hPa and 50 hPa (third and fourth rows, respectively). The FUB zonal wind profiles are also indicated for the four composites (green contour). The contour intervals for the Kelvin and MRG wave EP fluxes are 5 and 0.5 × 10^{-3} mPa/(m s^{-1}), respectively. Adapted from Figure 11 of Kim et al. (2019).*

This method identifies the Kelvin and MRG waves based on the polarization relation of the equatorial waves and the contrast in the characteristics of divergent/rotational modes (for details see *Kim and Chun*, 2015). After the Kelvin and MRG waves are identified and excluded from the total perturbations, the remaining component of the perturbations is decomposed into inertio-gravity (IG) waves, for $|k| > 20$ or $\omega > 0.4$ cycle day^{-1}, and Rossby waves otherwise.

Extending the results of *Kim and Chun* (2015) to include additional datasets, **Figure 9.36** shows monthly time series of the zonal momentum forcing by Kelvin, MRG, IG, and Rossby waves averaged over 5 °N - 5 °S at 30 hPa from 1981 to 2010. For ERA-Interim, MERRA-2, and JRA-55, model-level fields are used, and for MERRA and CFSR, the standard pressure-surface fields are used due to the data availability. Note that the results using the standard pressure-surface fields often show similar time evolutions but with smaller magnitudes of the forcing to those from the model-level fields (for example in the Kelvin and IG panels; for a comparison of model levels and pressure levels results using ERA-Interim, see **Figure 3** of *Kim and Chun*, 2015). The Kelvin waves exert large forcing during the easterly-to-westerly (E-W) transition phases with peak magnitudes of 5 - 13 m s^{-1} $month^{-1}$, where the E-W transition phases are defined as the period from maximum easterly to maximum westerly phases of the FUB wind (top panel of **Figure 9.36**; maximum E and W phases are indicated by dashed and solid vertical lines, respectively, in all panels).

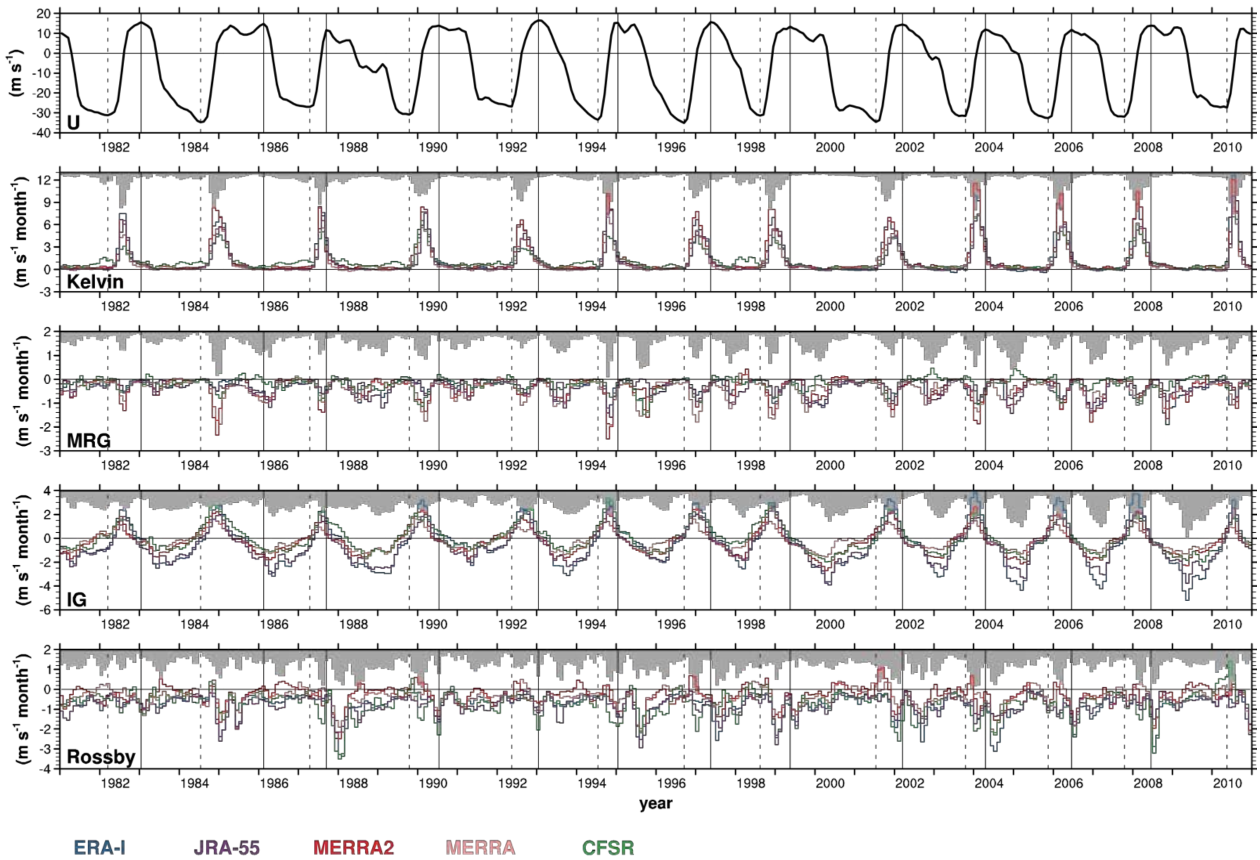


Figure 9.36: *EP flux divergence averaged over 5°N-5°S at 30hPa for the Kelvin, MRG, inertio-gravity, and Rossby waves in ERA-Interim, JRA-55, MERRA-2, MERRA, and CFSR, along with the FUB zonal wind. The results for ERA-Interim, JRA-55, and MERRA-2 are obtained using the model-level fields, and those for MERRA and CFSR are using the standard pressure-surface fields. The grey shading indicates the inter-reanalysis spread (maximum minus minimum) of each forcing. Note the different y-axis ranges for the different wave types. The months of the maximum easterly (westerly) phases are indicated by dashed (solid) lines. See text for details.*

The Kelvin wave forcing tends to tail off but remain non-zero as the wind reaches its W maximum, consistent with radiatively damped waves that do not meet critical levels. ERA-Interim and MERRA-2 tend to have larger Kelvin wave forcing than the others after around the year 2000, while before 2000 their forcing seems rather comparable to JRA-55 and MERRA. CFSR exhibits quite weak forcing in some years during the early period (1982, 1992, and 1994). In the westerly-to-easterly (W-E) transition phases (bracketed by solid and dashed vertical lines on the left and right in all panels of **Figure 9.36**), the Kelvin wave forcing is near zero, except in CFSR until 1998. This is another indication of the impact of the TOVS–ATOVS transition revealed in CFSR. The inter-reanalysis spread (grey shading) of the Kelvin wave forcing reaches $\approx 5\,\mathrm{m\,s^{-1}\,month^{-1}}$. The MRG wave forcing peaks during both transition phases. In general, magnitudes of MRG wave forcing are small (usually less than $\approx 2\,\mathrm{m\,s^{-1}\,month^{-1}}$) in all reanalyses, and their spread is comparable to the typical magnitudes of the forcing (0.5 - $2\,\mathrm{m\,s^{-1}\,month^{-1}}$). The forcing tends to be relatively large in MERRA-2 and MERRA compared to that in the others during the E-W transitions, whereas during the W-E transitions the relative magnitudes of the MRG wave forcing are not consistent among the reanalyses.

The IG wave forcing in **Figure 9.36** shows a very clear QBO variation. Peak eastward and westward forcing occurs during E-W and W-E transitions, respectively, and near the E and W maxima (dashed and solid vertical lines) the forcing is close to zero. This behaviour is consistent with waves that meet critical levels in the flow. The westward IG wave forcing tends to be larger in ERA-Interim and JRA-55 than in the others, with peak magnitudes of 2 - $5\,\mathrm{m\,s^{-1}\,month^{-1}}$, while the eastward forcing seems comparable among ERA-Interim, JRA-55, MERRA-2, and CFSR before around 2000. Although the IG waves include the smallest horizontal scales and highest frequencies ($|k| > 20$ or $\omega > 0.4$ cycle day^{-1}) no obvious relation is seen in **Figure 9.36** between reanalysis horizontal resolution and IG forcing strength; one reason for this could be that the five reanalyses shown all have similar resolutions[11]. Their horizontal resolutions are nevertheless higher than those often used in climate models (*e.g., Bushell et al.*, 2020) and hence it would be expected that some of the IG forcing seen in **Figure 9.36** would need to be parameterized in those models.

[11] From **Table 2** of *Fujiwara et al.* (2017) and **Table 2.2** of *Chapter 2*, the approximate grid spacings corresponding to forecast model resolutions are 79km (ERA-Interim), 55km (JRA-55), 74km (MERRA), 70km (MERRA-2), and 35km (CFSR).

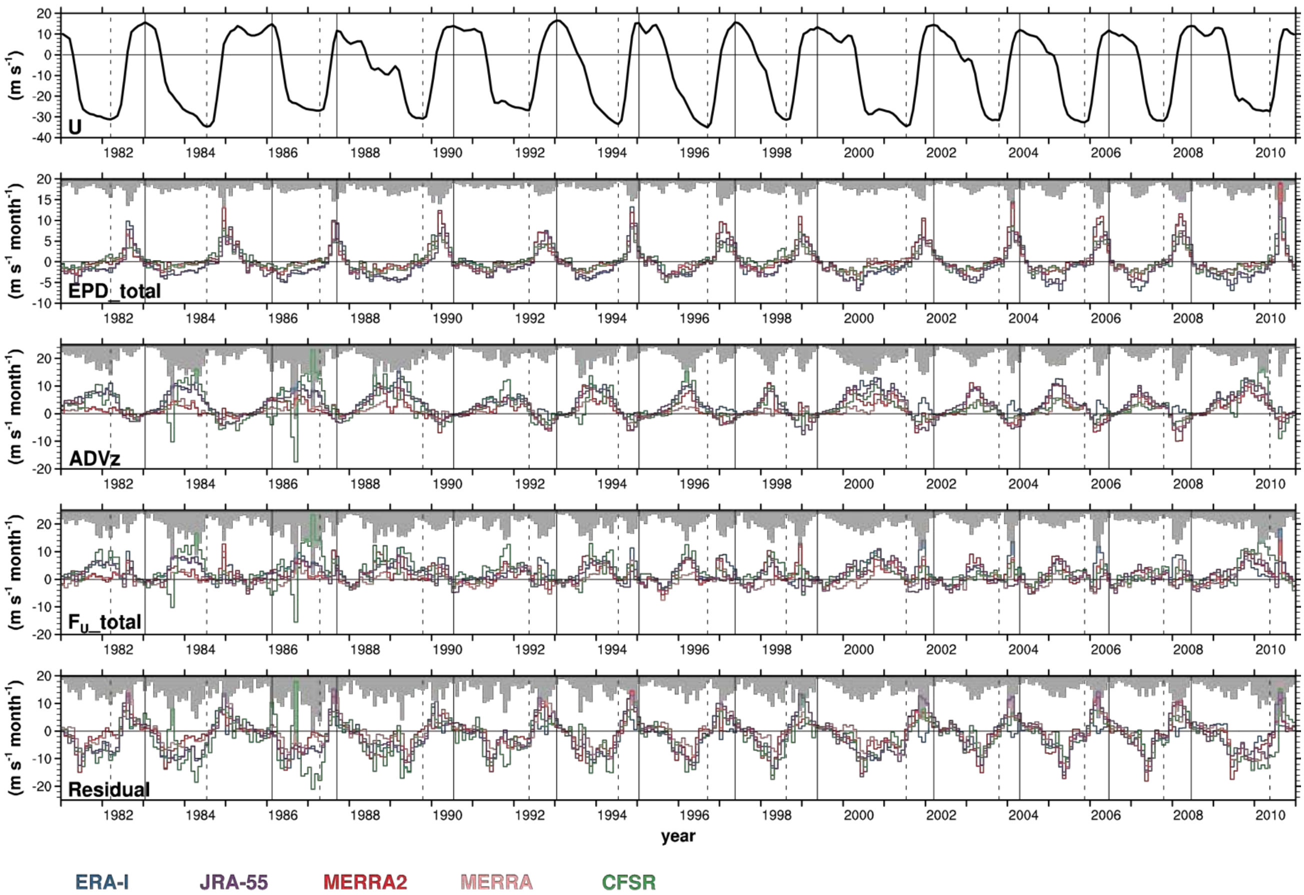


***Figure 9.37:** As **Figure 9.36**, but for the EP flux divergence from the total wave fields (EPD total), vertical advection of zonal wind (ADVz), total forcing from all resolved fields (F_u_total, i.e., sum of EPD, ADVz, meridional advection, and the Coriolis force), and residual of the zonal momentum budget.*

After 2000, ERA-Interim shows the largest IG wave forcing in both phases, which might be suggestive of an impact of the satellite transition on the IG waves assimilated in ERA-Interim. The inter-reanalysis spread is 1 - 2 m s^{-1} month^{-1} until 1998, and afterward it becomes larger during both transition phases (2 - 4 m s^{-1} month^{-1}) but smaller during the maximum westerly and easterly phases (*i.e.*, after 1998 the reanalyses are in better agreement that the IG forcing during wind maxima is close to zero). Finally, the Rossby wave (RW) forcing tends to be large during the solstices and less dependent on the QBO phase. However it does often show largest magnitudes when the FUB wind (top panel) is westerly, as would be expected for stationary ($c=0$) extratropical Rossby waves that cannot propagate into the tropics when tropical winds are easterly (i.e. when a critical line for these waves exists in the subtropics). (Note, however, that other kinds of waves, *e.g.,* nonstationary Rossby waves, could contribute to the RW term since it is defined here as the portion of the wave spectrum remaining after the KW, MRG and IG components are identified and removed.) The peak magnitudes of the RW forcing are ≈ 1 - 3.5 m s^{-1} month^{-1}, and the spread is up to 2 m s^{-1} month^{-1}. While weaker than typical KW forcing values, this is comparable to the size of the MRG and IG forcings.

Figure 9.37 shows time series of EPD calculated from the total perturbations (which is nearly the same as the sum of the four forcings shown in **Figure 9.36** with only negligible differences), vertical advection of zonal wind (ADV_z), total forcing of the zonal momentum from resolved fields (F_u_total, *i.e.*, EPD + ADV_z + ADV_y + COR, where ADV_y and COR are the meridional advection and Coriolis force, respectively), and the residual of the zonal momentum equation in each reanalysis (*i.e.*, zonal-wind tendency minus total forcing). The residual could comprise the zonal averages of parameterized gravity wave drag, analysis increment, and implicit/explicit diffusion in the models. The EPD exhibits time variations following the QBO phases with the same signs as those of the wind tendency. The eastward forcing peaks have much larger magnitudes (5 - 20 m s^{-1} month^{-1}) than the westward forcing peaks (2 - 7 m s^{-1} month^{-1}), due to the significant contribution by the Kelvin waves (**Figure 9.36**). The eastward (westward) forcing tends to be larger in ERA-Interim and MERRA-2 (ERA-Interim and JRA-55) than that in the other reanalyses, consistent with the results of the Kelvin (westward IG) wave forcing shown in **Figure 9.36**.

ADV_z has the opposite signs to EPD in the reanalyses except ERA-Interim, and its magnitudes are large in the W-E transition phases (≈ 10 m s^{-1} month^{-1}). The signs of ADVz during the E-W transition phases in ERA-Interim are sometimes positive, in particular after 2000.

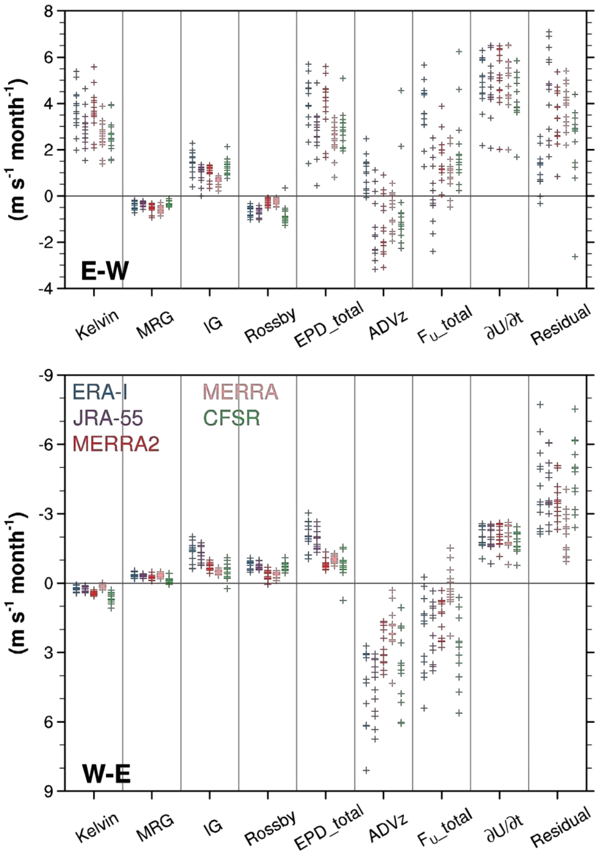


***Figure 9.38:** The eight forcing terms shown in **Figures 9.36** and **9.37** along with the zonal-mean zonal wind tendency in 5 °N - 5 °S at 30 hPa, averaged over the easterly-to-westerly (E-W, upper panel) and westerly-to-easterly (W-E, lower panel) transition phases for each QBO cycle during 1981 - 2010. The datasets used are the same as in **Figures 9.36** and **9.37**. Note that the y-axis directions and magnitudes differ between the upper and lower panels.*

It is because the residual-mean vertical velocity (w^*) in 5 °N - 5 °S becomes negative during these phases in ERA-Interim (not shown) due to the stronger response of the w^* anomaly to the QBO than in the other reanalyses (**Figure 9.25**). MERRA exhibits the smallest ADV_z in most years (**Figure 9.37**), because w^* in MERRA is relatively small (**Figure 9.24**) and also the wind shear is underestimated owing to the use of standard pressure-surface fields. The large fluctuations of ADV_z in CFSR in 1980s are attributed to a very large temporal fluctuation of the mean w^* in CFSR during this period (not shown).

The total forcing from resolved fields, F_u_ total (**Figure 9.37**, fourth row) exhibits positive peaks in both transition phases, dominated by the Kelvin wave forcing in the E-W transition phases and by ADV_z in the other phases. Note that ADV_y and COR are relatively small in the momentum budget of the equatorial lower stratosphere (not shown). The inter-reanalysis spread of the total forcing reaches ≈ 15 m s^{-1} month^{-1} in the E-W transition phases with similar contributions from the spreads of EPD and ADVz. In the opposite phases, the spread is roughly 5 m s^{-1} month^{-1} after 2001, whereas it is much larger in the earlier period.

The lack of westward forcing in the total forcing leads to large magnitudes of the westward forcing in the residual in all reanalyses (≈ 10 - 15 m s^{-1} month^{-1}, **Figure 9.37**, last row). This may imply that small-scale gravity waves that are unresolved or under-represented in the reanalyses could play a major role in the transition of the QBO phases from westerlies to easterlies, consistent with previous modelling and observational studies (*e.g.*, *Ern et al.*, 2014; *Kawatani et al.*, 2010). It is also possible that the current generation of reanalyses could under-represent the MRG and large-scale IG waves that could contribute to the westward forcing required for the QBO evolution, perhaps due to the vertical resolutions of the forecast models and/or observations used in the reanalyses being too coarse in the lower stratosphere to resolve waves with small vertical wavelengths (*Ern et al.*, 2014; *Richter et al.*, 2014). The eastward forcing in the residual has peak magnitudes of ≈ 10 - 15 m s^{-1} month^{-1} in the reanalyses except ERA-Interim. The residual in the E-W transition phases in ERA-Interim has relatively small magnitudes, especially after 2000, due to the positive ADVz in these phases (**Figure 9.37**, the third row) in addition to the relatively large Kelvin and IG wave forcing (**Figure 9.36**). The inter-reanalysis spread is similar to that of the total forcing, given that the mean wind tendency has much smaller spread than the forcing terms.

Figure 9.38 presents a summary of the eight forcing terms shown in **Figures 9.36** and **9.37**, along with the corresponding zonal-mean zonal wind tendency ($\partial\bar{u}/\partial t$ in 5 °N - 5 °S at 30 hPa), averaged over each E-W and W-E transition period (upper and lower panels, respectively) for each QBO cycle during 1981 - 2010. Large inter-cycle variability is evident in most of the forcing terms. In the E-W transition phases, the IG wave forcing resolved in the reanalyses is comparable to the sum of the MRG and Rossby wave forcing with the opposite signs. Due to their cancellation, EPD from all resolved waves has similar magnitude to that of the Kelvin wave forcing. The total forcing (F_u–total) is smaller than EPD by the negative ADVz in the four reanalyses except ERA-Interim, whereas in ERA-Interim it is larger than EPD by the positive ADVz (see also **Figure 9.37**). The residual is smallest (largest) in ERA-Interim (JRA-55), ranging between 0 and 2.5 m s^{-1} month^{-1} (1.5 and 7 m s^{-1} month^{-1}).

In the W-E transition phases (**Figure 9.38**), the magnitudes of the residual (1 - 8 m s^{-1} month^{-1}) are generally comparable to those in the opposite phases (except in ERA-Interim). The total forcing is positive (*i.e.*, opposite to the wind tendency) in most cases because of the small resolved wave forcing (EPD) and large ADVz. These yield much smaller zonal-wind tendency in the W-E transition phases, compared to the opposite phases. The large magnitudes of the residual in both phases suggest that a considerable amount of parameterized gravity wave drag may be required in order for the analysis increment to be small (*e.g.*, as in MERRA-2; see *Coy et al.*, 2016; *Molod et al.*, 2015). An important uncertainty in the momentum budget is due to the fields in the reanalyses that are not directly constrained by observations, as indicated by the large inter-reanalysis spread of ADV_z (**Figures 9.37** and **9.38**).

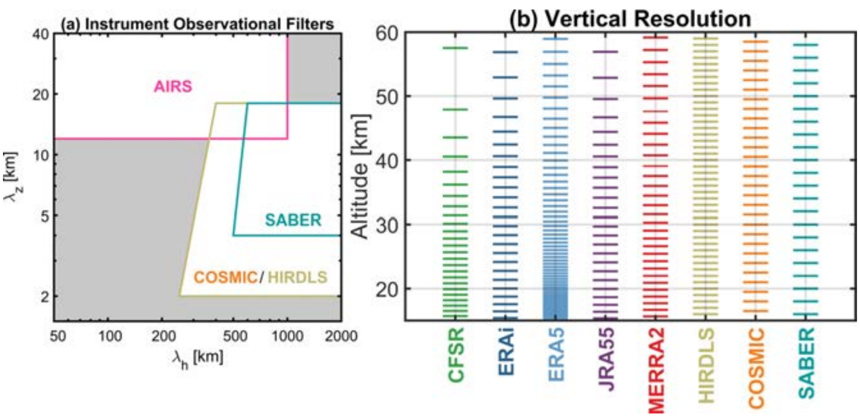


Figure 9.39: *(a) Schematic indicating approximate regions of spectral space (vertical wavelength, λ_z, and horizontal wavelength, λ_h) sampled by different satellite instruments. AIRS lower limit of horizontal resolution varies across the satellite track, but is approximately 50 km. (b) Vertical resolutions of reanalyses and satellite instruments. For COSMIC the exact level locations are arbitrary but the correct vertical resolution is indicated by the figure. Note that the term "COSMIC" is used here instead of the more general term "GNSS-RO" because COSMIC is the specific satellite mission that is examined here (other GNSS-RO observations such as CHAMP and GPS-MET are not examined here).*

9.3.3 Direct comparison with satellite data

Reanalyses are produced by assimilating an enormous variety and volume of observational data (*Fujiwara et al.*, 2017, and *Chapter 2*). The breadth of observations assimilated enables reanalyses to provide a comprehensive estimate of the atmospheric state, but has the drawback that it is difficult to find independent observations against which to validate reanalyses. In this section five modern full-input reanalyses are compared against four satellite observational datasets, two of which (SABER and HIRDLS) are not assimilated by any of the reanalyses.

Different portions of the atmospheric wave spectrum are observed by different satellite measurement techniques (*Alexander et al.*, 2010). **Figure 9.39** indicates schematically the regions of wave spectral space that are sampled by the four satellite datasets considered here: AIRS, COSMIC, HIRDLS and SABER. COSMIC and HIRDLS cover similar vertical and horizontal scales: from about 300–500 km to 2000 km horizontally, and about 2 km to 20 km vertically. These horizontal scales include gravity waves and inertio-gravity waves that contribute to forcing the QBO (*Baldwin et al.*, 2001). In the reanalyses, forcing by resolved waves at these scales was shown in **Figure 9.36** to vary strongly with the phase of the QBO [12]. Waves with fine vertical scales are also expected to be important in forcing the QBO because strong mean-flow shears can refract vertically propagating waves to small intrinsic zonal phase speeds and hence small vertical wavelengths (*Boville and Randel*, 1992). The resolution of SABER is similar to that of HIRDLS except that its finest resolved scales are approximately a factor of two coarser than those of HIRDLS. Finally, AIRS samples a somewhat different region of spectral space than the other instruments, covering finer horizontal scales (roughly 50 km to 1000 km) but coarser vertical scales (roughly 10 km to 40 km). The smallest horizontal scales sampled by AIRS are similar to the horizontal grid spacings of the reanalysis forecast models (see *footnote 11*) and hence at the very limit of waves that can possibly be resolved in the reanalyses.

The limited region of spectral space covered by each satellite instrument means that it is inappropriate to compare the satellite observations to diagnostics that utilize all available wavenumbers and frequencies in the reanalyses (as was done when comparing the reanalyses to each other in *Sections 9.3.1* and *9.3.2*). Here the waves in reanalyses are evaluated by sampling reanalysis temperatures in the same way as would be done by each satellite measurement technique (*i.e.*, to mimic as closely as possible what the satellite instrument would have "seen" if making measurements of the atmosphere as represented by each reanalysis). Wave activity is diagnosed using the gravity wave potential energy (GWPE), defined as $1/2(g/N)^2(T'/T)^2$, where g is the gravitational acceleration, N the buoyancy frequency, T the background temperature and T' the measured temperature anomaly diagnosed using the appropriate method for each kind of satellite measurement technique. For further details of the method, see *Wright and Hindley* (2018).

Figure 9.40 shows the time series of tropical GWPE at 32 hPa for the four different satellite measurement techniques as applied to the reanalyses, as well as the actual observational results. SABER observations (panel **a**, black line) show a clear QBO variation, with GWPE often peaking during E-W transitions (*i.e.,* westerly QBO onsets). Peaks during QBO transitions and minima during QBO phase maxima are consistent with gravity waves acting to force the descent of QBO shear zones, and are reminiscent of the IG forcing time variation seen in **Figure 9.36**. However there is also some tendency for peaks to align with the annual cycle, which could be difficult to separate from the QBO variation in the 15-year record shown here, especially since QBO phase transitions happen to show a strong seasonal alignment during the first half of this period (and note that SABER is the longest of the four satellite records).

[12] Note that the IG component of the QBO forcing shown in **Figure 9.36** was defined by $|k| > 20$ or $\omega > 0.4$ cycle day^{-1}. Since $k = 20$ at the equator corresponds to a horizontal wavelength $\lambda_h \approx 2000$ km, the IG term in **Figure 9.36** would include all horizontal scales indicated for the satellite observations by **Figure 9.39**.

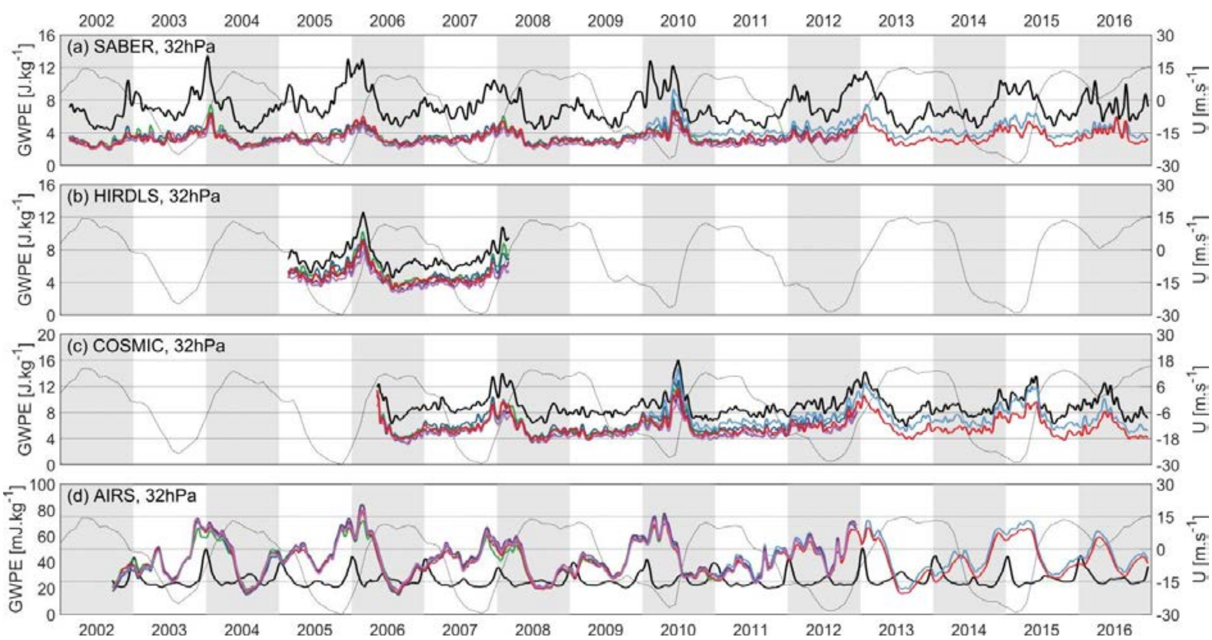


Figure 9.40: *Time series of gravity wave potential energy (GWPE) at 32 hPa, 5 °S - 5 °N in reanalyses and satellite observations, for the reanalyses ERA-Interim, MERRA-2, JRA-55, JRA-55C, CFSR, and ERA5. Actual observations are in black, and other line colours (for reanalyses) are as in* ***Figure 9.39(b)****. Thin grey lines indicate the 30 hPa QBO zonal wind, and alternating grey and white background designates calendar years. Note the y-axis units for AIRS (panel d) are a factor of 10^3 smaller than for the other panels.*

The corresponding GWPE from reaanalyses closely follows the time variations of observed GWPE, but is weaker than it (by about a factor of two) in all reanalyses. This suggests there is excessive damping of reanalysis waves in this spectral region, even though the smallest horizontal scale resolved by SABER, ≈ 500 km or $|k| \approx 80$, is about 6 times larger than the coarsest of the reanalysis gridscales.

The corresponding HIRDLS observational results (**Figure 9.40b**, black line) are unfortunately only available for the three-year period during which HIRDLS was active but show similar variation with QBO phase as the SABER observations. Clear peaks appear during the two E-W transitions, but not during the W-E transitions (except perhaps a hint of one at the start of 2005). The time series appears less noisy than its SABER counterpart during the same 2005 - 2007 period, which might be due to SABER being an older instrument not as well suited to measuring stratospheric temperatures as HIRDLS, with about 60 % larger errors (roughly 0.8 K in SABER vs. 0.5 K for HIRDLS). The reanalyses sampled as HIRDLS show larger GWPE than for SABER, agreeing better with the actual observations than they do for SABER. GWPE for COSMIC (**Figure 9.40c**) is fairly similar to that of HIRDLS over their short coincident period but roughly 10 % - 30 % larger than the HIRDLS results, for both observations and reanalyses. The longer COSMIC record shows a similar QBO variation of GWPE as HIRDLS and SABER: clear peaks during E-W transitions, and little systematic evidence of peaks during W-E transitions.

AIRS observations (**Figure 9.40d**, black line) differ markedly from the other instruments, showing a clear annual variation and ambiguous evidence of variation with QBO phase. The reanalyses sampled as AIRS, in contrast, agree well with each but not at all with the observations. Peaks in the GWPE occur during E-W transitions and to some extent also during W-E transitions. The observed annual variation is not reproduced. In contrast to the other cases (panels **a** - **c**) for AIRS the reanalyses tend to overestimate rather than underestimate the observed GWPE. The reasons for these large disagreements are presently unclear and are not considered further here.

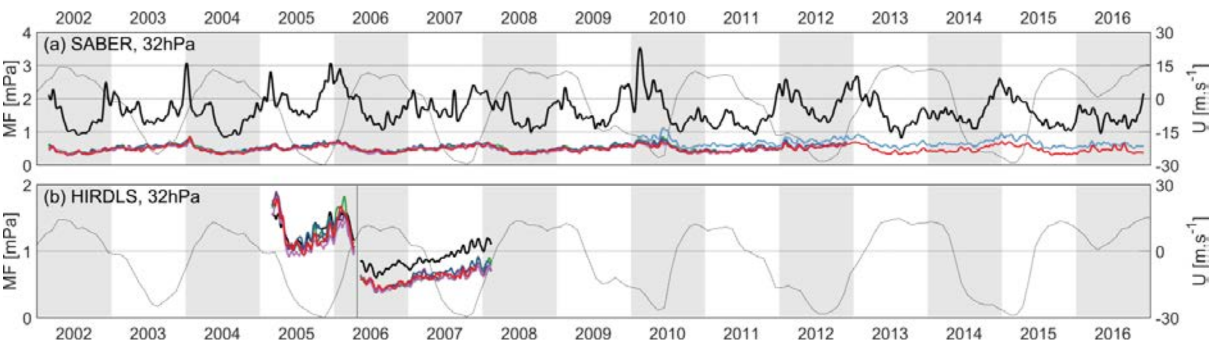


Figure 9.41: *As* ***Figure 9.40****, but for the time series of momentum flux (MF) in reanalyses and satellite observations.*

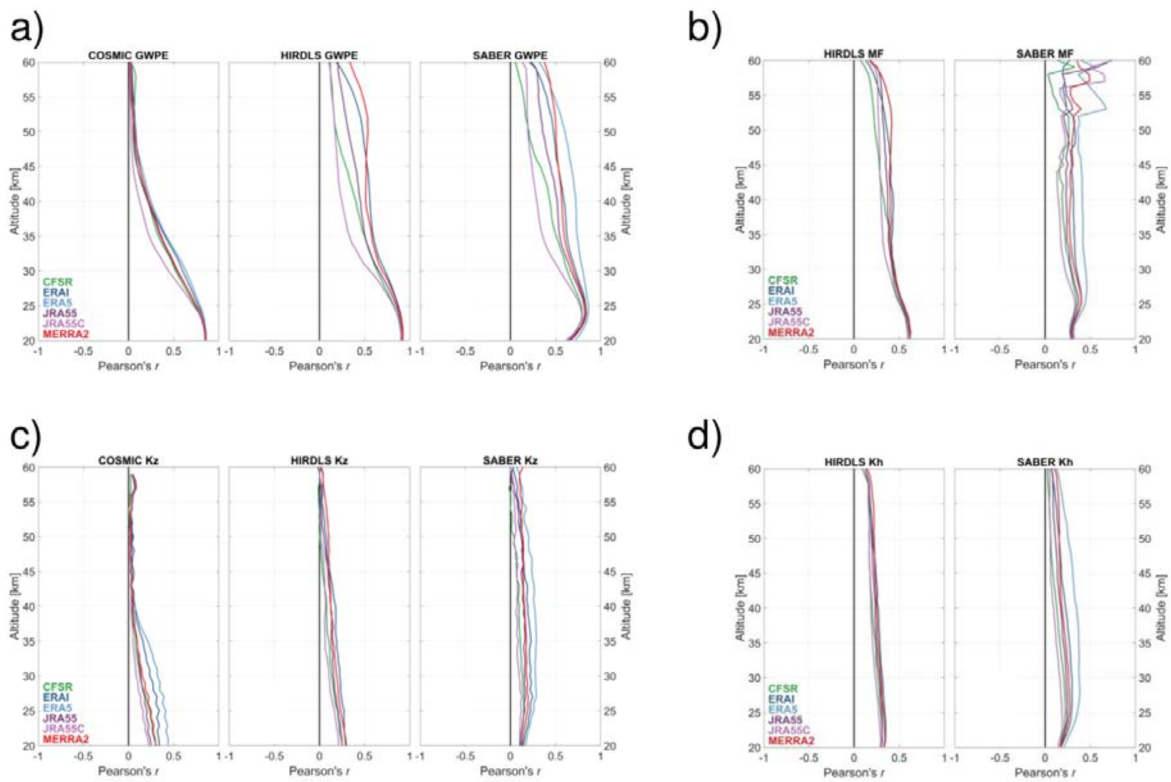


***Figure 9.42:** Vertical profiles of correlation coefficients between observations and reanalyses sampled in the same way as the observations for (a) gravity wave potential energy (GWPE), (b) momentum flux (MF), (c) vertical wavenumber (K_Z), and (d) horizontal wavenumber (K_h).*

An estimate of momentum flux associated with the measured waves can be made in the case of SABER and HIRDLS and is shown for the 32 hPa level in **Figure 9.41**. For HIRDLS, the reanalyses and observations agree well in the first portion of the record but in the second portion the reanalysis momentum fluxes are weaker than observed. This is caused by a change in the HIRDLS scan pattern in April 2006 that changed the inter-profile distance (it should not affect the GWPE because it is generated from individual profiles rather than along-track pairs). For SABER the reanalyses are much weaker than the observations, but very roughly seem to follow their time variation.

The correlation coefficient between observations and reanalyses for the different satellites and measured quantities, over the altitude range 20 - 60 km, is shown in **Figure 9.42**. For GWPE (panel a) the correlations near 30 hPa (≈ 25 km) for COSMIC, HIRDLS and SABER are fairly high, ≈ 0.8, as expected from **Figure 9.40**, confirming that the reanalyses capture the time variation of observed waves if not their magnitudes. The agreement degrades with increasing altitude, particularly for COSMIC. JRA-55C degrades more rapidly than the other reanalyses, as might be expected since it does not assimilate any satellite data. Up to about 25 km altitude all reanalyses have similar correlations (including JRA-55C) but inter-reanalysis differences are apparent at higher altitudes for HIRDLS and SABER.

For momentum fluxes, **Figure 9.42(b)** shows that correlations for HIRDLS are roughly 0.4 - 0.6 for all reanalyses over much of the altitude range, while SABER correlations are generally lower. The analysis methods also provide estimates of the dominant vertical wavenumbers for COSMIC, HIRDLS and SABER, and horizontal wavenumbers for HIRDLS and SABER (panels c and d, respectively). In all cases the correlations between reanalyses and observations are less than 0.5, suggesting that the reanalyses do not provide much useful information about these quantities. In summary, the time variation of SABER, HIRDLS and COSMIC GWPE is reproduced well by reanalyses at altitudes near 25 km (30 hPa) and below, although the magnitudes of GWPE in the reanalyses tend to be too low. It is notable that SABER and HIRDLS are not assimilated by the reanalyses and hence provide an independent validation of their GWPE. GNSS-RO data is assimilated in the reanalyses and could be one reason for the good agreement with SABER and HIRDLS in the lower tropical stratosphere. At higher altitudes (above 30 km, roughly 10 hPa) correlations tend to be lower (0.5 or less) and there are significant inter-reanalysis differences.

9.4 QBO teleconnections

There is a well-known impact of the QBO on the extra-tropical winter stratosphere dating back to classic papers by *Holton and Tan* (1980, 1982) who first noted that the NH polar vortex was stronger and less disturbed under QBO-W conditions than under QBO-E conditions, especially in early winter.

This impact is generally referred to as the "Holton-Tan" relationship, and has been studied by many subsequent authors (for reviews see *Anstey and Shepherd*, 2014; *Baldwin et al.*, 2001). Evidence is presented in *Section 9.4.1* for the continued existence of this relationship now that much longer data records are available. The consistency of the evidence between the different reanalysis datasets is also examined. Both composite analysis and a multi-linear regression (MLR) technique are used, noting that the latter aids in distinguishing the QBO signal from other sources of variability such as the ENSO, volcanic eruptions and the 11-year solar cycle. Sensitivity of the QBO signal to the data period and to the type of data assimilated by the reanalyses is also explored, at both equatorial and extra-tropical latitudes.

There is also evidence for a QBO impact on tropospheric winds and mean sea level pressure (MSLP) in the NH winter months (*Gray et al.*, 2018; *Anstey and Shepherd*, 2014; *Garfinkel and Hartmann*, 2011a,b; *Baldwin et al.*, 2001) and an influence of the QBO on the Madden Julian Oscillation (MJO) has also been recently observed (*e.g.*, *Son et al.*, 2017; *Nishimoto and Yoden*, 2017; *Yoo and Son*, 2016; *Marshall et al.*, 2016). Much of the research interest in QBO influence at the surface has been driven by its potential to extend seasonal predictability, since the QBO has relatively long period (*e.g.*, *Marshall and Scaife*, 2009). There are several potential routes for QBO influence at the surface. The polar route involves the Holton-Tan influence on vortex variability, which can then extend to the surface (*Kidston et al.*, 2015; *Baldwin and Dunkerton*, 2001). The subtropical route involves the direct modulation of the subtropical jet by the QBO-induced meridional circulation in the lower stratosphere. The tropical route is via the QBO modulation of temperatures (and hence static stability and wind shear) in the tropical lower stratosphere which can potentially influence tropical precipitation (*Gray et al.*, 2018; *Nie and Sobel*, 2015; *Garfinkel and Hartmann*, 2011a; *Liess and Geller*, 2012; *Ho et al.*, 2009; *Collimore et al.*, 2003; *Giorgetta et al.*, 1999). **Figure 1** of *Gray et al.* (2018) provides a schematic of these possible influence routes. *Section 9.4.2* examines more closely the QBO impact on tropospheric winds, *Section 9.4.3* places the QBO in the context of other major stratospheric forcings (solar, volcanic, ENSO), and *Section 9.4.4* examines the impact on the mean sea level pressure and precipitation fields, using the MLR technique.

9.4.1 Stratospheric teleconnections

Figure 9.43a,b shows the time-series of daily zonally-averaged zonal winds at 60°N, 10 hPa from the ERA-Interim and JRA-55 reanalysis datasets for 1979-2016, the post-satellite data era. QBO westerly (QBO-W) / easterly (QBO-E) composites are shown in red / blue and are defined by whether the equatorial zonal-mean zonal winds at 50 hPa in January are greater or less than zero (the results are relatively insensitive to a threshold of 3 m s^{-1} or 5 m s^{-1} instead of zero). The timeseries from the two datasets are indistinguishable, demonstrating how well the data assimilation captures the vortex behaviour at this level.

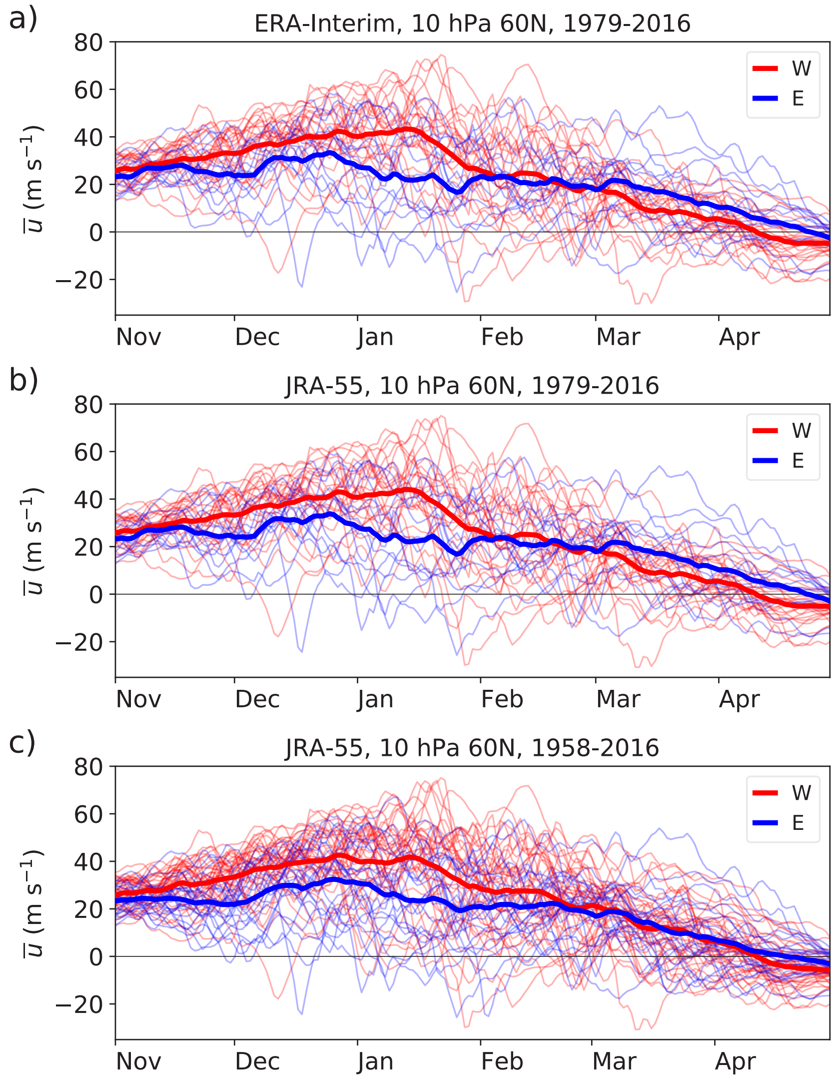


Figure 9.43: *Time series of daily zonally-averaged zonal winds (m s^{-1}) at 60°N, 10 hPa from each NH winter. (a) ERA-Interim, 1979-2016, (b) JRA-55, 1979-2016, (c) JRA-55, 1958-2016. Red and blue indicate years in which the equatorial QBO was westerly (W) and easterly (E), respectively, as determined by the sign of the equatorial zonal winds at 50 hPa in January. Thick red / blue lines show the ensemble-average.*

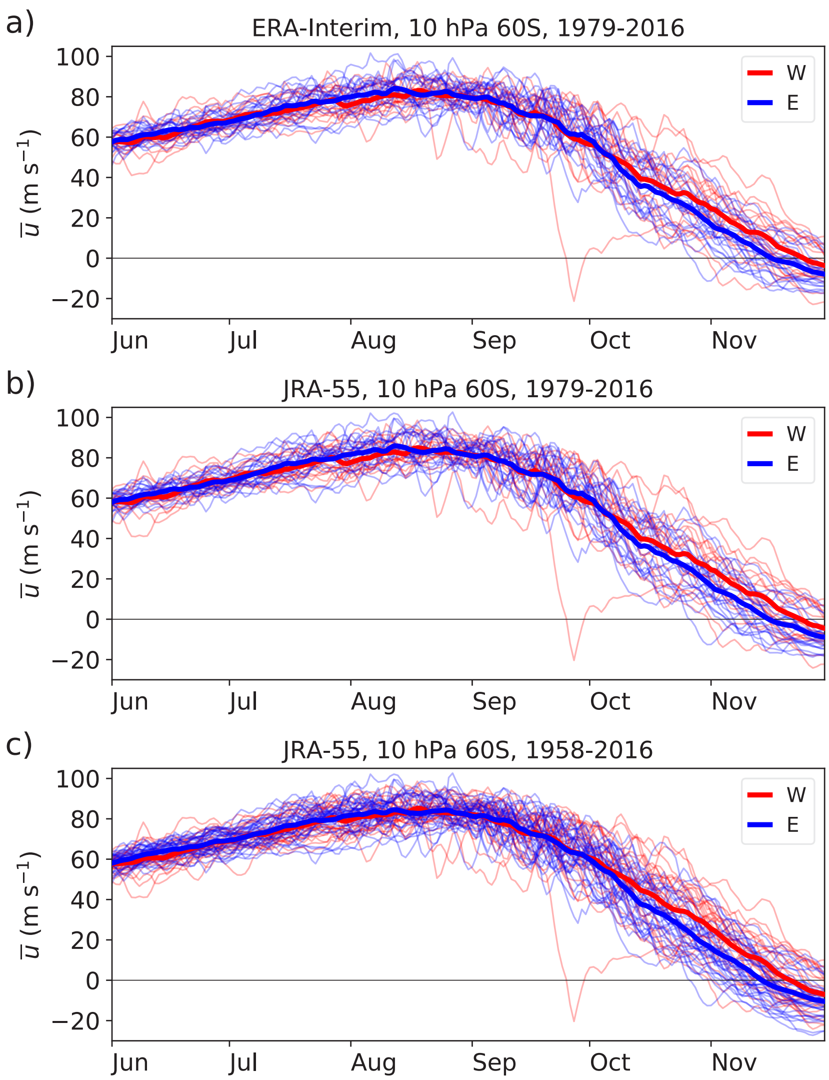


Figure 9.44: *As **Figure 9.43** but for 10 hPa 60 °S and QBO phase determined by the sign of the equatorial zonal winds at 20 hPa in July.*

Corresponding plots from the other reanalyses are virtually identical (not shown), and this is also true of the JRA-55C dataset that assimilates only conventional data (*i.e.*, no satellite data are assimilated). The good agreement of these reanalysis datasets indicates that the assimilated radiosonde data are likely the dominant influence on the polar vortex region at this level.

In early winter there is a clear separation of the composite-means of the two QBO phases, particularly in January, showing a stronger, less disturbed (and hence colder) NH vortex in QBO-W years than in QBO-E years. This confirms the signal first identified by Holton and Tan. In late winter (March, April) this behaviour reverses, and the QBO-W vortex is weaker and more disturbed, although the difference between the composite-means is much smaller than in January. The QBO-W composite-mean also reaches the zero level earlier in April, suggesting a possible QBO modulation of the final warming date.

JRA-55 is the reanalysis with the longest period of available data for which a consistent assimilation system has been employed. **Figure 9.43c** shows the full period 1958 - 2016. As for the shorter period, the QBO-W polar vortex is stronger than the QBO-E vortex in early winter, and this effect appears to extend for slightly longer, into February (for an estimation of the statistical significance of this February feature see next section). However, the reversal of the QBO impact in late winter is no longer evident, suggesting that this feature is sensitive to the length of the data period. The apparent QBO influence on the final warming date in early spring is nevertheless still evident. Both these late winter / early spring characteristics require more years of data in order to test their robustness.

Figure 9.44 shows the corresponding time-series plots for the SH polar vortex (60 °S, 10 hPa). There is no evidence of a QBO influence on the strength of the early winter vortex. However, there is an apparent QBO influence in late winter, and consequently a QBO impact on the final warming date. While this is still evident in the longer 1958 - 2016 period, verification of a robust signal requires additional years. Some analyses (*e.g.*, the MLR analyses shown in **Figure 9.47**) suggest there may be more sensitivity at higher levels *e.g.*, 1 - 3 hPa at the core of the vortex.

Corresponding plots for 1 hPa for the 1979 - 2016 and 1958 - 2016 periods are shown in **Figure AS9.17**. At this higher level, the 1979 - 2016 period has a suggestion of a QBO response earlier in winter (June - July) but this is not present in the longer period. However, there is an obvious westerly bias in the earlier data at these high levels (see *e.g.*, the difference in June values between **Figures AS9.17a** and **AS9.17b**) so using data from the pre-satellite era is not recommended for analysis of the SH QBO response in the upper stratosphere.

Figure 9.45 shows the corresponding latitude-time evolution of the NH 10 hPa composite QBO response over a nine-month period for the same reanalyses and time periods as in **Figure 9.43** (1979 - 2016 ERA-Interim and JRA-55, and 1958 - 2016 JRA-55). In all cases the response is first apparent at low latitudes before rapidly extending poleward in October and November. From November the high-latitude response strengthens, peaks in January, and subsequently decays. This confirms that the latitude of 60 °N used in **Figures 9.43** and **9.44** is representative of the composite response at latitudes poleward of 40 °N.

The ERA-Interim and JRA-55 1979 - 2016 evolutions are extremely similar at all latitudes, as might be expected from **Figure 9.8** that showed appreciable inter-reanalysis disagreement of 10 hPa monthly-mean zonal-mean zonal wind only in the tropics (15 °S - 15 °N). The composite for the extended JRA-55 record (1958 - 2016, **Figure 9.45**, bottom) differs in its late-winter response, consistent with the corresponding 60 °N figures, but otherwise shows similar features as the 1979 - 2016 period composite.

The corresponding 10 hPa composites for the SH are shown in **Figure 9.46**. As in the NH a response first appears at low latitudes in early winter, but unlike the NH it is mainly confined equatorward of 60 °S until late winter when, beginning in September, positive wind anomalies migrate poleward, culminating in the peak high-latitude response during November. These features are very similar in both ERA-Interim and JRA-55 for the 1979 - 2016 period, indicating that during the satellite era the available observations are sufficient to strongly constrain the two reanalyses. The validity of pre-satellite reanalysis products in the SH is more questionable given the much sparser radiosonde coverage of the SH compared to the NH. Nevertheless, at least for the composite-mean response to the QBO, JRA-55 for 1958 - 2016 (**Figure 9.46**, bottom) shows very similar behaviour as the 1979 - 2016 period.

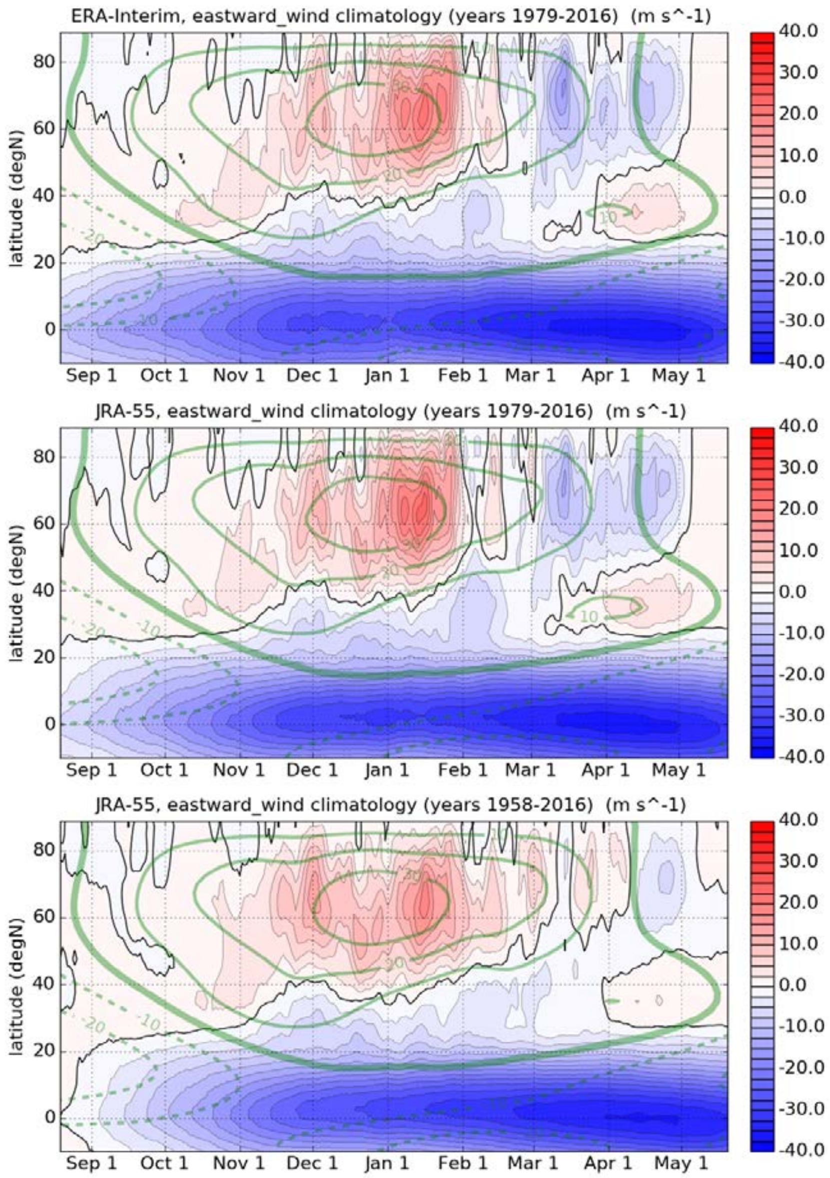


Figure 9.45: *Time-series of QBO composite difference (QBO-W minus QBO-E) of daily zonal-mean zonal wind at 10 hPa, 10 ° S - 90 ° N during NH winter (contour interval 2.5 m s*$^{-1}$*). Top: ERA-Interim, 1979 - 2016. Middle: JRA-55, 1979 - 2016. Bottom: JRA-55, 1958 - 2016. QBO phase is defined by the sign of 50 hPa January monthly-mean equatorial wind. Green contours show climatological 10 hPa zonal-mean zonal wind (contour intervals 10 m s*$^{-1}$*, westerly solid, easterly dashed, zero-wind line thick solid).*

This initial assessment of the impact of the QBO on the polar vortex suggests that the results are more sensitive to the data period employed than to the choice of reanalysis dataset. This conclusion corresponds well with the results from the S-RIP Chapter 6 (extratropical stratosphere-troposphere coupling) where many more detailed aspects of the NH polar vortex variability are diagnosed. Conclusions drawn from a simple composite analysis may be compromised by aliasing problems due to the presence of variability from other sources, especially when the data period is short. In order to address this source of uncertainty, we now employ a MLR analysis that includes indices to represent variability associated with the 11-yr solar cycle, ENSO, volcanic eruptions and a linear trend, as well as the QBO. The MLR analysis was performed at each latitude / pressure level using the monthly-mean zonal-mean zonal winds. The primary results are shown for the 4 most recent reanalysis datasets (JRA-55, ERA-Interim, MERRA-2, CFSR) but results from the older reanalyses are shown, where appropriate, in *Appendix A9.1*. The regression analysis covered the period 1980 - 2016 for the 4 most recent reanalyses and 1980 - 2012 for the older reanalyses (1980 was chosen as the common starting year, to accommodate MERRA; the ERA-40 analysis extends only to 2002). The QBO index is defined by the (contemporaneous) FUB equatorial wind time-series at 50 hPa. This level was chosen as the commonly available level that optimises the NH QBO response at higher latitudes (*Baldwin and Dunkerton*, 1998). Note that, in contrast, the 20 hPa equatorial winds are employed to optimise the SH winter response (next sub-section) and the 70 hPa equatorial winds are employed to optimise the tropospheric response in *Section 9.4.2*, following *Gray et al.* (2018).

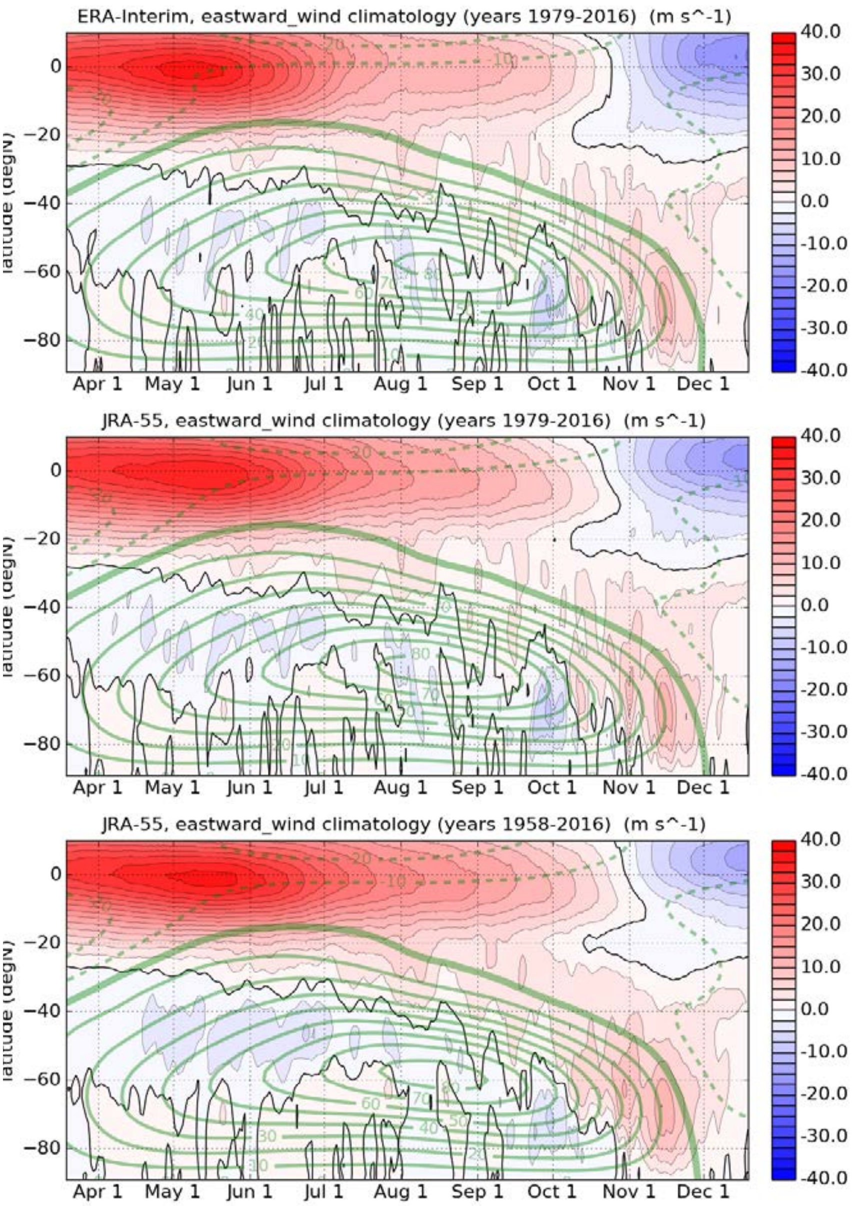


Figure 9.46: *As **Figure 9.45** but for 90°S - 10°N during SH winter, with QBO phase defined by the sign of 20 hPa July monthly-mean equatorial wind.*

Figure 9.47 shows the 1980 - 2016 QBO response (QBO-W minus QBO-E difference) for the NH winter hemisphere (October - March) from the four individual reanalyses datasets JRA-55, ERA-Interim, MERRA-2 and CFSR. The corresponding results for the older reanalysis datasets (MERRA, ERA-40, JRA-25, NCEP-DOE R2 and NCEP-NCAR R1) are shown in **Figure AS9.18**. **Figure 9.47** confirms many of the initial impressions gained from the analysis of daily data at only one polar vortex location. Firstly, the QBO signal in all four datasets is almost identical, especially in the extra-tropics and up to 10 hPa, confirming the earlier conclusion that the choice of dataset is immaterial for the purposes of examining the Holton-Tan relationship in the NH over this data period.

The 4-reanalysis mean of the QBO signal is shown in **Figure 9.48**, top row; the regression analysis was performed on the individual reanalysis datasets and then the regression coefficients, statistical significance values and climatological fields were averaged to produce this final figure. The figure clearly shows the Holton-Tan relationship, and is consistent with the daily data composite analysis in **Figure 9.43**. The more careful extraction of the QBO signal from the ENSO, volcanic and solar influences using the MLR approach shows very clearly that the polar vortex response in both December and January is statistically significant at the 99 % level. The reversed QBO response in later winter is also evident, for example in March at the 95 % significance level.

There are, however, some inter-reanalysis differences in the QBO signal, especially in the upper equatorial stratosphere (see **Figure 9.47**). This likely reflects uncertainty due to (a) the relatively poor vertical resolution of the assimilated satellite datasets, (b) differences in the satellite datasets that are assimilated and (c) how well the assimilation model itself is able to represent the processes that give rise to the QBO (and SAO) at these levels. These differences in the upper equatorial stratosphere are further highlighted in **Figure 9.48**: while the top row shows the average of the QBO signal from the 4 reanalyses, the 2nd row shows their standard deviation (SD). The SD is small nearly everywhere apart from the equatorial stratosphere above 30 hPa. In the same Figure (3rd row) we also show the 4-dataset average of the interannual variability in each month, to help assess how these inter-reanalysis differences in the QBO signal compare with the background year-to-year variations. As expected, there is large interannual variability in the region of the QBO at equatorial latitudes above 70 hPa and at polar latitudes associated with the variability of the polar jets. In the lowermost row of **Figure 9.48** we show the inter-reanalysis SD of the QBO signal as a percentage of the interannual variability (*i.e.*, 2nd row divided by 3rd row, times 100). This highlights that the inter-reanalysis differences are of the same order of magnitude as the interannual variability in the upper equatorial stratosphere. It also highlights that the inter-reanalysis differences extend down into the equatorial troposphere and there are also differences in the QBO responses in the SH (*e.g.*, in October), perhaps not surprisingly, given the sparsity of the data available for assimilation.

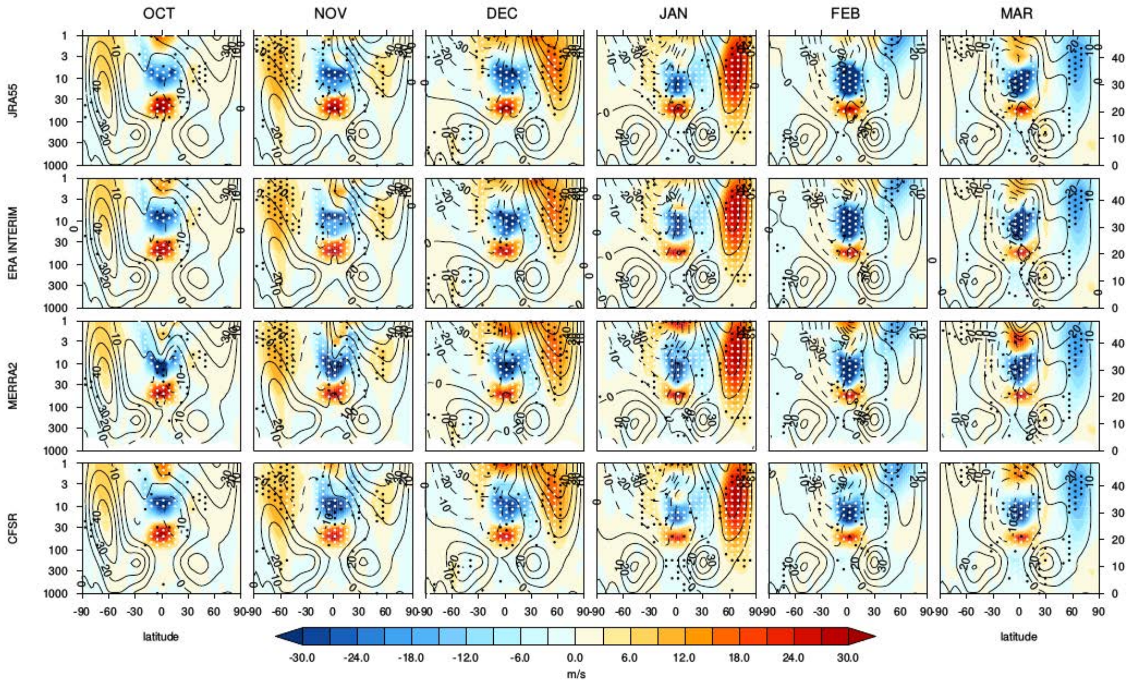


Figure 9.47: *Height-latitude cross-sections of the NH winter (October - March) QBO response in zonally-averaged zonal winds (m s^{-1}) from the regression analysis of the four recent reanalyses: JRA-55, ERA-Interim, MERRA-2 and CFSR for the period 1980 - 2016. The QBO index was based on the FUB equatorial zonal winds at 50 hPa. The regression coefficients have been scaled to show the typical QBO-W minus QBO-E difference in zonal winds (to aid comparison with studies that employ composite difference techniques). Black (white) dots denote statistical significance at the 95 % (99 %) level. The appropriate monthly climatological wind fields are superimposed with contour spacing of 10 m s^{-1}.*

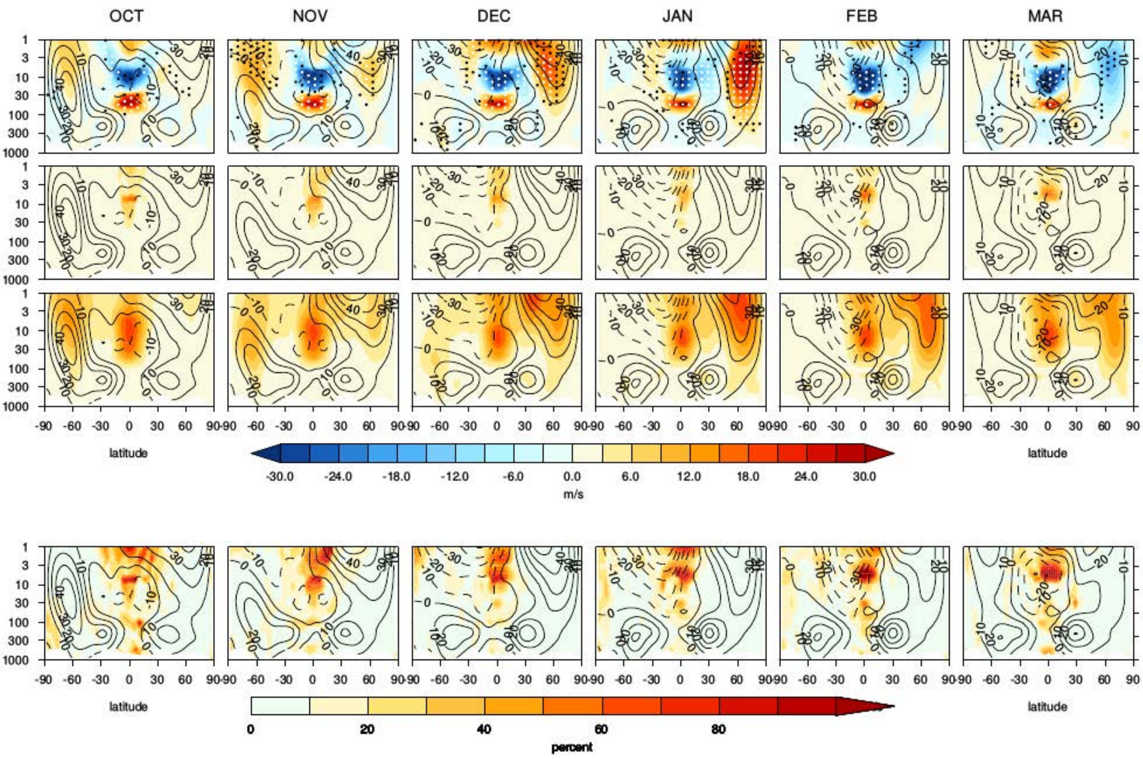


Figure 9.48: *1st row: Average NH winter QBO signal in zonally averaged zonal winds (m s^{-1}) from the four recent reanalyses (JRA-55, ERA-Interim, MERRA-2, CFSR) i.e. the average of the fields shown in **Figure 9.47**, for the period 1980 - 2016 (with averaged climatology and statistical significance levels overlaid). 2nd row: inter-reanalysis standard deviation (m s^{-1}) of the QBO signals from the 4 reanalyses. 3rd row: average of the interannual standard deviation (m s^{-1}) from the 4 reanalyses. Bottom row: inter-reanalysis standard deviation in the QBO signal as a percentage of the interannual variability (i.e., row 2 divided by row 3, multiplied by 100).*

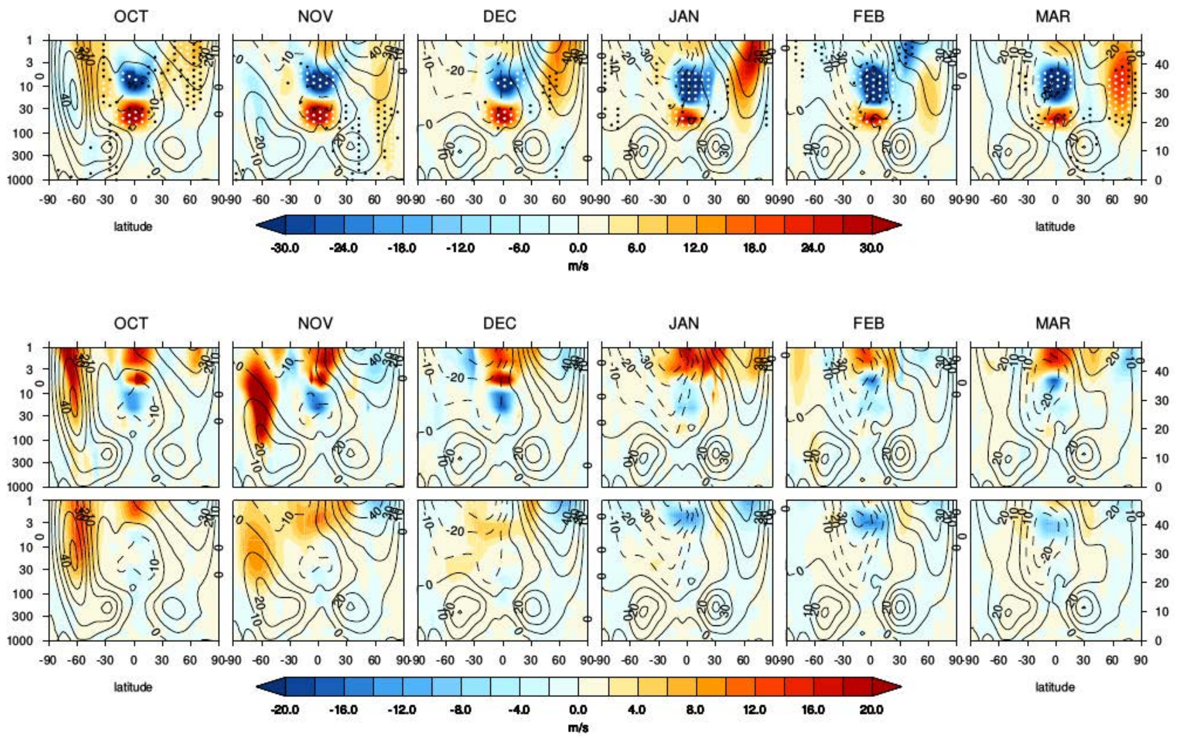


Figure 9.49: *Comparison of NH winter (October - March) QBO signal in zonally-averaged zonal winds from ERA-40 versus JRA-55 for the period 1958 - 1979. Top row: the averaged QBO signal (m s^{-1}) from the 2 reanalysis datasets. Middle row: difference (m s^{-1}) in the 2 QBO signals (ERA-40 minus JRA-55). Bottom row: difference (m s^{-1}) between the 2 climatological fields (ERA-40 minus JRA-55). The average of the 2 climatological wind fields is overlaid on each plot (contour interval of 10 m s^{-1}).*

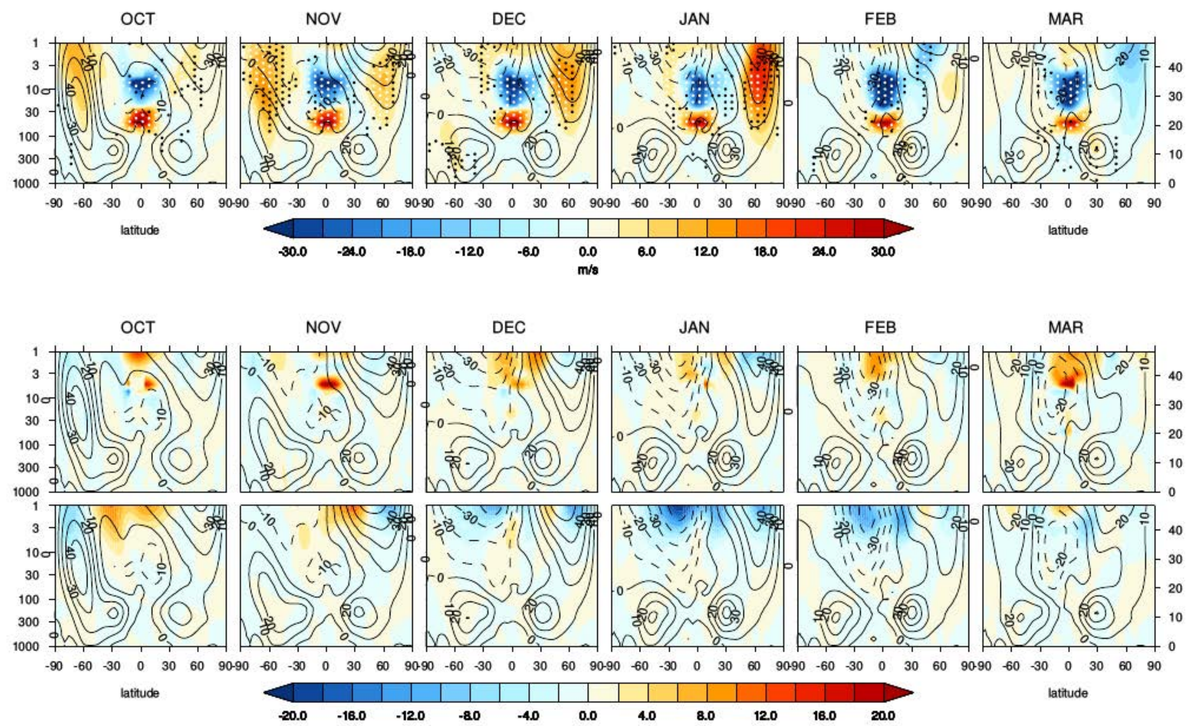


Figure 9.50: *As Figure 9.49 but comparing the NH winter difference between the QBO signals from JRA-55 versus JRA-55C (JRA-55 minus JRA-55C) for the period 1973 - 2012.*

To examine differences in the QBO signal in the pre-satellite and post-satellite era, **Figure 9.49** shows the regression-based QBO response from ERA-40 and JRA-55 for the period 1958 - 1979. The top panel shows the average response (the separate responses are shown in **Figure AS9.19**) together with the ERA-40 minus JRA-55 difference in the QBO signal (middle row) and the difference in their monthly climatologies (bottom row). A similar comparison between JRA- 55 and NCEP-NCAR *R1* is provided in **Figures AS9.20** and **AS9.21**. Not surprisingly, the main differences in all three fields are found in the upper stratosphere, particularly in equatorial regions, and in the SH polar regions where there are fewer constraining data available. Interestingly, the differences in the climatological fields are not of the same sign in all months. Comparison with **Figure 9.48** shows the difference in the QBO signals analysed using pre-satellite and post-satellite era data. The early winter NH polar response is similar, although less significant in the early period, but the late winter responses are very different. This highlights that the late-winter NH QBO response is sensitive to the selected data period.

To further explore the influence of satellite data assimilation we compare results from the MLR analysis of JRA-55 and JRA-55C. **Figure 9.50** shows the mean of the QBO signals from the two reanalyses (top row), the difference between the QBO signals from the 2 datasets (middle row) and the difference between the background climatology of the 2 datasets (bottom row). The comparison is carried out for the period 1973 - 2012 which is the maximum period of overlap of these two reanalysis datasets (see **Figure AS9.22** for results from the individual regression analyses). As expected, the main differences between the two QBO signals are in the upper stratosphere, where the satellite data are most important. The largest differences are in the equatorial region. This is likely because the vertical depth of the equatorial QBO signal is relatively shallow, and involves large vertical wind shears that the satellite data assimilation is poor at capturing. The polar vortex structure, in comparison, is more barotropic and is relatively well characterised by the assimilation of radiosonde data; indeed the small differences between the JRA-55 and JRA-55C QBO signals at higher latitudes leads to the conclusion that the QBO response at NH high latitudes seen in the reanalyses does not rely on the assimilation of satellite data.

The corresponding comparison of the MERRA and MERRA-2 reanalysis datasets is shown in **Figures AS9.23** and **AS9.24**. We note that several of the improvements in MERRA-2 have a potential for influencing the representation of the QBO, including (a) the assimilation of MLS satellite data above 5 hPa which is likely to improve the vertical shears in this region because of its limb-sounding nature, and (b) the ability of the underlying model to self-generate its own QBO. The main differences between the two reanalyses are again in the equatorial region, and extend down as far ≈ 50 hPa. In the austral winter period the differences consist of a relatively straightforward westerly bias in the MERRA dataset in the upper equatorial stratosphere and a shift in the height distribution of the QBO, but in the boreal winter the height pattern of the differences are more complicated and suggest the presence of a number of different influences.

The sensitivity of the NH QBO polar (Holton-Tan) response to the length of the data period is underlined in **Figure 9.51** which shows a comparison of the QBO response from the JRA-55 reanalysis, which is the longest available dataset that uses a consistent underlying model, for the whole period 1958 - 2016 compared with the shorter post-satellite period 1980 - 2016 shown in **Figure 9.47** (the difference fields are provided in **Figure AS9.25**). The QBO signal from the longer period is essentially the same as in the shorter period in mid-winter (December–January) with slightly reduced amplitude, especially in January. However, the late-winter response with a weakened NH polar vortex in February–March is much weaker in the extended period and is no longer statistically significant. Given that the vortex response is represented well by the assimilation of only conventional observations (albeit these were less extensive in the pre-satellite era) the disappearance of the late-winter signal in the longer period is unlikely to be due to the lack of satellite data in the earlier period.

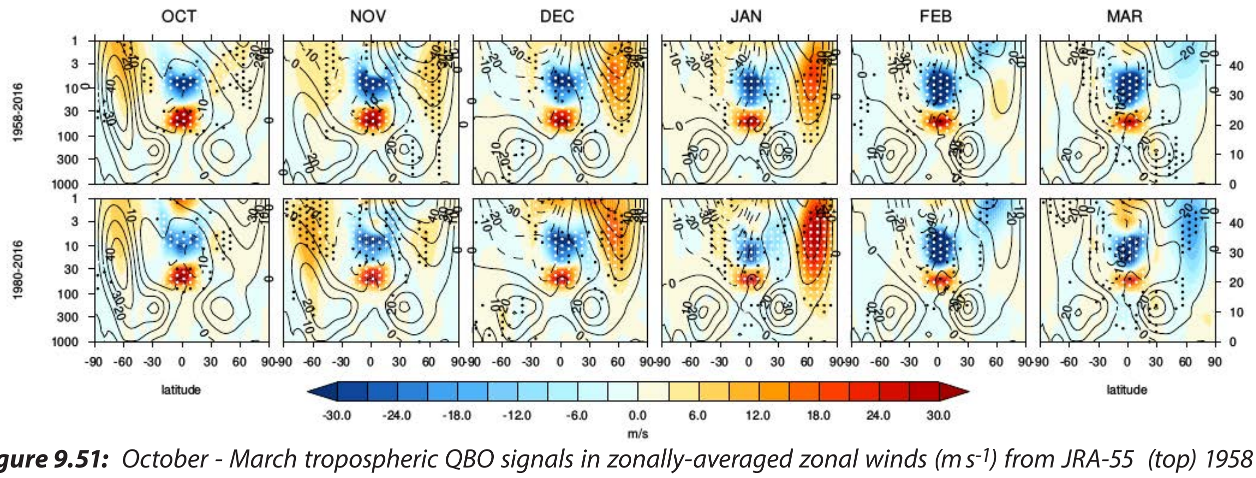


Figure 9.51: *October - March tropospheric QBO signals in zonally-averaged zonal winds (m s^{-1}) from JRA-55 (top) 1958 – 2016, (bottom) 1980 - 2016.*

It is most likely a reflection of the true nature of the QBO relationship in these months i.e. that it is not statistically significant and additional years will be required to determine whether the signal is real or not. We therefore recommend the use of the longest data period available for studies of the NH winter QBO response.

On the other hand, the NH tropospheric response in March in **Figure 9.51**, with a dipole structure between 30 - 60 °N showing the jet strengthened in the subtropics and weakened in midlatitudes, remains a persistent feature in the longer period analysis. This is despite the lack of a significant vortex response, suggesting that it is unlikely to be directly associated with the vortex response. In November, a similar tropospheric NH dipole response is also more apparent in the longer period. These and other tropospheric QBO signals are discussed further in *Section 9.4.2*.

The 4-reanalysis average (JRA-55, ERA-Interim, MERRA-2, CFSR) of the 1980 - 2016 QBO response from the regression analysis for the SH winter months is shown in **Figure 9.52** (top row), together with its SD (2nd row), interannual variability (3rd row) and the SD expressed as a percentage of the interannual variability (bottom row). The QBO index in the regression analysis was defined by the (contemporaneous) FUB equatorial wind time-series at 20 hPa, to optimise the SH polar response. (Separate results for each individual reanalysis dataset, and for older datasets, are provided in **Figures AS9.26** and **AS9.27** respectively.) **Figure 9.52** can be compared with the NH responses shown in **Figure 9.48** (but note that the signals in the overlapping months are slightly different because of the difference in QBO indices employed). As in the NH, the polar vortex is stronger, less disturbed, and hence colder, under QBO-W conditions, for example in October - November at the 95 - 99 % statistical significance level. The main inter-reanalysis differences are in the upper equatorial stratosphere, with very little variations in the polar vortex response.

An examination of the SH QBO signal from the pre-satellite years is shown in **Figure 9.53**, which shows the comparison between ERA-40 and JRA-55 for 1958 - 1979. As well as large differences in the upper stratosphere, there are also large differences at SH high latitudes in the climatologies (bottom row), which are reflected to some extent in the QBO signals also. In general though, the QBO responses in the two datasets have a similar pattern, and the JRA-55 signals are generally larger and more significant (see **Figure AS9.28**). Comparison of NCEP-NCAR *R1* and JRA-55 (see **Figures AS9.29** and **AS9.30**) also shows similar features.

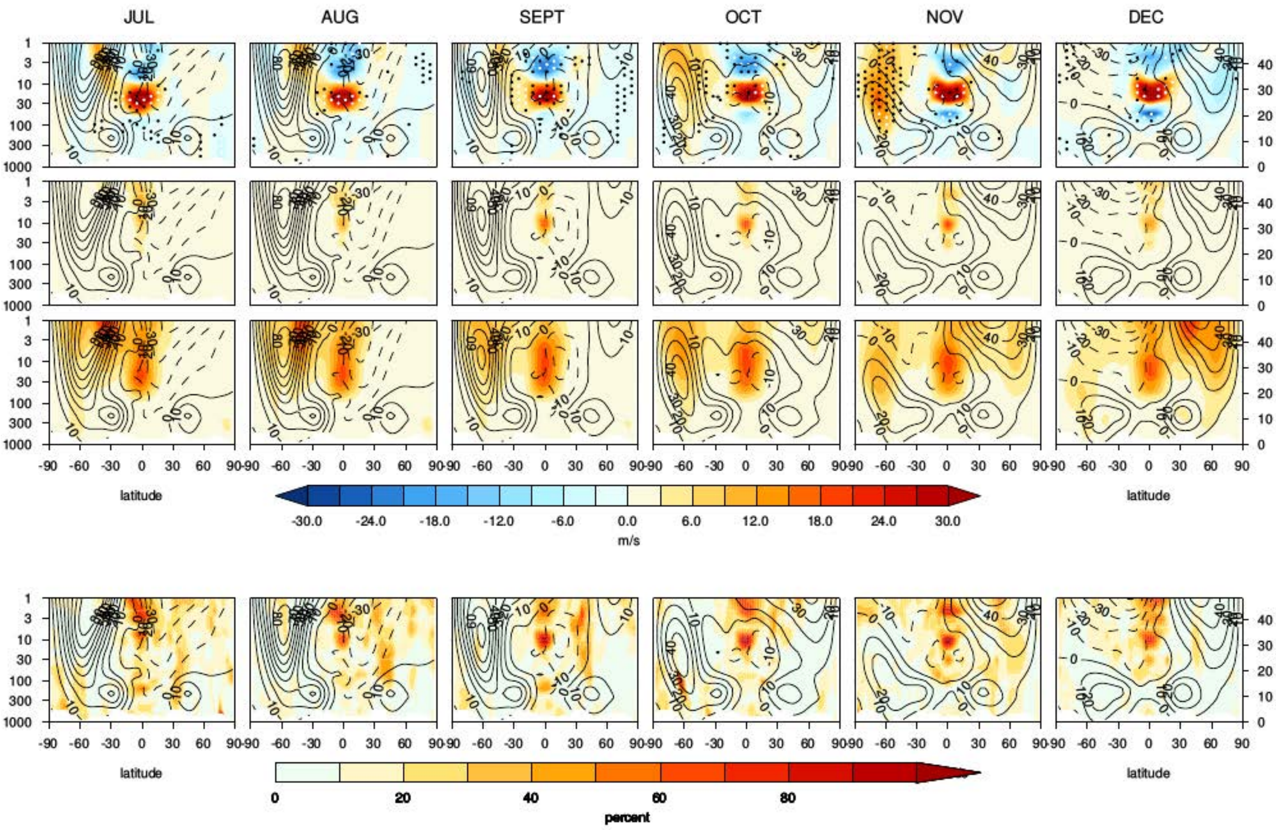


Figure 9.52: *1st row: Average SH winter QBO signal in zonally averaged zonal winds (m s⁻¹) from the four recent reanalyses (JRA-55, ERA-Interim, MERRA-2, CFSR) for the period 1980 - 2016 (with averaged climatology and statistical significance levels overlaid). 2nd row: standard deviation (m s⁻¹) of the QBO signals from the 4 reanalyses. 3rd row: average of the interannual standard deviation (m s⁻¹) from the 4 reanalyses. Bottom row: standard deviation in the QBO signal as a percentage of the interannual variability (i.e., row 2 divided by row 3 times 100).*

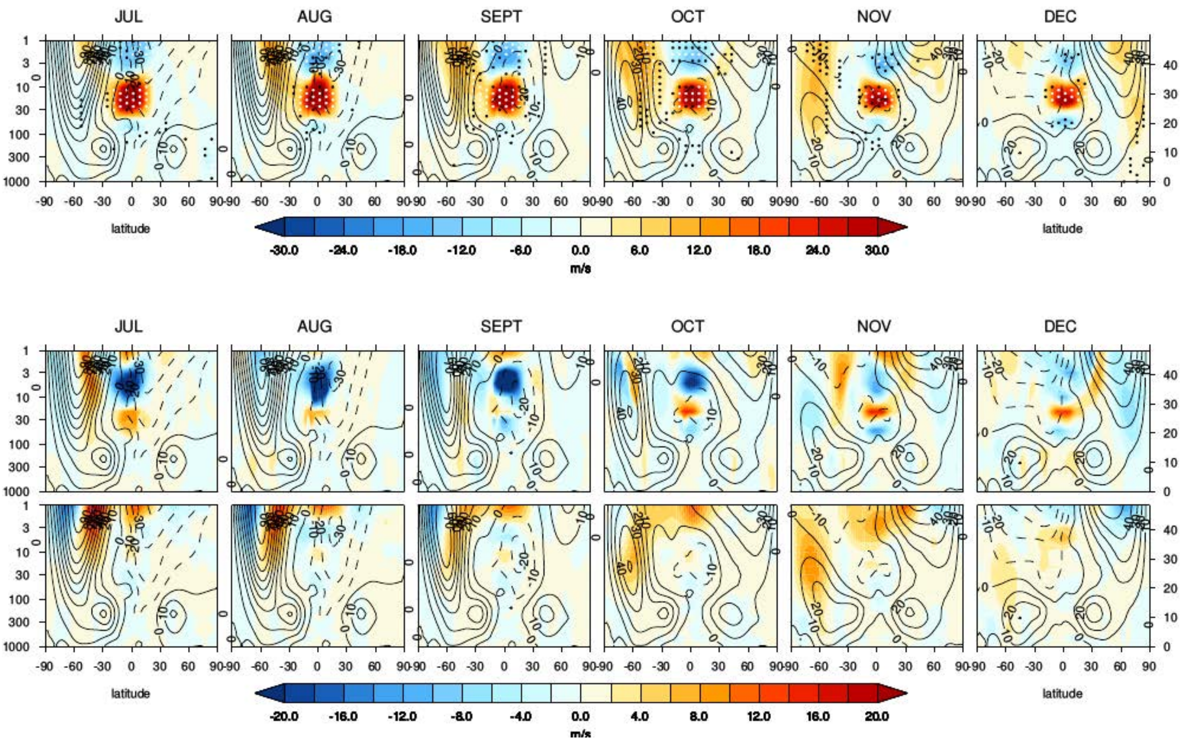


Figure 9.53: *Comparison of SH winter (July–December) QBO signal in zonally-averaged zonal winds from the ERA-40 versus JRA-55 datasets for the period 1958 - 1979. Top row: the average QBO signal (m s*$^{-1}$*) from the two reanalysis datasets. Middle row: difference (m s*$^{-1}$*) in the two QBO signals (ERA-40 minus JRA-55). Bottom row: difference (m s*$^{-1}$*) between the two climatological fields (ERA-40 minus JRA-55). The average of the two climatological wind fields is overlaid on each plot (contour interval of 10 m s*$^{-1}$*). The QBO index in the regression analysis was based on the FUB equatorial zonal winds at 20 hPa (and not 50 hPa as was the case for the NH analysis).*

Comparison of the JRA-55 and JRA-55C (the latter assimilates only conventional observations, *i.e.*, there is no assimilation of satellite data) for the SH winter period (see **Figures AS9.31** and **AS9.32**) show similar differences to those discussed for the NH winter, *i.e.,* the main impacts on the background climatology are in the upper equatorial stratosphere and the SH but this has fairly minimal impact on the extracted QBO signal. Similar conclusions are also drawn from the MERRA vs MERRA-2 comparison (**Figures AS9.33** and **A9.34**). The main differences are at equatorial latitudes, likely due to a combination of the improved satellite data assimilation in MERRA-2 together with improvements in the underlying model that enable it to self-generate a QBO.

Figure 9.54 shows the 1980 - 2016 versus the full 1958 - 2016 from JRA-55. While the general pattern of response is similar between the two data periods the statistical significance of the QBO impact on the SH vortex is substantially reduced *e.g.,* in October - November at 50 - 60 °S above 30 hPa. Given the lack of available observations above 10 hPa in the pre-satellite era it is unclear whether these differences arise from this lack of input data or whether the signal from the shorter period is simply an artefact of the analysis. Further years of data will be required to clarify this.

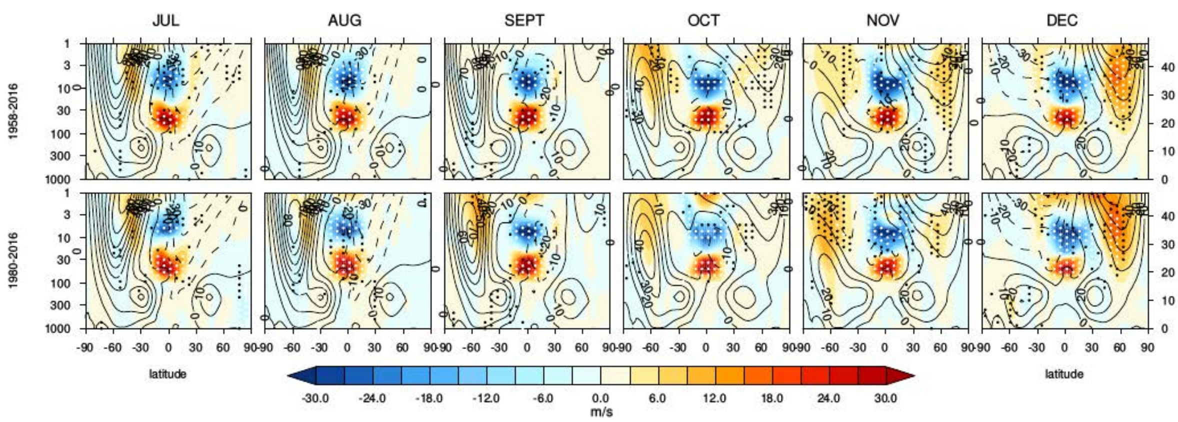


Figure 9.54: *Comparison of the SH winter QBO signals in zonally-averaged zonal winds (m s*$^{-1}$*) from JRA-55 over the extended period 1958 - 2016 versus 1980 - 2016.*

9.4.2 Tropospheric teleconnections

The average QBO signals from the four recent reanalyses (JRA-55, ERA-Interim, MERRA-2, CFSR) for the period 1980 - 2016 are shown in **Figures 9.55** and **9.56** (separate reanalyses are shown in **Figures AS9.35** and **AS9.36**). The plots are essentially those in **Figure 9.48**, with the vertical scale and contour levels adjusted to focus on the troposphere, except that the QBO index in the regression analysis is defined by the FUB equatorial wind time-series at 70 hPa to optimise the tropospheric responses, following *Gray et al.* (2018).

As in **Figure 9.48**, the top rows of **Figures 9.55** and **9.56** show the 4-reanalysis average QBO response. The standard deviation (SD) of the QBO signal (2nd row) and the interannual SD (3rd row) are shown, and also the QBO SD as a percentage of the interannual SD (bottom row). The analysis indicates several interesting tropospheric responses to the QBO. Throughout boreal winter (December–April) there is an easterly wind anomaly of up to ≈ 4 - 5 m s^{-1} in the tropical upper troposphere underlying the QBO-W phase in the lower stratosphere, and the statistical significance of this anomaly reaches 99 % in several of the months. This is accompanied by a strengthening of the subtropical jet in the winter hemisphere *e.g.,* near 30 ° N in February - March and 30 ° S in August–September. At NH polar latitudes there is a hint of a positive response underlying the positive polar vortex anomaly e.g. in December–January at 50 - 60 ° N and this is later replaced by a negative anomaly in March which may be associated with the polar vortex anomaly (although note the earlier discussion on the lack of robustness of this late-winter stratospheric response of the polar vortex).

The inter-reanalysis SD over the 1980 - 2016 post-satellite era is relatively small (see also **Figures AS9.35** and **AS9.36**). In order to examine the QBO signal with as many years as possible, **Figures 9.57** and **9.58** show the tropospheric QBO response from the JRA-55 reanalysis for the period 1958 - 2016 (see **Figures AS9.37** and **AS9.38** for differences in QBO signals and climatologies). While the main pattern of response is essentially the same, the amplitude and significance values of the signals are sensitive to the length of the data period.

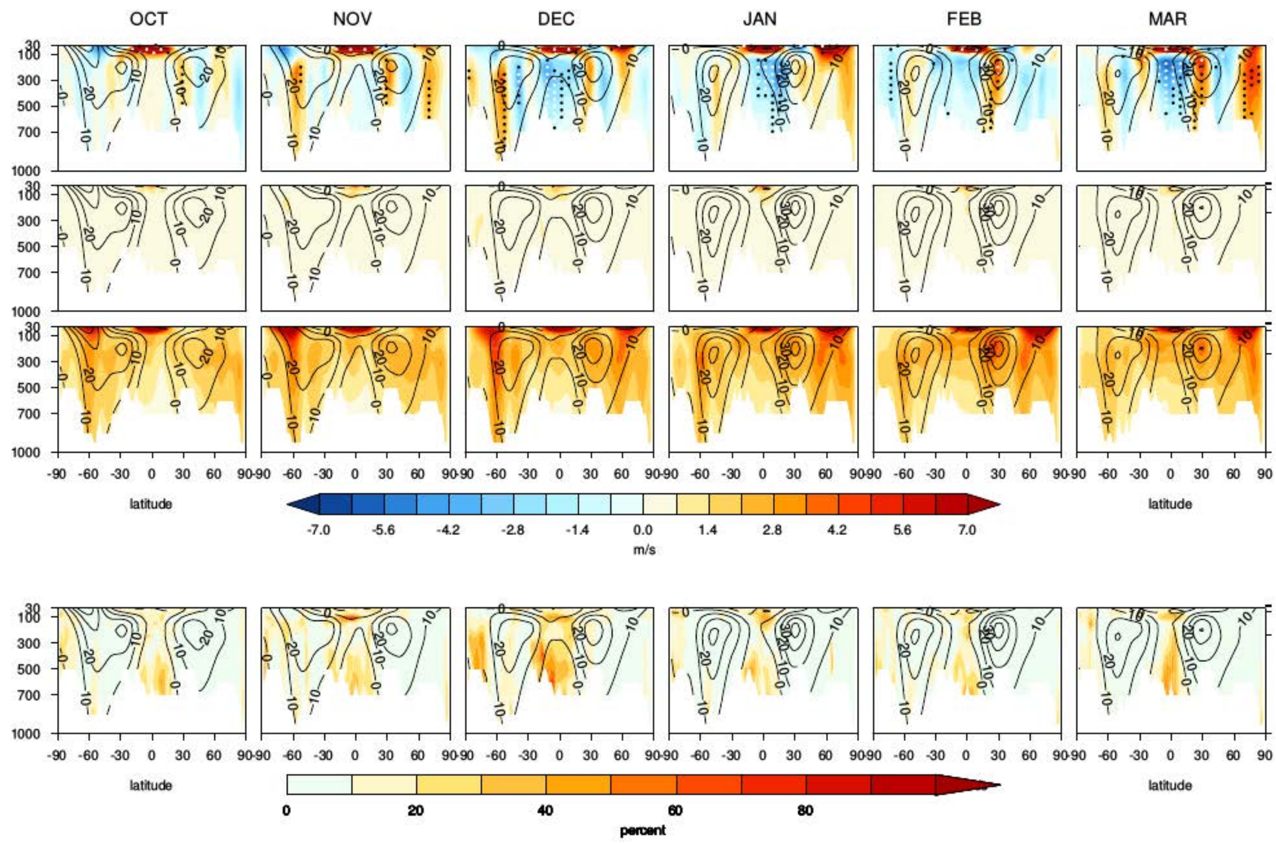


***Figure 9.55:** As **Figure 9.48** but highlighting the tropospheric response. 1st row: the average QBO signal in the troposphere for the months October - March over the period 1980 - 2016 (together with the averaged climatologies and statistical significance levels) from the regression analysis of the four recent reanalyses shown in **Figure 9.47** (JRA-55, ERA-Interim, MERRA-2, CFSR). 2nd row: standard deviation (m s^{-1}) of the QBO signals from the 4 reanalyses. 3rd row: average of the interannual standard deviation (m s^{-1}) from the 4 reanalyses. Bottom row: standard deviation in the QBO signal as a percentage of the interannual variability (i.e., row 2 divided by row 3, multiplied by 100). The QBO index in the regression analysis was based on the FUB equatorial zonal winds at 70 hPa in order to maximise the tropospheric response (and not 50 hPa or 20 hPa as was the case for the NH / SH winter analysis shown previously).*

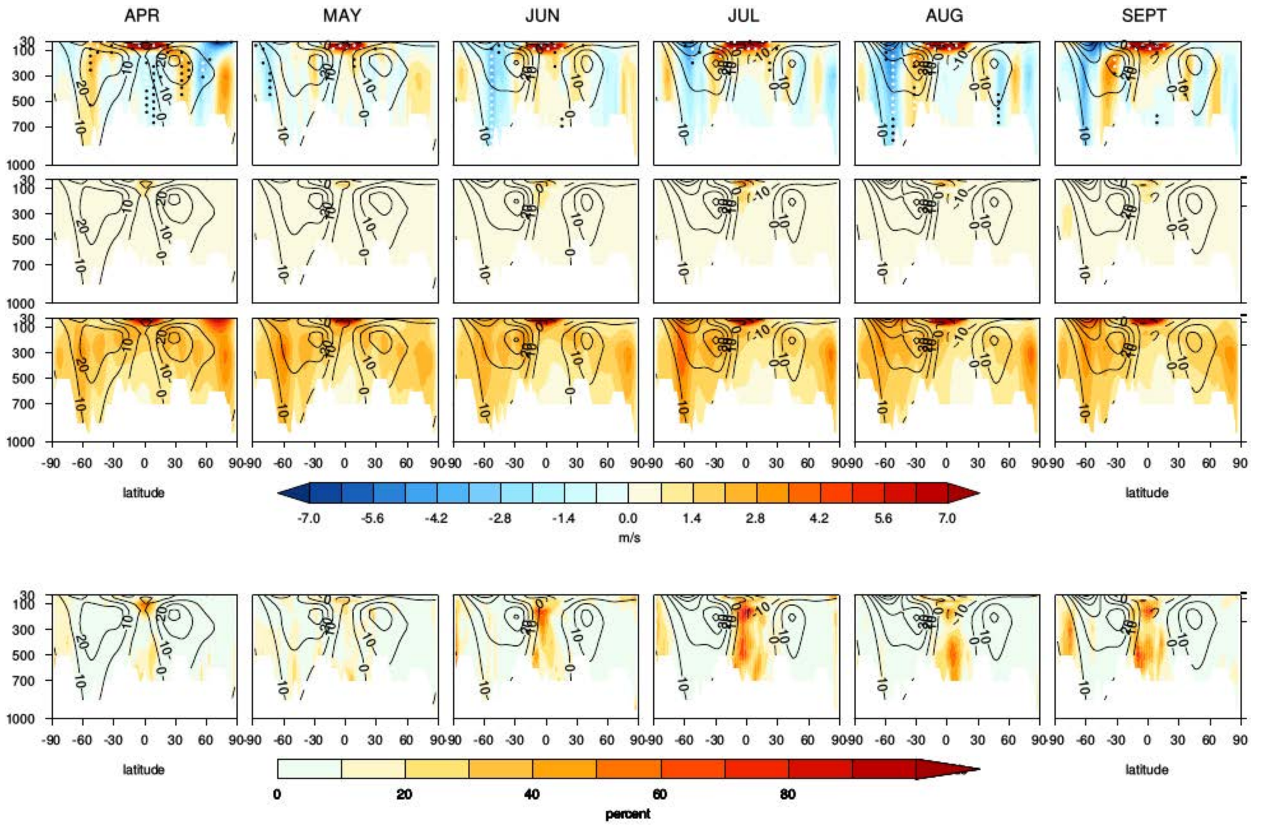


Figure 9.56: *As* ***Figure 9.55*** *but for the months April - September.*

For example, in the longer period the easterly response in the upper troposphere at the equator is substantially reduced in amplitude / significance in January - February, although the March signal is still robust; the November response in the NH mid-latitudes is no longer significant but the February NH subtropical response has increased in significance. Similarly, there are some changes to the the midlatitude responses in May - July between the two periods. There are also some small differences when compared with the results of *Gray et al.* (2018, see the lowermost row of their Figure 5) who combined the ERA-40 and ERA-Interim datasets to achieve a similar length dataset for 1958 - 2016 (but show results only for November - March). For example, the SH mid-latitude response in December from JRA-55 (1958 - 2016) is similar in pattern but is not significant in the ERA-40 / ERA-Interim analysis. The sensitivity of the QBO response to the data period suggests caution is required in their interpretation and additional years are required to verify whether these are real or not.

9.4.3 QBO teleconnections in context

In order to place the amplitude of the QBO signal into context, **Figures 9.59** and **9.60** shows the ENSO, volcanic and 11-yr solar signals from the MLR analysis of JRA-55 over the period 1958 - 2016 for each month of the year.

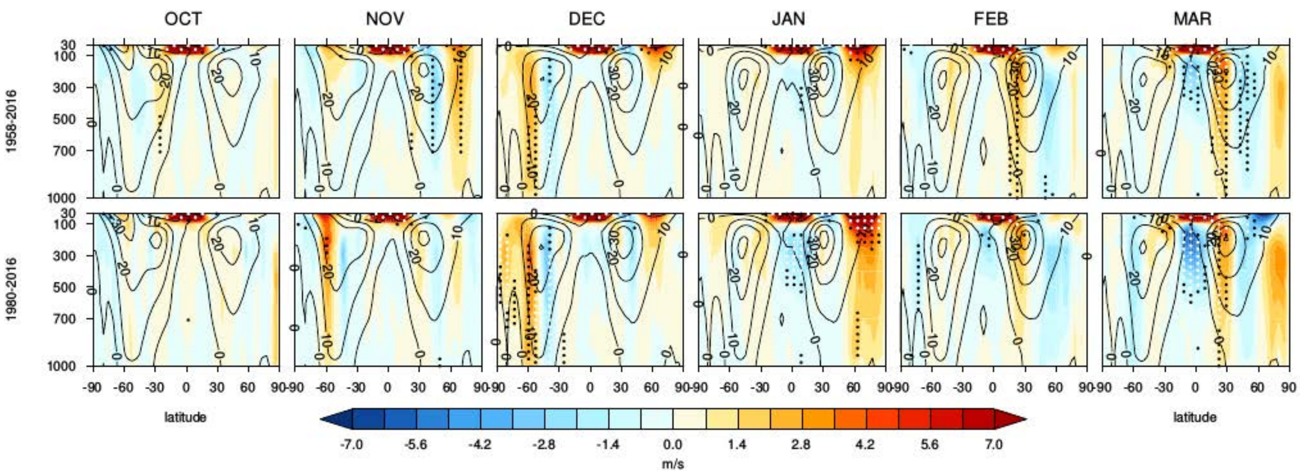


Figure 9.57: *October - March tropospheric QBO signals in zonally-averaged zonal winds (m s^{-1}) using JRA-55 over (top) the full period 1958 - 2016 versus (bottom) 1980 - 2016.*

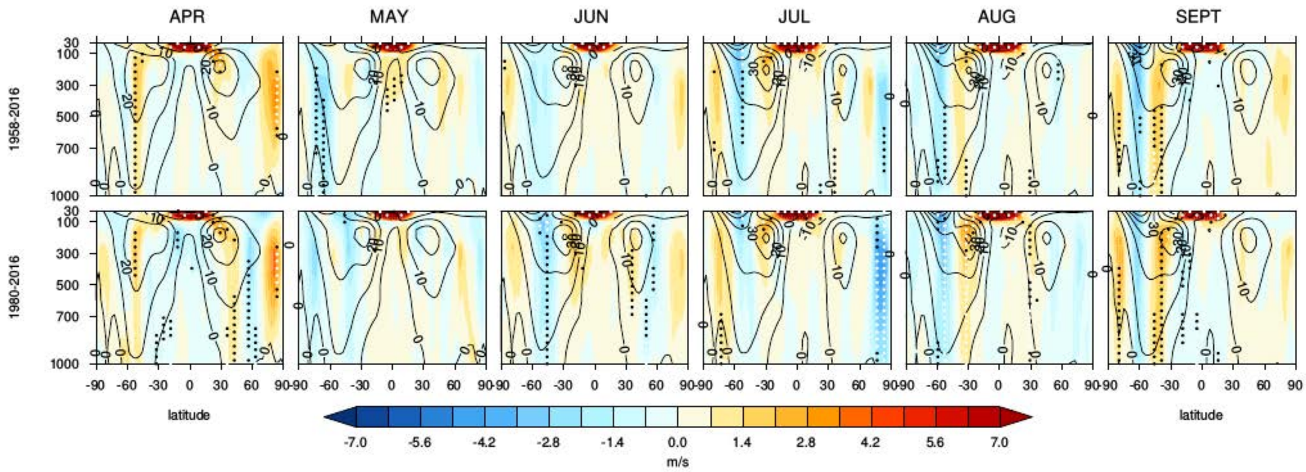


Figure 9.58: *As **Figure 9.57** but for the months April - September.*

The extended 1958 - 2016 period was chosen to maximise the number of solar cycles within the data period. The QBO index was defined as the contemporaneous FUB equatorial wind time-series at 50 hPa. In all cases the signal has been re-scaled to show the maximum likely amplitude *i.e.*, the difference between solar max and solar min in the largest amplitude solar cycle, the difference between the most extreme El Niño / La Niña, and the response to the largest volcanic eruption. (See **Figures AS9.39 - AS9.44** for the corresponding plots from the four recent reanalyses JRA-55, ERA-Interim, MERRA-2 and CFSR for the common data period 1980 - 2016, as an indication of the inter-reanalysis differences).

The solar cycle response (top row) is particularly uncertain because of the short data period relative to the period of the cycle and the lack of satellite data in the early period in the upper stratosphere. There are subtropical westerly anomalies in the upper stratospheric winter months of both hemispheres *e.g.*, at 20 °S near 1 hPa in June - August and near 20 °N near 3 hPa in November - December under solar maximum conditions, but these are barely significant. At polar latitudes the only apparent response is a weakened vortex in later winter (*e.g.*, in February in the NH and September - October in the SH). This is inconsistent with proposed mechanisms for solar influence on the vortex which predicts a strengthened polar vortex under solar max conditions (*Matthes et al.*, 2004; *Kodera and Kuroda*, 2002). A strengthened vortex is seen in January in the shorter postsatellite period (**Figure AS9.39**) but this is not statistically significant. There is good inter-reanalysis consistency between the signals in the shorter post-satellite era (**Figures AS9.39** and **AS9.40**) but nevertheless these signals are substantially reduced in amplitude and significance in the longer period.

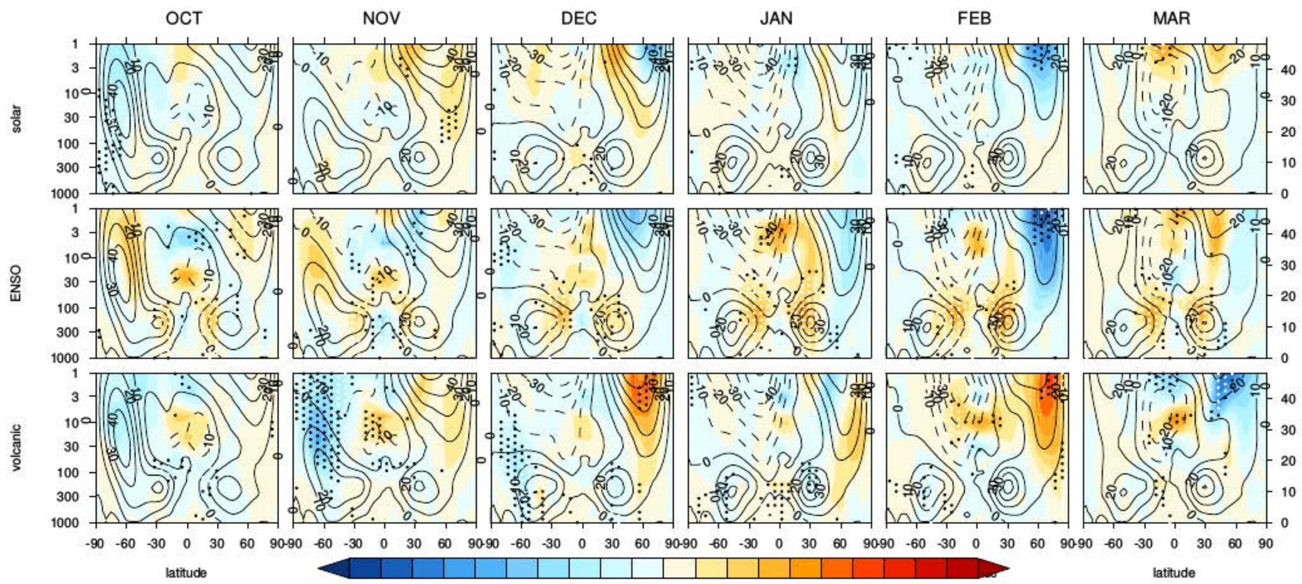


Figure 9.59: *11-yr solar cycle (top row), ENSO (middle) and volcanic signal (bottom row) in zonally averaged zonal winds (m s^{-1}) for October - March from the regression analysis of the JRA-55 dataset for 1958 - 2016 (with climatology and statistical significance levels overlaid). (The QBO index in the regression analysis was based on the FUB equatorial zonal wind time-series at 50 hPa).*

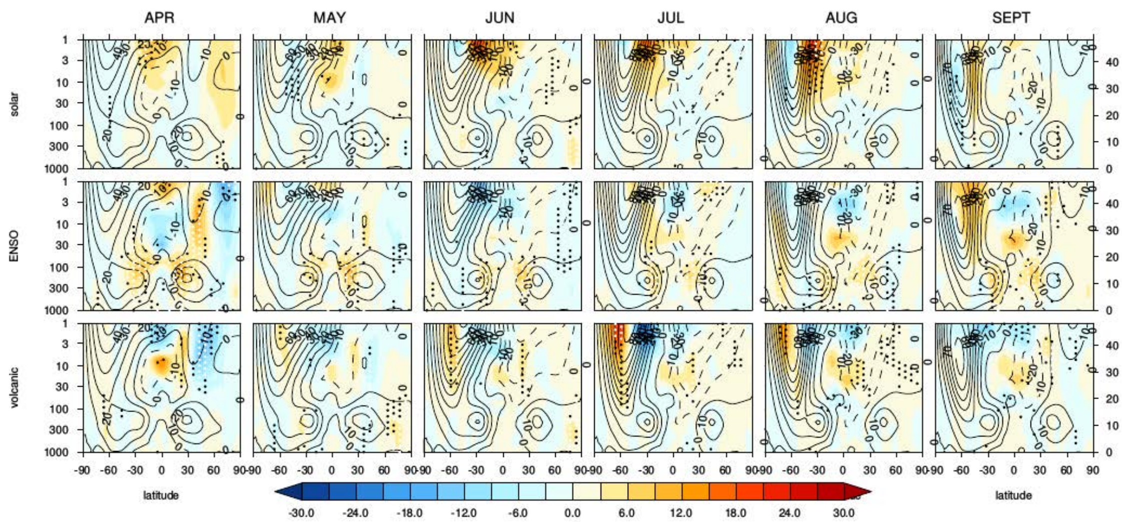


Figure 9.60: *As* ***Figure 9.59****, but for April - September.*

Comparisons in the previous sections have indicated that the assimilation of only conventional data in the pre-satellite era is sufficient to capture the polar vortex quite well (especially in the NH where the conventional data coverage is better). The consequent conclusion is that employing the longest possible data period, including both the pre- and post- satellite era, is preferable. The veracity of proposed mechanisms for solar cycle influence on the polar vortex therefore remains undetermined, and requires more years of data before they can be confirmed.

The ENSO response (middle row) is clearly evident in the tropical troposphere in all boreal winter months, as expected, and the subtropical westerly anomaly under ENSO conditions extends well into the stratosphere e.g. near 30°N in January - February. There is also a weakened NH winter polar vortex response in mid-late winter, in agreement with previous studies (see *e.g.*, *Butler and Polvani*, 2011); the December - February (DJF) months show a consistent weakening but only the February response is statistically significant. However, note that because of the consistency in the sign of the response, the DJF-averaged response is also likely to be significant. In the shorter postsatellite era (**Figures AS9.41** and **AS9.42**) there are similar responses; the amplitude of the responses has weakened considerably in the longer period but the significance of the February response is increased. This is a further demonstration of the difficulty of identifying a robust response in the presence of substantial background variability but in this case, in contrast to the solar cycle response, the longer data period confirms the signal and increases our confidence that it is real. There is also the suggestion of a weakened SH vortex in December in the shorter post-satellite era, but the amplitude and significance of this is reduced in the longer period, possibly due to the paucity of assimilated data in the early period. Additional years of observations will be required to confirm (or otherwise) this signal.

The volcanic response shows a strengthening of the NH mid-winter (December - February) polar vortex followed by a weaker vortex in March - April. This is in good agreement with previous studies that have shown a weakened vortex following major volcanic eruptions (*Stenchikov et al.*, 2006; *Shindell et al.*, 2004; *Robock*, 2000). A similar pattern is seen in the SH with mid-winter strengthening (June - August) followed by weakening in late winter (November - December). The latter suggests a possible influence on the timing of the final warming in each hemisphere but note that even the longer data period includes only 3 major equatorial volcanic eruptions with substantial amounts of aerosol reaching the stratosphere, so these results must be treated with caution.

9.4.4 Surface teleconnections

Sea Level Pressure

Figure 9.61 shows the QBO signal (QBO-W minus QBO-E) in mean sea level pressure for the period 1958–2016 from the regression analysis of the JRA-55 dataset. The JRA-55 dataset was examined because it provides the longest available data period using the same reanalysis system. The QBO index was defined as the contemporaneous FUB equatorial wind time-series at 50 hPa. **Figure 9.61** can be compared directly with **Figure 7** (5th row) of *Gray et al.* (2018), who examined the QBO in MSLP for the same period by combining the ERA-40 (1958 - 1978) with the ERA-Interim (1979 - 2016) datasets. The results are remarkably similar, demonstrating that either dataset is adequate for this purpose. The main responses are (a) a positive North Atlantic Oscillation (NAO)-like response in January, in which the southern node is statistically significant at the 95 % level but the northern node response is insignificant (nevertheless it shows the correct polarity for a positive NAO response; we note that the background variability increases substantially at higher latitudes); (b) a dipole response over the Pacific in March, with a region of reduced MSLP in QBO-W over the North Pacific and increased MSLP over the Equatorial Pacific. This response is similar to that found in other studies; for further discussion see *Gray et al.* (2018).

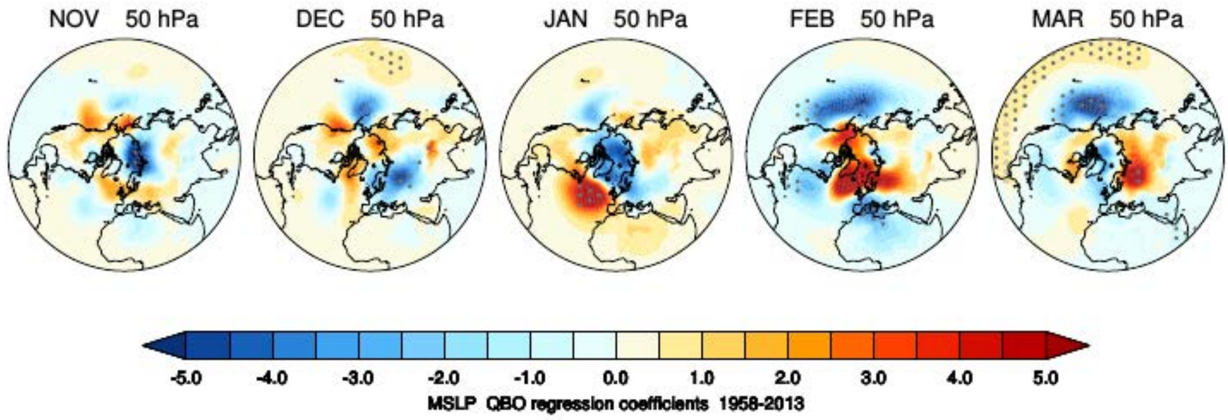


Figure 9.61: *Polar stereographic view of the NH winter (November - March) QBO response in mean sea level pressure (hPa) from the regression analysis of the JRA-55 dataset for the period 1958 - 2013. The QBO index was based on the FUB equatorial zonal winds at 50 hPa. The regression coefficients have been scaled to show the typical QBO-W minus QBO-E difference in zonal winds (to aid comparison with studies that employ composite difference techniques). Black (white) dots denote statistical significance at the 95 % (99 %) level.*

Precipitation

The QBO impact on tropical precipitation is examined in two of the modern reanalysis datasets for which a long data period is available *i.e.*, JRA-55 and the concatenated ERA-40 (1958 - 78) and ERA-Interim (1979 - 2016) datasets. The analysis follows earlier work that examined the signals in individual datasets and/or models (*e.g.*, *Gray et al.*, 2018; *Nie and Sobel*, 2015; *Liess and Geller*, 2012; *Ho et al.*, 2009; *Collimore et al.*, 2003; *Giorgetta et al.*, 1999) **Figure 9.62** shows an amended version of **Figure 9** from *Gray et al.* (2018) in which the annual-mean QBO signal in total precipitation is shown for a variety of different datasets using a QBO index that consists of the time-series of equatorial zonal winds at a single level (30 hPa, 50 hPa or 70 hPa) taken from the FUB zonal wind dataset. The GPCC observations (1979 - 2016; see *Section 9.1.2*) are shown in composite-difference form (1st column) as well the results from the MLR analysis (2nd column) in which the influences from ENSO, solar and volcanic forcings have been removed.

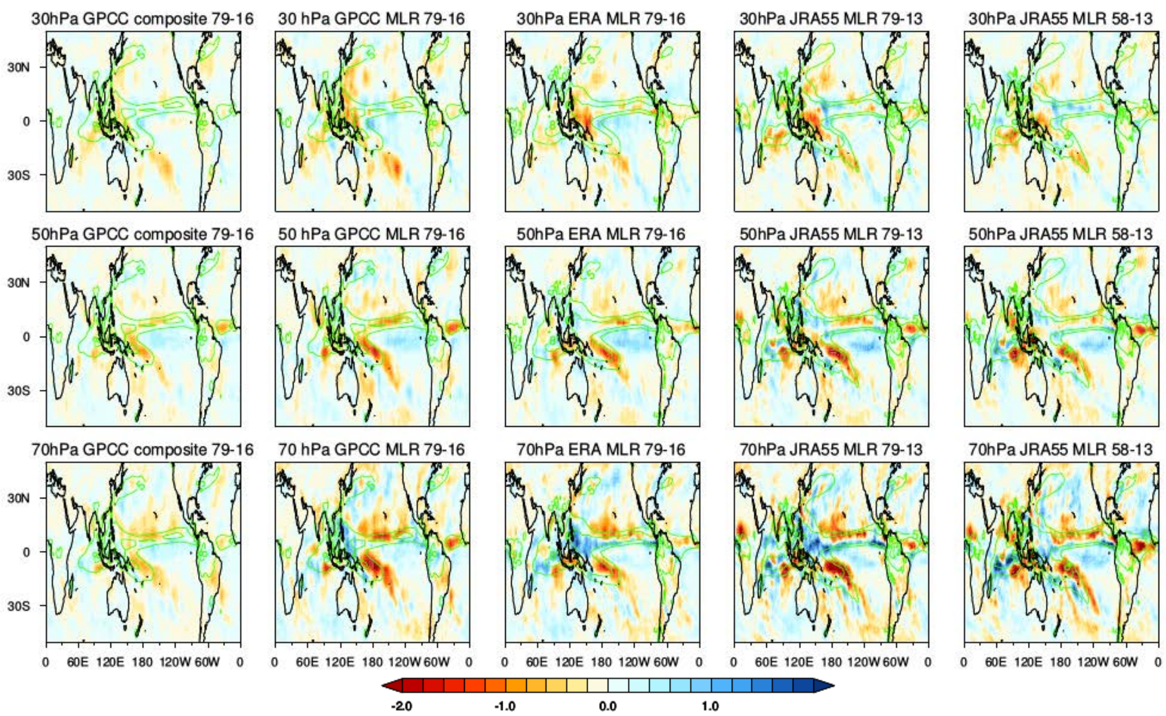


Figure 9.62: *Latitude-longitude distributions of QBO response in annual-averaged total precipitation (mm day⁻¹) using a QBO index defined as the time-series from a single level of the FUB zonally-averaged zonal wind dataset at the equator: 30 hPa (top row), 50 hPa (middle row) and 70 hPa (bottom row). 1st column: QBO-W minus QBO-E composite difference from the GPCC dataset (1979 - 2016); 2nd column: corresponding response but from the MLR analysis of the GPCC dataset (1979 - 2016); 3rd column: MLR analysis of the ERA-Interim dataset (1979 - 2016); 4th column: MLR analysis of the JRA-55 dataset (1979 - 2013); 5th column: MLR analysis of the JRA-55 dataset (1958 - 2013). Green contours indicate the climatological distribution for comparison.*

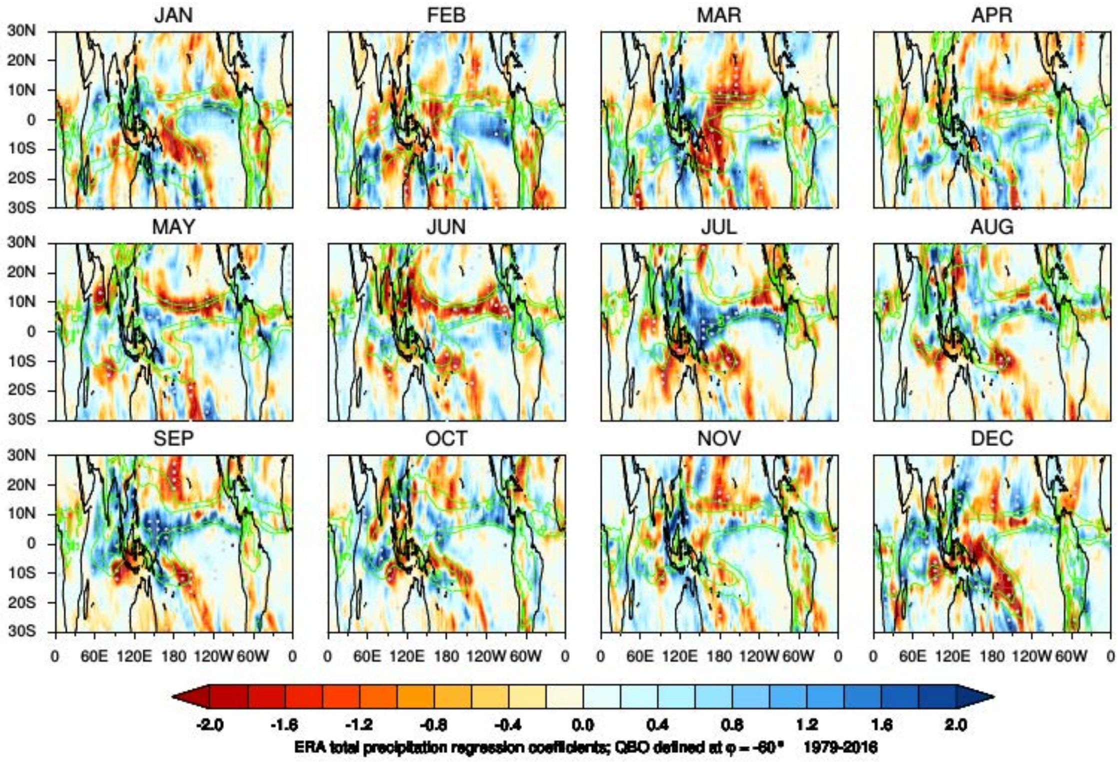


Figure 9.63: *Latitude-longitude distribution of QBO response in monthly-averaged total precipitation (mm day-1) for ERA-Interim, 1979 - 2016, with a QBO index defined by the EOF phase angle -60°, which is approximately equivalent to defining the QBO by the time-series at 70 hPa (as in the lower row of **Figure 9.62**). Filled contours show the regression coefficient for total precipitation, consistent with Gray et al. (2018). Green contours indicate the climatological distribution for comparison.*

The responses are largest when using a QBO index from the lowermost stratosphere at 70 hPa. The composite and MLR responses are similar in pattern but the MLR results have increased amplitudes and a clearer change in sign of the response over the Maritime Continent between the 30 hPa and 70 hPa levels (estimates of confidence levels are provided in later figures).

The corresponding annual-mean MLR responses from the ERA (3rd column) and JRA-55 (4th column) reanalyses for approximately the same period (1979 - 2016 and 1979 - 2013 respectively) are very similar to the GPCC responses and provide encouragement that the reanalyses can be used for investigation of the QBO signal in precipitation, despite the well-known difficulties associated with the representation of precipitation in the reanalyses. For completeness, the MLR response for the longer JRA-55 period (1958 - 2013) is also provided (5th column) and shows that the response is coherent across the different periods although the amplitude is slightly weaker, perhaps due to the poorer data coverage in the earlier period.

The major annual-mean QBO responses across all of these precipitation datasets are (a) increased precipitation over the eastern Maritime Continent for a QBO index at 70 hPa (centred around 150 °E over the equator) and (b) decreased precipitation along the band of maximum precipitation associated with the ITCZ; the latter suggests either an amplitude change (50 hPa) or a slight southward shift (70 hPa) of the ITCZ depending on the level of the QBO index.

Figures 9.63 and **9.64** shows the individual monthly-averaged precipitation signals that have contributed to the annual-mean responses. Instead of using a single level to define the QBO (such as the 30, 50 and 70 hPa levels used in **Figure 9.62**) we employ an EOF-based representation of the FUB equatorial wind time-series, that allows us to analyse the response to a particular vertical profile of the QBO rather than a single level. Results are shown for the two reanalysis datasets for the period since 1979, using an EOF phase angle (-60 °) that roughly equates to choosing a single-level indicator at 70 hPa; see *Gray et al.* (2018) for further details. While the individual months are clearly noisier, there is nevertheless reasonable agreement between the two datasets. The JRA-55 responses are slightly larger in amplitude (note the difference in scales), likely because the background climatological fields are larger (see **Figure 20** of (*Kobayashi et al.*, 2015). Both reanalyses show that the increase over the eastern Maritime Continent comes primarily from July - September.

A corresponding analysis of the convective component of the total precipitation from the two reanalysis datasets (**Figures AS9.45** and **AS9.46**) indicates that the QBO response is primarily in the convective component, since the total and convective precipitation responses are almost identical. Also shown in *Appendix A9.1* are the convective precipitation responses for the full available period from 1958 using the same EOF phase angle of -60 ° (**Figures AS9.47** and **AS9.48**). The two reanalyses show overall similar response patterns, although there are discrepancies in some months *e.g.*, March.

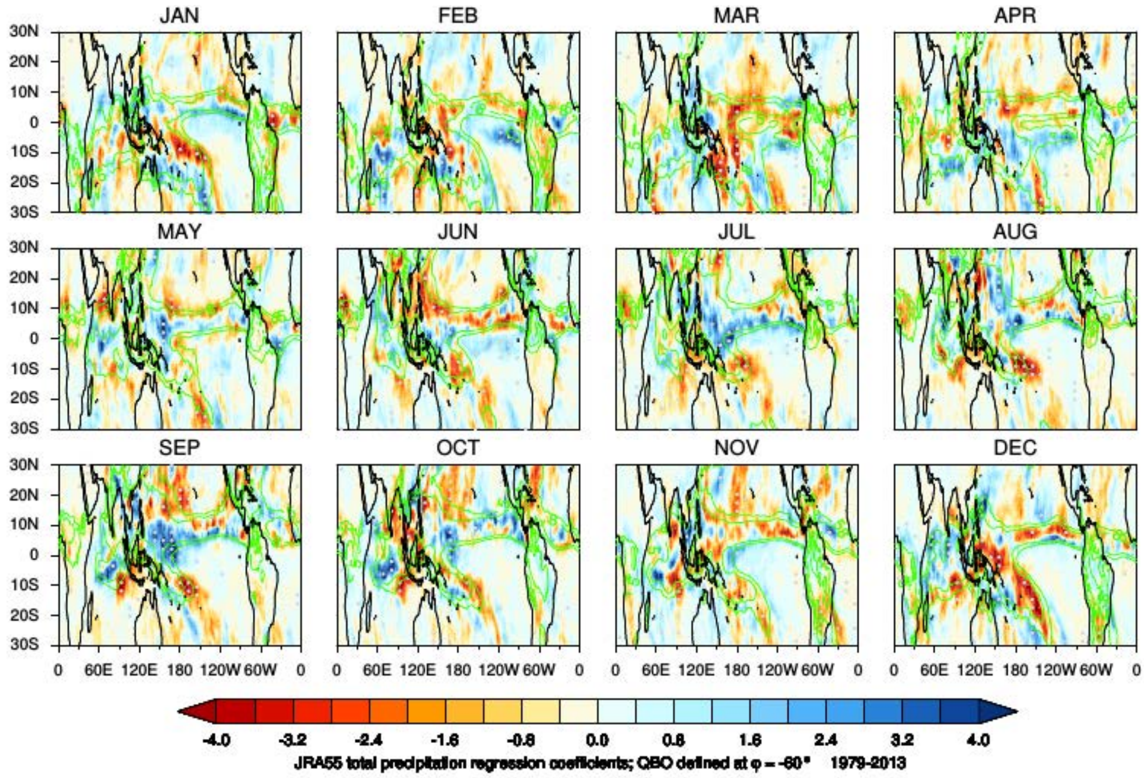


Figure 9.64: *As* ***Figure 9.63****, but for JRA-55, 1979 - 2013.*

The amplitudes are again slightly greater in the JRA-55 dataset and also the statistical significance of the responses (the latter is perhaps unsurprising since this dataset is more coherent than the combined ERA-40 / ERA-Interim dataset).

As noted in *Gray et al.* (2018) the QBO precipitation response along the main band of precipitation associated with the ITCZ is more clearly seen in the ERA dataset when the EOF phase-angle of +30° is employed to characterise the QBO; this is roughly equivalent to using the 50 hPa single-level index but also characterises the QBO profile with maximum vertical shear at the 70 hPa level. The corresponding analyses for 1958 - 2013 for this phase angle from both reanalysis datasets are shown in **Figures AS9.49** and **AS9.50** (note that small differences with **Figure 11** from *Gray et al.* (2018) are due to differences in the time period analysed, which has been truncated to 2013 in this report, to match the available data period of the JRA-55 dataset). Again, the response patterns are similar, but the JRA-55 patterns are less coherent than in the ERA dataset, and e.g. in July there is disagreement between the sign of the response over the equatorial Pacific.

9.5 Summary, key findings, and recommendations

Here we provide a concise summary of the main results from each section of the chapter, indicating which key figures illustrate these results.

9.5.1 Summary for monthly-mean equatorial variability

- Almost all of the reanalyses agree reasonably well with the FUB winds, and hence with each other, on the evolution of the zonal wind QBO. The older NCEP reanalyses (NCEP-NCAR and NCEP-DOE) are an exception, although even in these cases the phase of the QBO is usually correct; the main error is that the QBO wind amplitude is substantially underestimated (by up to a factor of 2, depending on the altitude considered; **Figure 9.20b**). We attribute the good representation to the primary importance of tropical radiosonde wind observations in constraining the tropical stratospheric winds up to altitudes of 10 hPa. This is evidenced by the excellent agreement between JRA-55 and JRA-55C reanalyses (**Figure 9.5**), as well as by the fact that extended reanalyses such as ERA-40 and JRA-55 agree well with each other and the FUB winds in the pre-satellite era. The importance of wind observations is anticipated on the basis that the QBO mechanism requires a zonal momentum source (as analysis increments due to wind observations would provide) and that previous studies have indicated the importance of wind observations for good representation of the tropical winds (*e.g.*, *Hersbach et al.*, 2017; *Kobayashi et al.*, 2014).

- The inter-reanalysis spread has decreased over time (**Figure 9.4**), which is consistent with increasing availability of observations to constrain the reanalyses. However the differences between JRA-55 and JRA-55C do not show any long-term trend (**Figure 9.5**), indicating that the increasing amount of satellite data assimilated into JRA-55 over the 1973 - 2012 period does not improve the agreement between the two reanalyses, bolstering the conclusion that satellite observations are much less important than conventional observations for the QBO.

- Most inter-reanalysis spread occurs during QBO phase transitions, and in particular during QBO-W (westerly phase) onsets (**Figure 9.4**), during which the phase onset is often delayed by ≈ 1 - 2 months in comparison to FUB winds. QBO-W onsets are also delayed with respect to the MERRA-2 reanalysis (**Figures 9.13, 9.14**), which uses a forecast model that spontaneously generates a QBO, mainly due to tuning of the non-orographic gravity wave drag parameterization (*Coy et al.*, 2016). Hence we attribute the systematically delayed QBO-W onsets in the other reanalyses (*i.e.*, all of them except for MERRA-2) to the lack of a sufficiently strong westerly momentum source in the tropical stratosphere, which can only be provided by wave drag.

- There is substantial uncertainty (*i.e.*, inter-reanalysis spread) in the strength and spatial structure of zonal winds in the tropical upper troposphere and tropical tropopause region, both for the zonal-mean (**Figure 9.8**) and the zonally varying (**Figure 9.9**) component. This has implications for the modelling of tropical wave propagation in terms of how these background winds influence (filter) the upward propagation of waves that force the QBO and SAO, including parameterized gravity waves (since small changes in wave filtering at lower altitudes can have substantial effects on wave forcing at higher altitudes).

- There is uncertainty regarding how much zonal asymmetry is present in the QBO, especially at 70 hPa, given that the assimilation of radiosonde winds in the tropics is dominated by the contribution from the Singapore station; the inter-reanalysis spread is greatest over the oceans where there is a lack of radiosonde observations (**Figure 9.11**). Introduction of more spatially homogeneous coverage of wind data could address this. Although the inter-reanalysis spread has reduced over time (as noted above), its spatial pattern remains unchanged (**Figure 9.12**), and is especially evident at 70 hPa (where the flow is less zonally symmetric than at higher levels).

- The vertical velocity anomaly associated with the QBO is comparable to background vertical velocity, though the magnitudes of both vary among reanalyses (**Figures 9.24, 9.25**).

- There is good representation of the QBO temperature anomaly evolution when compared with sondes and GNSS-RO (note that all reanalyses considered here assimilate radiosondes, and the "modern four" assimilate GNSS-RO data although over slightly different periods). Peak-to-peak QBO zonal-mean temperature variations are ≈ 2 K and 1 K at 70 hPa and near the tropical tropopause (100 hPa), respectively, corresponding to roughly 25 - 30 % and 15 - 20 % the size of the annual cycle. Zonal asymmetries are evident in the temperature signal, with QBO amplitude in the Indonesian region roughly 30 % larger than the zonal-mean amplitude. Comparison with GNSS-RO, which are spatially homogeneous, suggests that this is a real feature rather than an artefact of the strong influence of Singapore observations on reanalysis QBOs. This may have implications for QBO influence on convection / precipitation.

9.5.2 Summary for tropical waves and QBO forcing

- There is good agreement between the reanalyses on the relative contributions of the various tropical waves to the forcing of the QBO (**Figure 9.38**). The greatest inter-reanalysis spread is in the Kelvin wave contribution during the QBO-W descending phase. There is significant natural variability (*i.e.*, from one QBO cycle to the next) in the various contributions. The vertical advection term differs widely between reanalyses, including in its sign (consistent with large inter-reanalysis differences in vertical velocity, **Figure 9.24**).

- Although the assimilation of satellite observations does not have a major impact on the representation of the QBO wind evolution (see *Section 9.2* summary), it nevertheless has an indirect impact via improved representation of the different components of the waves that force the QBO, which may contribute to improvements in details such as the spread in the timing of the QBO phase changes referred to above. There is clear evidence that the representation of tropical waves in the reanalyses has changed after the introduction of the AMSU satellite observations in ≈1998 (**Figure 9.34**) and assuming that the observations are more accurate in the latter period we recommend that the more recent data are used for studies of wave diagnostics.

- There are also clear differences in the wave characteristics when derived on model versus pressure surfaces (**Figures 9.32, 9.33**) - recommendation is the use of model levels wherever possible and be aware of limitations if pressure levels are used. Qualitative results are similar in the two cases, but for quantitative results model levels are better, so as not to lose information due to vertical interpolation.

- Comparison of the wave characteristics with satellite observations (HIRDLS, SABER, COSMIC, and AIRS) shows consistency between the reanalyses and high correlation in the lower tropical stratosphere with all the observations except AIRS (**Figures 9.40, 9.42**). The correlations with HIRDLS and SABER are notable because these observations are not assimilated by any of the reanalyses and thus provide independent validation of the reanalyses. Reanalysis momentum fluxes in the lower tropical stratosphere correlate well with HIRDLS but less well with SABER. This suggests good estimates from the reanalyses since HIRDLS is generally regarded as a better instrument than SABER.

9.5.3 Summary for QBO teleconnections

- There is good inter-reanalysis agreement on the representation of the QBO influence on the NH winter polar vortex. A clear impact is evident in early winter (November - January), with a stronger (colder) vortex when the lower stratospheric winds are in the QBO westerly (QBO-W) phase than the QBO easterly (QBO-E) phase, the expected well-known Holton-Tan effect (**Figure 9.43**). An apparent late winter reversal of this response (February - March) seen in the 1979 - 2016 analysis is not robust since it no longer appears when the longer 1958 - 2016 period is analysed, highlighting the importance of using as long a data record as possible. There is some suggestion of a QBO impact on the timing of the final NH warming, with an earlier reversal of the winds under QBO-W conditions, but more years of data are required to verify this.

- There is no evidence for a QBO influence on the early winter or midwinter strength of the SH vortex. The final warming of the SH vortex occurs later during QBO-W than QBO-E, when the QBO phase is defined using 20 hPa QBO winds. Although the lack of observations to constrain reanalyses in the SH stratosphere during the pre-satellite era suggests caution when examining the vortex response during the extended record 1958 - 2016, the response is very similar to that obtained using the satellite-era only (**Figures 9.44**).

- In boreal winter (December - April) there is a QBO impact on the strength of the tropical upper tropospheric winds of ≈ 4 - $5\,\mathrm{m\,s^{-1}}$, of opposite sign to the overlying QBO phase in the lower stratosphere (**Figures 9.55**, **9.56**). This is accompanied by a strengthening of the subtropical jet in the winter hemisphere, near 30°N in February - March (**Figure 9.55**) and 30°S in August - September (**Figure 9.56**). There is good agreement of this signal over the period 1980 - 2016 in the four recent full-input reanalyses (ERA-Interim, MERRA-2, JRA-55, CFSR) but some details of the response are not robust when the longer period 1958 - 2016 is examined. A longer record is therefore required to verify whether this signal is real or not.

- A QBO modulation of mean sea level pressure (MSLP) is found in NH winter over the extended 1958 - 2016 period in the JRA-55 reanalysis (**Figure 9.61**), almost identical to the response found by *Gray et al.* (2018) who examined a similarly long data record by combining the ERA-40 and ERA-Interim reanalyses. This demonstrates that choosing either of these methods for achieving a long data period is adequate for MSLP. The main QBO-W minus QBO-E responses are (a) a positive NAO-like response in January and (b) a dipole response over the Pacific in March, with a region of reduced MSLP in QBO-W over the North Pacific and increased MSLP over the Equatorial Pacific.

- A QBO modulation of tropical precipitation is observed in both JRA-55 and ERA-Interim reanalyses over the satellite era, and both compare well with independent GPCC satellite observations (**Figure 9.62**). The response is mostly robust to the inclusion of pre-satellite years using JRA-55. The major QBO-W minus QBO-E responses are (a) increased precipitation over the eastern Maritime Continent and (b) decreased precipitation along both the band of maximum precipitation associated with the ITCZ in the tropical Pacific Ocean, as well as in the South Pacific Ocean northeast of Australia. Since this response occurs in a region known to be strongly influenced by ENSO (*e.g.*, *Son et al.*, 2017), care is required to separate the QBO response, for example by using a multi-linear regression approach as was done here. The overall strongest response is found when the QBO is characterized using the 70 hPa equatorial winds.

9.5.4 Recommendations

In this final section we provide recommendations, based on the results described in the chapter, on which reanalyses are appropriate to use for various diagnostics of the QBO and tropical stratospheric variability. A summary of recommendations is given in **Figure 9.65**, classifying each reanalysis for each diagnostic into one of five cases, and discussion of the recommendations follows below. The large number of unevaluated cases in **Figure 9.65** (tan colour) indicates simply that the required data was not available, or that it was judged not worthwhile to perform the diagnostic for that reanalysis; the relevant sections, indicated by the "Section" column at left, provide more information. For the particular case of ERA5, the time period that would be required for most diagnostics (*e.g.*, the 1980 - 2012 period for the standard QBO metrics of *Section 9.2.2*) was not available at the time of writing the report.

The comparison with satellite instruments in *Section 9.3.3* is an exception because it required only recent years overlapping with the satellite records, which were available earlier than the 1979 - onward period of ERA5.

Recommendations for reanalysis users:

- For determination of QBO phase – that is, whether the prevailing tropical stratospheric zonal-mean zonal winds are westerly or easterly – most reanalyses are suitable but they may disagree near the transition times. The reanalysis that agrees best with the FUB zonal wind record at 30 hPa is MERRA-2, suggesting that it might have the most accurate transition times at these altitudes provided that FUB can be assumed representative of the zonal mean. However, MERRA-2 may be a poor choice for determining the 10 hPa QBO phase as it appears to have unusual features earlier in its record. Nevertheless, almost all reanalyses correlate very highly with FUB winds and with each other, so that almost any of them (except perhaps the older ones: NCEP-NCAR, NCEP-DOE, JRA- 25) are very suitable for determination of QBO phase at any level in the tropical lower stratosphere. Agreement between reanalyses is best at the middle QBO levels, 20 - 50 hPa.

- For characterization of the QBO (amplitude, period, *etc.*) conventional-input reanalyses represent the QBO well provided tropical radiosonde data are assimilated. JRA-55C appears to be as suitable for examining the QBO as JRA-55 (*Kobayashi et al.*, 2014), although its record is slightly shorter. Surface-input reanalyses such as ERA-20C have not been considered here and should not be used to examine the QBO. If a QBO exists in such reanalyses then it will be entirely produced by the forecast model, and even if the model's QBO is realistic, the lack of assimilated tropical stratospheric wind observations implies that it will generally not reproduce the observed QBO phase.

- For comparison of QBO characteristics with climate models (amplitude, period, *etc.*) the modern reanalyses are most suitable because they have improved QBO representations. JRA-55 is a good choice because it provides the longest record of any full-input modern reanalysis, thus providing the most statistically robust estimates of QBO characteristics. MERRA-2 may also be a good choice because its representation of the QBO does not rely on the data assimilation to correct a severe model bias, *i.e.*, the lack of a QBO, but (at least based on the diagnostics presented here) this may not be important for most applications, and the caveat about the 10 hPa level winds (see above) should also be noted. CFSR is less suitable than JRA-55, MERRA-2 or ERA-Interim because it underestimates the QBO amplitude compared to other reanalyses.

- For studies of tropical stratospheric temperature and meridional wind spectra the modern reanalyses are all similar and therefore all are suitable. However, estimates of QBO wave forcing (*i.e.*, Eliassen-Palm flux divergence) show more variation across reanalyses and it is not clear which of these are most accurate. Comparisons of QBO forcing in climate models with reanalyses should take account of the inter-reanalysis spread by using more than one reanalysis where possible, and should also note the very large natural variability of QBO forcing terms. The vertical advection term is particularly uncertain because vertical velocity in the lower tropical stratosphere shows large inter-reanalysis variations. For the most quantitatively accurate wave diagnostics, model levels are better than pressure levels since wave quantities can be effectively damped by vertical interpolation. However, pressure-levels diagnostics were found to capture the same qualitative variations as model-levels diagnostics. The post-1998 period is likely more reliable for evaluating wave spectra and QBO wave forcing.

- For investigation of QBO teleconnections, including impacts on the winter polar vortex strength, tropospheric circulation, surface pressure and precipitation, the length of the data period needs to be as long as possible in order to maximise the signal-to-noise ratio, for example using the JRA- 55 dataset for the period 1958 onwards or concatenating the ERA-40 and ERA-Interim datasets (although for some studies even this may not be sufficient). Particular care is required to distinguish surface impacts of the QBO from ENSO impacts. While using pre-satellite era data to extend the data period is especially recommended for analysis of features at levels below ≈ 10 hPa since these have the benefit of conventional data input, extra care is required in the interpretation of QBO impacts at levels higher than 10 hPa. Comparing results from the pre- and post-satellite era separately is recommended. However, when examining QBO influence on the SH polar vortex, pre-satellite data should be used with caution because of the poor coverage of ground-based data.

Recommendations for reanalysis data providers:

- We recommend that reanalysis centres include 15 hPa and 40 hPa levels as standard output levels. The QBO amplitude peaks at 15 hPa in the FUB data, so model-reanalysis comparisons require this level for accurate validation of the models. The 40 hPa level, which is also in the FUB data, is highly correlated with the NH polar vortex response, and was the level at which the unusual easterly layer (the "QBO disruption") first emerged during 2015/16 NH winter.

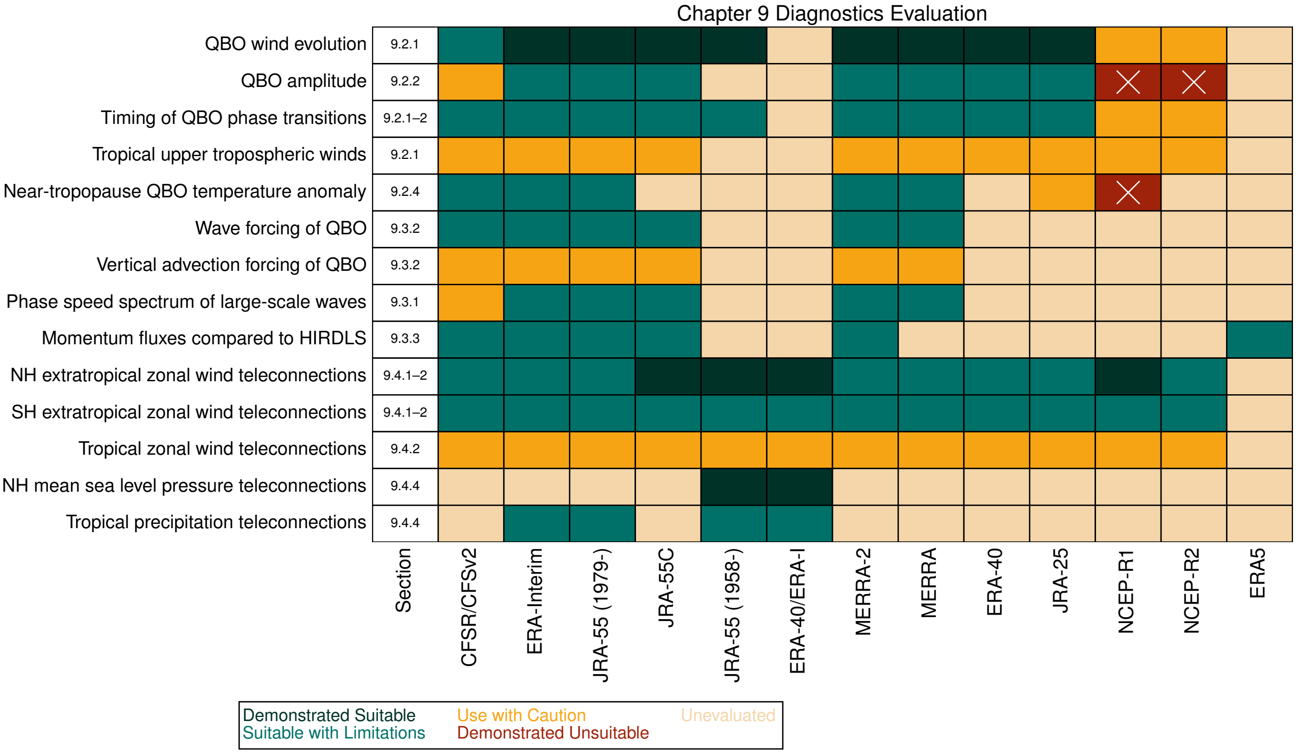


Figure 9.65: *Appropriateness of different reanalyses for various diagnostics presented in Sections 9.2-9.4. The "Section" column at left indicates where in the chapter each diagnostic is described. (Note, NCEP-NCAR is NCEP-NCAR and NCEP-DOE is NCEP-DOE).*

Code availability

QBO metrics calculations (**Figure 9.15**) were based on code available at **https://github.com/vschenzinger/QBO-Metrics**.

Acknowledgements

We thank the reanalysis centres for providing their support and data products, and we thank the British Atmospheric Data Centre (BADC) of the UK Centre for Environmental Data Analysis (CEDA) for providing the S-RIP group workspace on their JASMIN service. We are grateful to the UK Natural Environment Research Council (NERC) for funding QBO workshops in 2015 and 2016 that helped support the analysis presented in this chapter. We acknowledge the EGU journal Atmospheric Chemistry and Physics for allowing reproduction or adaptation of published figures under the Creative Commons Attribution 4.0 License (**https://creativecommons.org/licenses/by/4.0/**).

References

Alexander, M.J., *et al.*, 2010: Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Quart. J. Roy. Meteor. Soc.*, **136**, 1103 - 1124,doi: 10.1002/qj.637.

Alexander, S.P., *et al.*, 2016: Southern Hemisphere Extratropical Gravity Wave Sources and Intermittency Revealed by a Middle-Atmosphere General Circulation Model. *J. Atmos. Sci.*, **73**, 1335 - 1349, doi: 10.1175/JASD-15-0149.1.

Andrews, D.G., J.R. Holton,and C.B. Leovy, 1987: Middle Atmosphere Dynamics. Academic Press, San Diego, 1987.

Anstey, J.A. and T.G. Shepherd, 2014: High-latitude influence of the quasi-biennial oscillation. *Quart. J. Roy. Meteor. Soc.*, **140**, 1–21, doi: 10.1002/qj.2132.

Baldwin, M.P. and T.J. Dunkerton, 1998: Biennial, quasi-biennial, and decadal oscillations of potential vorticity in the northern stratosphere. *J. Geophys. Res.*, **103**, 3919 - 3928, doi: 10.1029/97JD02150.

Baldwin, M.P. and T.J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581 - 584, doi: 10.1126/science.1063315.

Baldwin, M.P., *et al.*, 2001: The quasi-biennial oscillation. *Rev. Geophys.*, **39**, 179 - 229, oi: 10.1029/1999RG000073.

Boville, B. and W.J. Randel, 1992: Equatorial waves in a stratospheric GCM: effects of vertical resolution. *J. Atmos. Sci.*, **49**, 785 - 801, doi: 10.1175/1520-0469(1992)049<0785:EWIASG>2.0.CO;2.

Bushell, A.C., *et al.*, 2020: Evaluation of the Quasi-Biennial Oscillation in global climate models for the SPARC QBOinitiative. *Quart. Jo Roy. Meteor.Soc.*, 1 - 31, doi: 10.1002/qj.3765.

Butchart, N., *et al.*, 2018: Overview of experiment design and comparison of models participating in phase 1 of the SPARC Quasi-Biennial Oscillation initiative (QBOi). *Geosc. Mod. Dev.*, **11**, 1009 - 1032, doi: 10.5194/gmd-11-1009-2018.

Butler, A.H. and L.M. Polvani, 2011: El Niño, La Niña, and stratospheric sudden warmings: A reevaluation in light of the observational record. *Geophys. Res. Lett.*, **38**, doi: 10.1029/2011GL048084.

Collimore, C.C., *et al.*, 2003: On the relationship between the QBO and tropical deep convection. *J. Climate*, **16**, 2552 - 2568, doi: 10.1175/1520-0442(2003)016<2552:OTRBTQ>2.0.CO;2.

Coy, L., *et al.*, 2016: Structure and Dynamics of the Quasi-Biennial Oscillation in MERRA-2. *J. Climate*, **29**, 5339 - 5354, doi: 10.1175/JCLI-D-15-0809.1.

Coy, L., P.A. Newman, S. Pawson, and L.R. Lait, 2017: Dynamics of the Disrupted 2015/16 Quasi-Biennial Oscillation, *J. Climate*, **30**, 5661 - 5674, doi: 10.1175/JCLI-D-16-0663.1.

Crooks, S.A. and L.J. Gray, 2005: Characterization of the 11-Year Solar Signal Using a Multiple Regression Analysis of the ERA-40 Dataset, *J. Climate*, **18**, 996 - 1015, doi: 10.1175/JCLI-3308.1.

Das, S.S., *et al.*, 2016: Four-decadal climatological intercomparison of rocketsonde and radiosonde with different reanalysis data: results from Thumba Equatorial Station. *Quart. J. Roy. Meteor. Soc.*, **142**, 91 - 101, doi: 10.1002/qj.2632.

Dunkerton, T.J., 1990: Annual variation of deseasonalized mean flow acceleration in the equatorial lower stratosphere. *J. Meteor. Soc. Japan*, **68**, 499 - 508, doi: 10.2151/jmsj1965.68.4_499.

Dunkerton, T.J. and D.P. Delisi, 1985: Climatology of the equatorial lower stratosphere. *J. Atmos. Sci.*, **42**, 376 - 396, doi: 10.1175/1520-0469(1985)042<0376:COTELS>2.0.CO;2.

Ern, M. and P. Preusse,2009: Wave fluxes of equatorial Kelvin waves and QBO zonal wind forcing derived from SABER and ECMWF temperature space-time spectra. *Atmos. Chem. Phys.*, **9**, 3957 - 3986, doi: 10.5194/acp-9-3957-2009.

Ern, M., *et al.*, 2008: Equatorial wave analysis from SABER and ECMWF temperatures. *Atmos. Chem. Phys.*, **8**, 845 - 869, doi: 10.5194/acp-8-845-2008.

Ern, M., *et al.*, 2014: Interaction of gravity waves with the QBO: A satellite perspective. *J. Geophys. Res. Atmos.*, **119**, 2329 - 355, doi: 10.1002/2013JD020731.

Fujiwara, M., *et al.*, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. A*tmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp-17-1417-2017.

Fujiwara, M., *et al.*, 2020: Singapore upper air station visited by SPARC researchers. *SPARC Newsletter No*. **54**, pp. 42 - 47, 2020, available at: **https://www.sparc-climate.org/publications/newsletter/**.

Garfinkel, C.I. and D.L. Hartmann, 2011a: The Influence of the Quasi-Biennial Oscillation on the Troposphere in Winter in a Hierarchy of Models. Part I: Simplified Dry GCMs. *J. Atmos. Sci.*, **68**, 1273 - 1289, doi: 10.1175/2011JAS3665.1.

Garfinkel, C.I. and D.L. Hartmann, 2011b: The Influence of the Quasi-Biennial Oscillation on the Troposphere in Winter in a Hierarchy of Models. Part II: Perpetual Winter WACCM Runs. *J. Atmos. Sci.*, **68**, 2026 - 2041, doi: 10.1175/2011JAS3702.1.

Giorgetta, M.A., L. Bengtsson, and K. Arpe, 1999: An investigation of QBO signals in the east Asian and Indian monsoon in GCM experiments. *Clim. Dyn.*, **15**, 435 - 450, doi: 10.1007/s003820050292.

Gray, L.J., *et al.*, 2013: A lagged response to the 11 year solar cycle in observed winter Atlantic/European weather patterns. *J. Geophys. Res. Atmos.*, **118**, 13,405 - 13,420, doi: 10.1002/2013JD020062.

Gray, L.J., T.J. Woollings, M. Andrews, and J. Knight, 2016: Eleven-year solar cycle signal in the NAO and Atlantic/European blocking. *Quart. J. Roy. Meteor. Soc.*, **142**, 1890 - 1903, doi: 10.1002/qj.2782.

Gray, L.J., *et al.*, 2018: Surface impacts of the Quasi Biennial Oscillation. *Atmos. Chem. Phys.*, **18**, 8227 - 8247, doi: 10.5194/acp-18-8227-2018.

Hampson, J. and P. Haynes, 2004: Phase Alignment of the Tropical Stratospheric QBO in the Annual Cycle. *J. Atmos. Sci.*, **61**, 2627 - 2637, oi: 10.1175/JAS3276.1.

Hersbach, H., *et al.*, 2017: The potential value of early (1939 - 1967) upper-air data in atmospheric climate reanalysis. *Quart. J. Roy. Meteor. Soc.*, **143**, 1197 - 1210, doi: 10.1002/qj.3040.

Ho, C.-H., H.-S. Kim, J.-H. Jeong, and S.-W. Son, 2009: Influence of stratospheric quasi-biennial oscillation on tropical cyclone tracks in the western North Pacific. *Geophys. Res. Lett.*, **36**, doi: 10.1029/2009GL037163.

Holton, J.R. and H.C. Tan, 1980: The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb. *J. Atmos. Sci.*, **37**, 2200 - 2207, doi: 10.1175/1520-0469(1980)037<2200:TIOTEQ>2.0.CO;2.

Holton, J.R. and H.C. Tan, 1982: The quasi-biennial oscillation in the Northern Hemisphere lower stratosphere. *J. Meteor. Soc. Japan*, **60**, 140 - 147, doi: 10.2151/jmsj1965.60.1_140.

Horinouchi, T., *et al.*, 2003: Tropical cumulus convection and upward-propagating waves in middle-atmospheric GCMs. *J. Atmos. Sci*., **60**, 2765 - 2782, doi: 10.1175/1520-0469(2003)060<2765:TCCAUW>2.0.CO;2.

Huesmann, A.S. and M.H. Hitchman, 2001: The stratospheric quasi-biennial oscillation in the NCEP reanalyses: Climatological structures. *J. Geophys. Res.*, **106**, 11 859 - 11 874, doi: 10.1029/2001JD900031.

Kalnay, E., *et al.*, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.*, **77**, 437 - 472, doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kawatani, Y. and K. Hamilton, 2013: Weakened stratospheric quasi-biennial oscillation driven by increased tropical mean upwelling. *Nature*, 497, 478-481, doi: 10.1038/nature12140.

Kawatani, Y., *et al.*, 2010: The roles of equatorial trapped waves and internal inertia-gravity waves in driving the quasi-biennial oscillation. Part I: zonal mean wave forcing. *J. Atmos. Sci.*, **67**, 963 - 980, doi: 10.1175/2009JAS3222.1.

Kawatani, Y., *et al.*, 2016: Representation of the tropical stratospheric zonal wind in global atmospheric reanalyses. Atmos. Chem. Phys., **16**, 6681 - 6699, doi: 10.5194/acp-16-6681-2016.

Kidston, J., *et al.*, 2015: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nat. Geosci*., **8**, 433 - 440, doi: 10.1038/NGEO2424.

Kiladis, G.N., *et al.*, 2009: Convectively coupled equatorial waves. *Rev. Geophys.*, **47**, RG2003, doi: 10.1029/2008RG000266.

Kim, D., *et al.*, 2014: Representation of tropical subseasonal variability of precipitation in global reanalyses. *Clim. Dyn.*, **43**, 517 - 534, doi: 10.1007/s00382-013-1890-x.

Kim, J.-E. and M.J. Alexander, 2013: Tropical Precipitation Variability and Convectively Coupled Equatorial Waves on Submonthly Time Scales in Reanalyses and TRMM. *J. Clim.*, **26**, 3013 - 3030, doi: 10.1175/JCLI-D-12-00353.1.

Kim, Y.-H. and H.-Y. Chun, 2015: Momentum forcing of the quasi-biennial oscillation by equatorial waves in recent reanalyses, *Atmos. Chem. Phys.*, **15**, 6577–6587, **doi: 10.5194/acp-15-6577-2015**.

Kim, Y.-H., *et al.*, 2019: Comparison of equatorial wave activity in the tropical tropopause layer and stratosphere represented in reanalyses. *Atmos. Chem. Phys*., **19**, 10 027 - 10 050, doi: 10.5194/acp-19-10027-2019.

Kobayashi, C., *et al.*, 2014: Preliminary Results of the JRA-55C, an Atmospheric Reanalysis Assimilating Conventional Observations Only. *SOLA*, **10**, 78 - 82, doi: 10.2151/sola.2014-016.

Kobayashi, S., *et al.*, 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *J. Meteor. Soc. Japan*, **93**, 5 - 48, doi: 10.2151/jmsj.2015-001.

Kodera, K. and Y. Kuroda, 2002: Dynamical response to the solar cycle. *J. Geophys. Res*., **107**, 4749, doi: 10.1029/2002JD002224.

Liess, S. and M.A. Geller, 2012: On the relationship between QBO and distribution of tropical deep convection. *J. Geophys. Res*., **117**, D03108, doi: 10.1029/2011JD016317.

Lindzen, R.S., and J.R. Holton, 1968:. A Theory of the Quasi-Biennial Oscillation. *J. Atmos. Sci.*, **25**, 1095 - 1107. doi: 10.1175/1520-0469(1968)025%3C1095:ATOTQB%3E2.0.CO;2

Long, C.S., *et al.*, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14 593 - 14 629, doi: 10.5194/acp-17-14593-2017.

Lott, F., *et al.*, 2014: Kelvin and Rossby-gravity wave packets in the lower stratosphere of some high-top CMIP5 models. *J. Geophys. Res. Atmos.*, **119**, 2156 - 2173, doi: 10.1002/2013JD020797.

Marshall, A.G. and A.A. Scaife, 2009: Impact of the QBO on surface winter climate. *J. Geophys. Res.*, **114**, D18110, doi: 10.1029/2009JD011737.

Marshall, A.G., H.H. Hendon, and G. Wang, 2016: On the role of anomalous ocean surface temperatures for promoting the record Madden-Julian Oscillation in March 2015. Geophys. Res. Lett., **43**, 472 - 481, doi: 10.1002/2015GL066984.

Martineau, P., 2017: S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels. *Centre for Environmental Data Analysis*, doi: 10.5285/b241a7f536a244749662360bd7839312.

Maruyama, T., 1991: Annual and QBO-synchronized variations of lower stratospheric equatorial wave activity over Singapore during 1961 - 1989. *J. Meteor. Soc. Japan*, *69*, 219 - 231, doi: 10.2151/jmsj1965.69.2_219.

Matsuno, T., 1966: Quasi-Geostrophic Motions in the Equatorial Area. *J. Meteor. Soc. Japan*. *Ser. II*, **44**, 25 - 43, doi: 10.2151/jmsj1965.44.1 25, 1966.

Matthes, K., *et al.*, 2004.: Improved 11-year solar signal in the Freie Universitat Berlin Climate Middle Atmosphere Model (FUB-CMAM). *J. Geophys. Res.*, **109**, D06101, doi: 10.1029/2003JD004012.

Mitchell, D.M., *et al.*, 2015: Solar signals in CMIP-5 simulations: the stratospheric pathway. *Quart. J. Roy. Meteor. Soc.*, **141**, 2390 - 2403, doi: 10.1002/qj.2530.

Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geosc. Mod. Dev.*, **8**, 1339 - 1356, doi: 10.5194/gmd-8-1339-2015.

Naujokat, B., 1986: An update of the observed quasi-biennial oscillation of the stratospheric winds over the tropics. *J. Atmos. Sci.,* **43**, 1873 - 1877, doi: 10.1175/1520-0469(1986)043<1873:AUOTOQ>2.0.CO;2.

Newman, P.A., L. Coy, S. Pawson, and L.R. Lait, 2016: The anomalous change in the QBO in 2015–2016. *Geophys. Res. Lett.*, **43**, 8791 - 8797, doi: 10.1002/2016GL070373.

Nie, J. and A.H. Sobel, 2015: Responses of Tropical Deep Convection to the QBO: Cloud-Resolving Simulations. *J. Atmos. Sci.*, **72**, 3625 - 3638, doi: 10.1175/JAS-D-15-0035.1.

Nishimoto, E. and S. Yoden, 2017: Influence of the Stratospheric Quasi-Biennial Oscillation on the Madden–Julian Oscillation during Austral Summer. *J. Atmos. Sci.*, **74**, 1105 - 1125, doi: 10.1175/JAS-D-16-0205.1.

Osprey, S.M., *et al.*, 2016: An unexpected disruption of the atmospheric quasi-biennial oscillation. *Science*, **353**, 1424 - 1427, doi: 10.1126/science.aah4156.

Pawson, S. and M. Fiorino, 1998: A comparison of reanalyses in the tropical stratosphere. Part 2: the quasi-biennial oscillation. *Clim. Dyn.*, **14**, 645 - 658, doi: 10.1007/s00382005024.

Pfeifroth, U., R. Mueller, and B. Ahrens, 2013: Evaluation of Satellite-Based and Reanalysis Precipitation Data in the Tropical Pacific. *J. Appl. Meteor. Clim.*, **52**, 634 - 644, doi: 10.1175/JAMC-D-12-049.1.

Plumb, R.A. and R.C. Bell, 1982: A model of the quasi-biennial oscillation on an equatorial beta-plane. *Quart. J. Roy. Meteor. Soc.*, **108**, 335 - 352, doi: 10.1002/qj.49710845604.

Polavarapu, S., T.G. Shepherd, Y. Rochon, and S. Ren, 2005: Some challenges of middle atmosphere data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3513 - 3527, doi: 10.1256/qj.05.87.

Politowicz, P.A. and M.H. Hitchman, 1997: Exploring the effects of forcing quasi-biennial oscillations in a two-dimensional model. *J. Geophys. Res. Atmos.*, **102**, 16 481 - 16 497, doi: 10.1029/97JD00693.

Randel, W., *et al.*, 2004: The SPARC intercomparison of middle atmosphere climatologies. *J. Clim.*, **17**, 986 - 1003, doi: 10.1175/1520-0442(2004)017<0986:TSIOMC>2.0.CO;2.

Rayner, N.A., *et al.*, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.*, **108**, doi: 10.1029/2002JD002670.

Richter, J.H., A. Solomon, and J. Bacmeister, 2014: On the simulation of the quasi-biennial oscillation in the Community Atmosphere Model. *J. Geophys. Res.*, **119**, 3045 - 3062, doi: 10.1002/2013JD021122.

Robock, A., 2000: Volcanic eruptions and climate. *Rev. Geophys.*, **38**, 191 - 219, doi: 10.1029/1998RG000054.

Saha, S., *et al.*, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Met. Soc.*, **91**, 1015 - 1058, doi: 10.1175/2010BAMS3001.1.

Sato, M., J.E. Hansen, M.P. McCormick, and J.B. Pollack, 1993: Stratospheric aerosol optical depths, 1850 - 1990. *J. Geophys. Res. Atmos.*, **98**, 22 987 - 22 994, doi: 10.1029/93JD02553.

Schenzinger, V., S. Osprey, L. Gray, and N. Butchart, 2017: Defining metrics of the Quasi-Biennial Oscillation in global climate models. *Geosc. Mod. Dev.*, **10**, 2157 - 2168, doi: 10.5194/gmd-10-2157-2017.

Schneider, U., *et al.*, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Clim.*, **115**, 15 - 40, doi: 10.1007/s00704-013-0860-x.

Shibata, K. and M. Deushi, M., 2005: Radiative effect of ozone on the quasi-biennial oscillation in the equatorial stratosphere. *Geophys. Res. Lett.*, **322**, doi: 10.1029/2005GL023433.

Shindell, D.T., G.A. Schmidt, M.E. Mann, and G. Faluvegi, 2004: Dynamic winter climate response to large tropical volcanic eruptions since 1600. *J. Geophys. Res. Atmos.*, **109**, doi: 10.1029/2003JD004151.

Son, S.-W., *et al.*, 2017: Stratospheric Control of the Madden–Julian Oscillation. *J. Clim.*, **30**, 1909 - 1922, doi: 10.1175/JCLI-D-16-0620.1.

Stenchikov, G., *et al.*, 2006: Arctic Oscillation response to volcanic eruptions in the IPCC AR4 climate models. *J. Geophys. Res. Atmos.*, **111**, D07107, doi: 10.1029/2005JD006286.

Tegtmeier, S., *et al.*, 2020: Zonal Asymmetry of the QBO Temperature Signal in the Tropical Tropopause Region. *Geophys. Res. Lett.*, **47**, e2020GL089533, doi: 10.1029/2020GL089533.

Wang, Y.-M., J.L. Lean, N.R. Sheeley, Jr., 2005.: Modeling the Sun's Magnetic Field and Irradiance since 1713. *Astrophys. J.*, **625**, 522, doi: 10.1086/429689.

Wright, C.J. and N.P. Hindley, 2018: How well do stratospheric reanalyses reproduce highresolution satellite temperature measurements? *Atmos. Chem. Phys.*, **18**, 13703 - 13731, doi: 10.5194/acp-18-13703-2018.

Yang, G.-Y., B.J. Hoskins, and J.M. Slingo, 2011: Equatorial Waves in Opposite QBO Phases. *J. Atmos. Sci.*, **68**, 839 - 862, doi: 10.1175/2010JAS3514.1.

Yang, G.-Y., B. Hoskins, and L. Gray, 2012: The Influence of the QBO on the Propagation of Equatorial Waves into the Stratosphere. *J. Atmos. Sci.*, **69**, 2959 - 2982, doi: 10.1175/JAS-D-11-0342.1.

Yoo, C. and S.-W. Son, 2016: Modulation of the boreal wintertime Madden-Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophys. Res. Lett.*, **43**, 1392 - 1398, doi: 10.1002/2016GL067762.

Appendix A: Additional Figures

A9.1 Supplemental

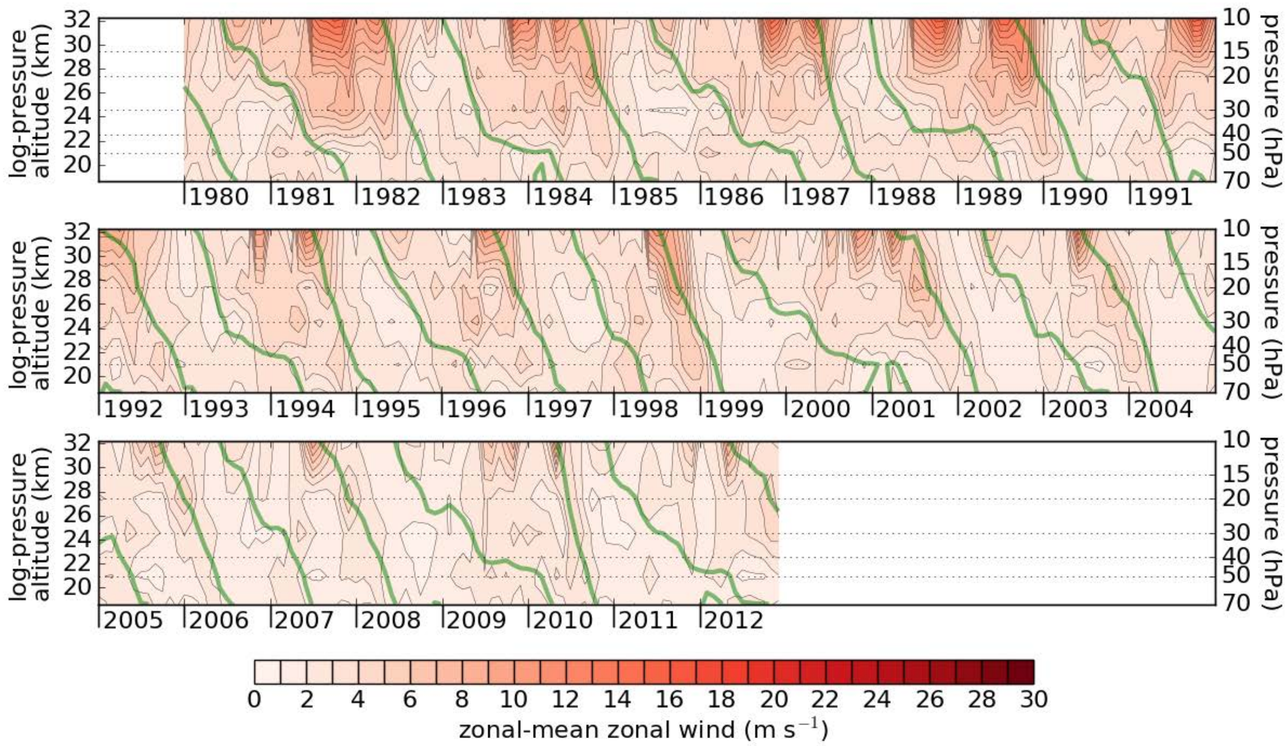


Figure AS9.1: *Inter-reanalysis standard deviation (SD) of monthly-mean 2°S-2°N zonal-mean zonal wind (as in* ***Figure 9.4****) for the common year range 1980-2012 of nine reanalyes: ERA-Interim, MERRA, MERRA-2, JRA-25, JRA-55, JRA-55C, CFSR, NCEP-NCAR, NCEP-DOE. ERA-40 is excluded because it ends in 2002. Thick green contours show the zero-wind line of the REM for these 9 reanalyses.*

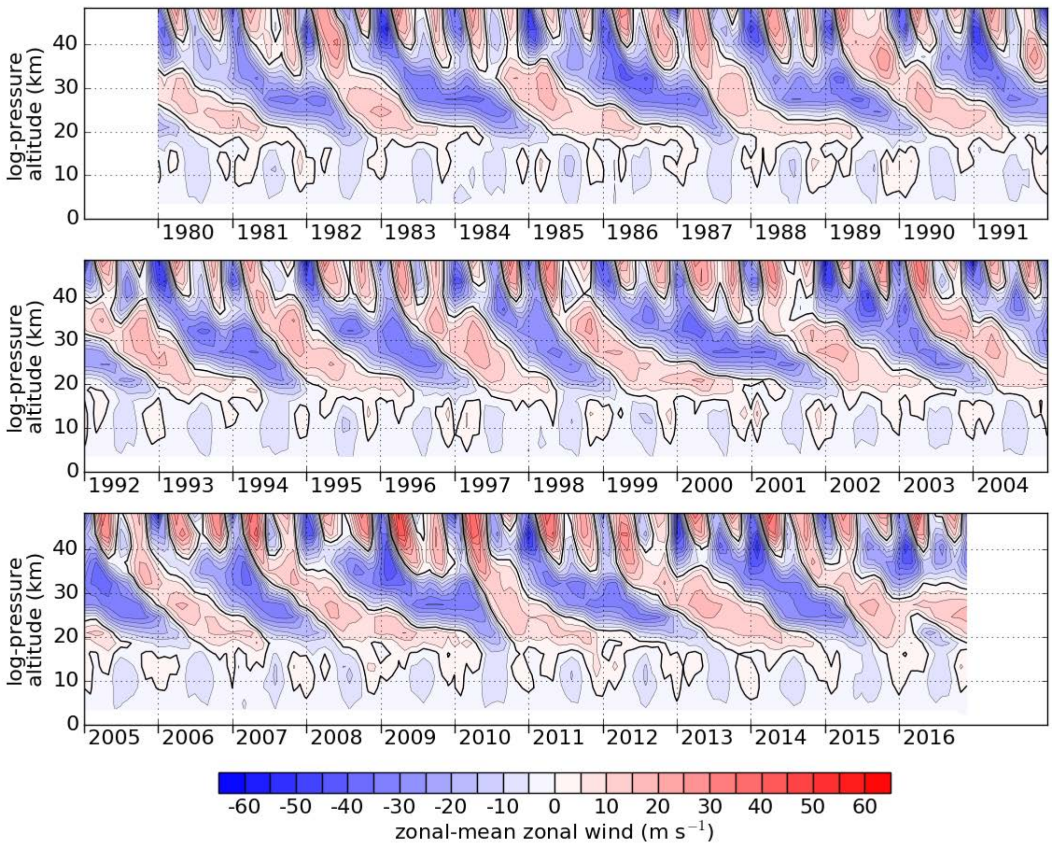


Figure AS9.2: *REM of monthly-mean 2°S-2°N zonal-mean zonal wind for ERA-Interim, JRA-55, MERRA-2 and CFSR as in* ***Figure 9.3****, but showing the 1000-1 hPa altitude range.*

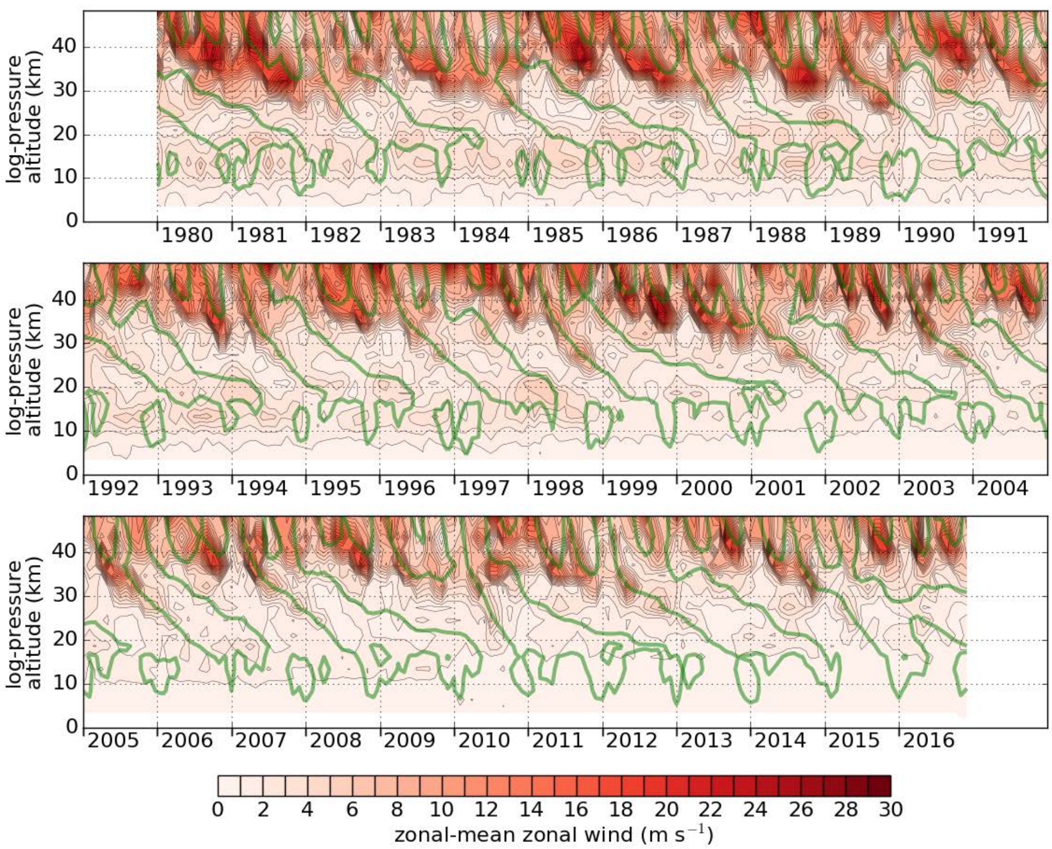


Figure AS9.3: *Inter-reanalysis SD of monthly-mean 2°S-2°N zonal-mean zonal wind for ERA-Interim, JRA-55, MERRA-2 and CFSR as in* ***Figure 9.4****, but showing the 1000-1 hPa altitude range. Thick green contours show the zero-wind line of the REM (****Figure AS9.2****).*

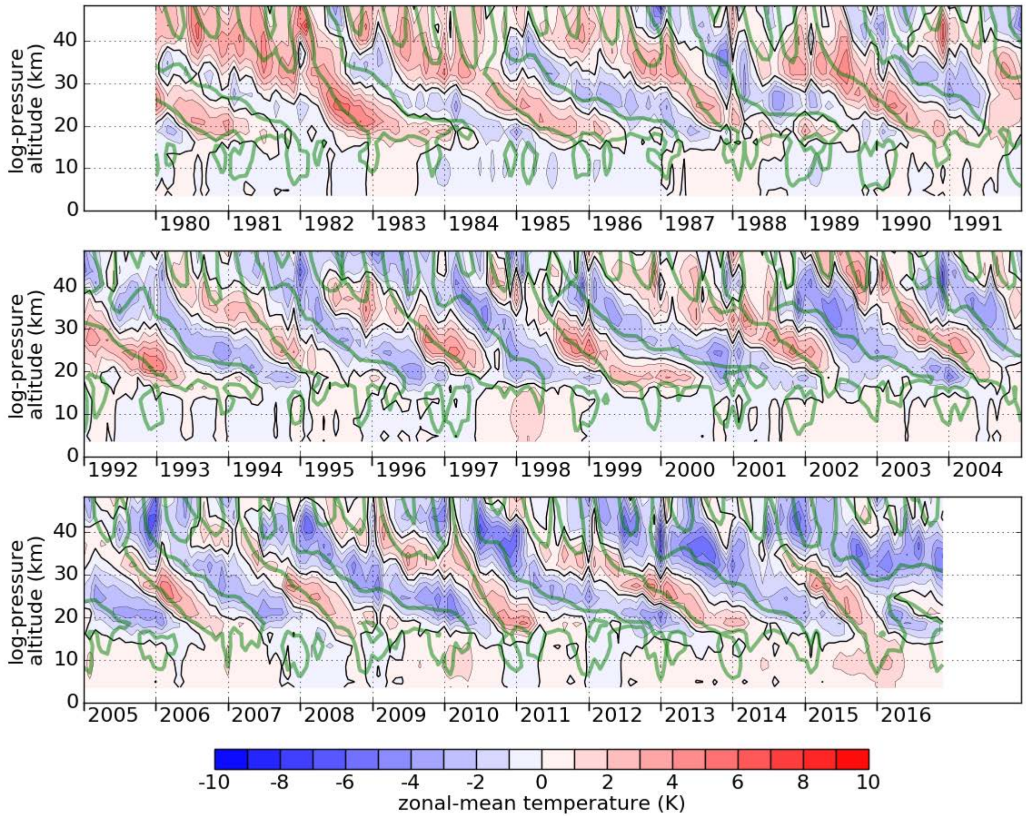


Figure AS9.4: *REM of deseasonalized monthly-mean 2°S - 2°N zonal-mean temperature for ERA-Interim, JRA-55, MERRA-2 and CFSR as in **Figure 9.6**, but showing the 1000 - 1 hPa altitude range. Thick green contours show the zero-wind line of the zonal wind REM (**Figure AS9.2**).*

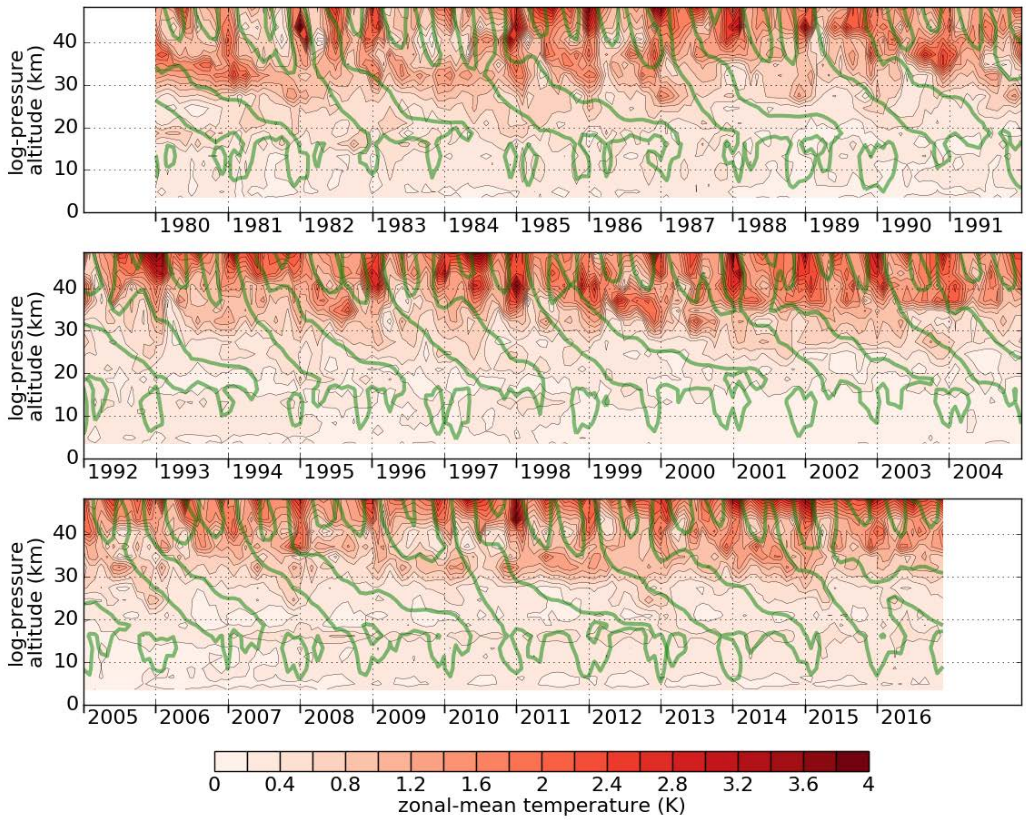


Figure AS9.5: *Inter-reanalysis SD of deseasonalized monthly-mean 2°S-2°N zonal-mean temperature for ERA-Interim, JRA-55, MERRA-2 and CFSR as in **Figure 9.4**, but showing the 1000 - 1 hPa altitude range. Thick green contours show the zero-wind line of the zonal wind REM (**Figure AS9.2**).*

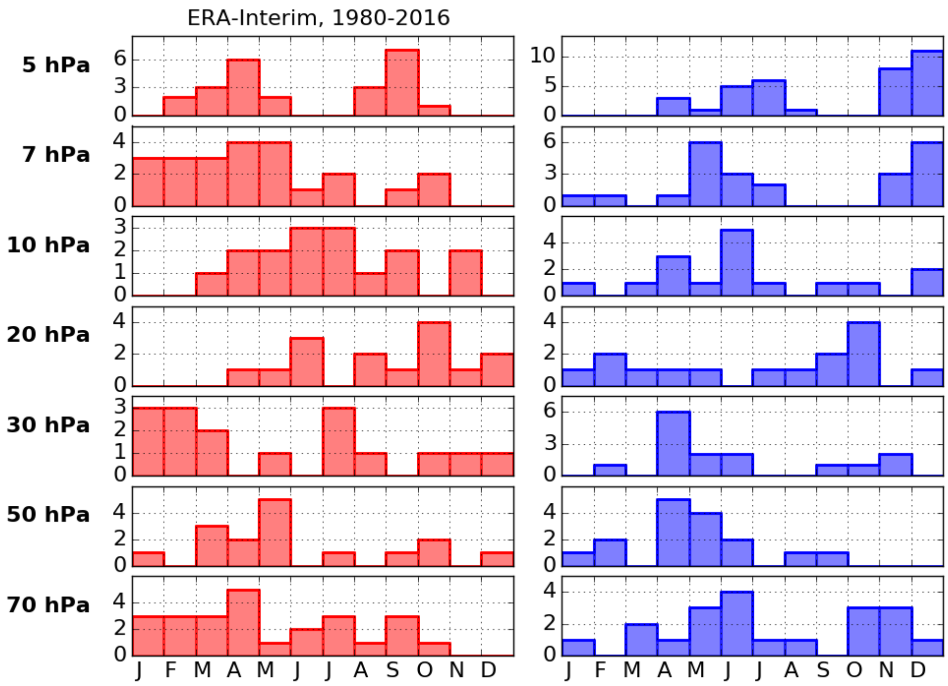


***Figure AS9.6:** Seasonal distribution of QBO phase onsets during the 1980 - 2016 period as in **Figure 9.17**, for just ERA-Interim. Timing of onsets is diagnosed from the monthly-mean 5° S - 5° N zonal-mean zonal wind, interpolated to the time of the zero crossing. Red bars (left column) indicate QBO-W onsets and blue bars (right column) indicate QBO-E onsets.*

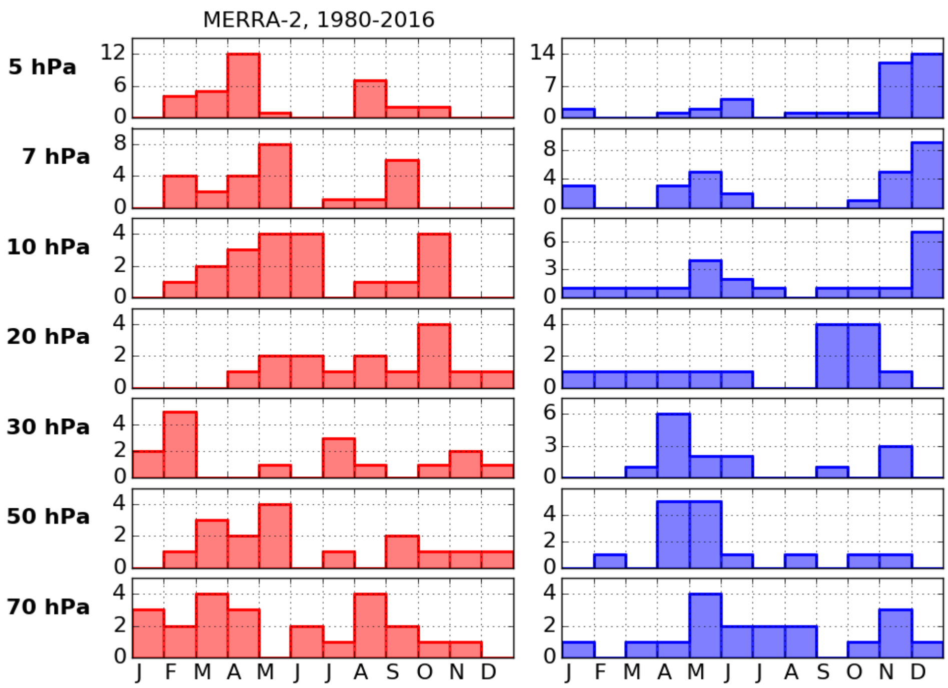


***Figure AS9.7:** As **Figure AS9.6**, but for MERRA-2.*

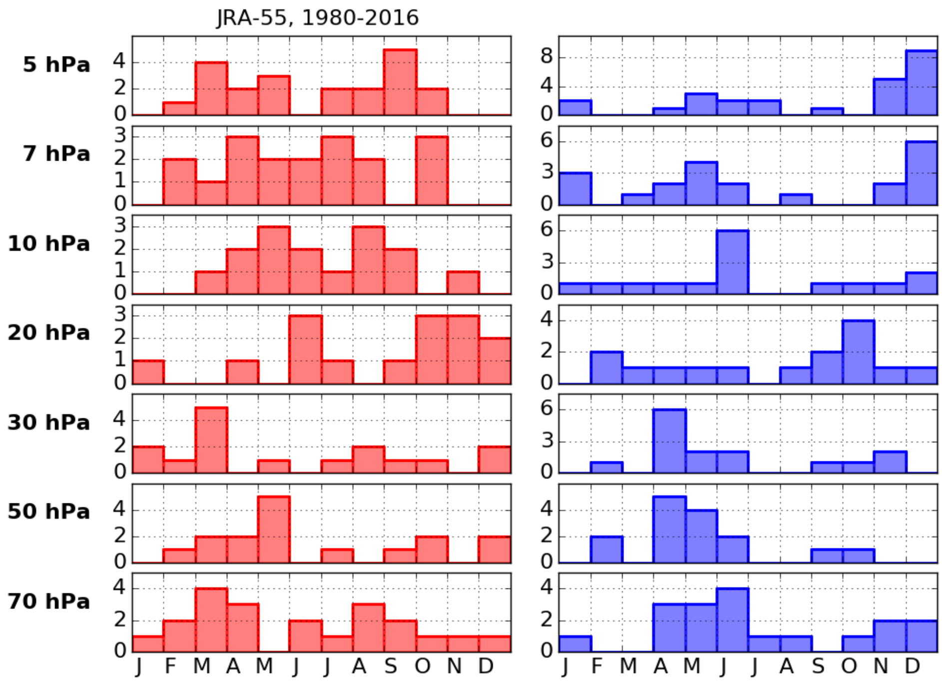


Figure AS9.8: *As **Figure AS9.6**, but for JRA-55.*

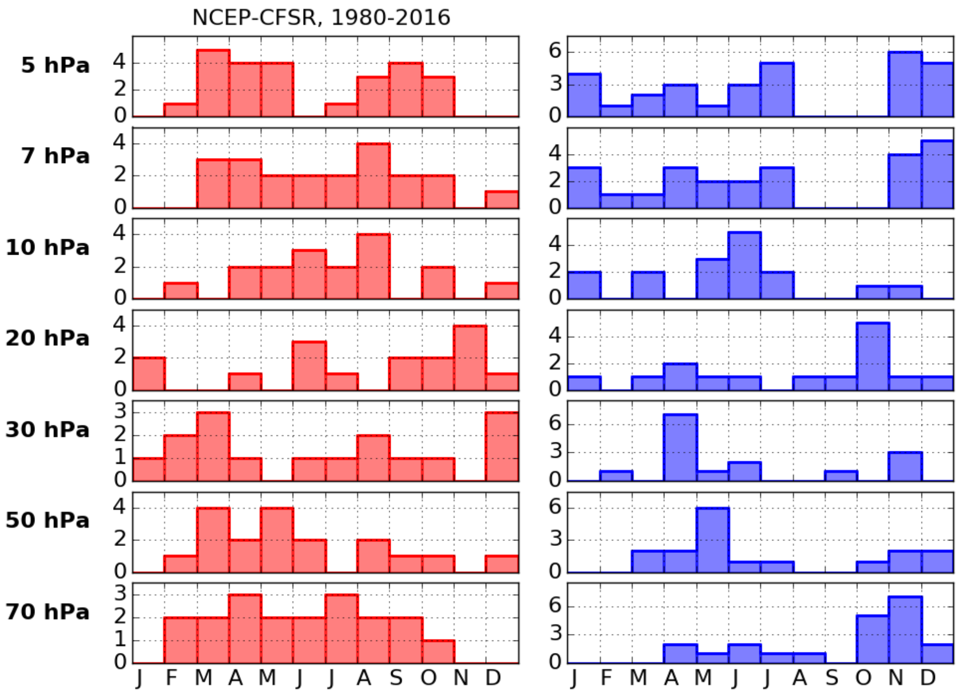


Figure AS9.9: *As **Figure AS9.6**, but for CFSR.*

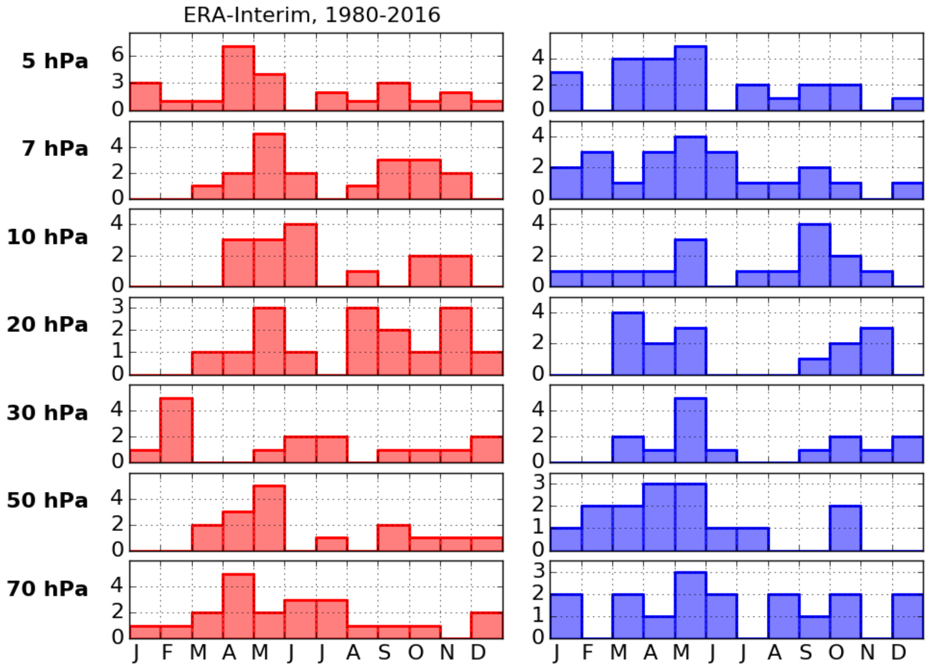


Figure AS9.10: *As Figure AS9.6, but onsets are defined using deseasonalized monthly-mean 5° S - 5° N zonal-mean zonal wind.*

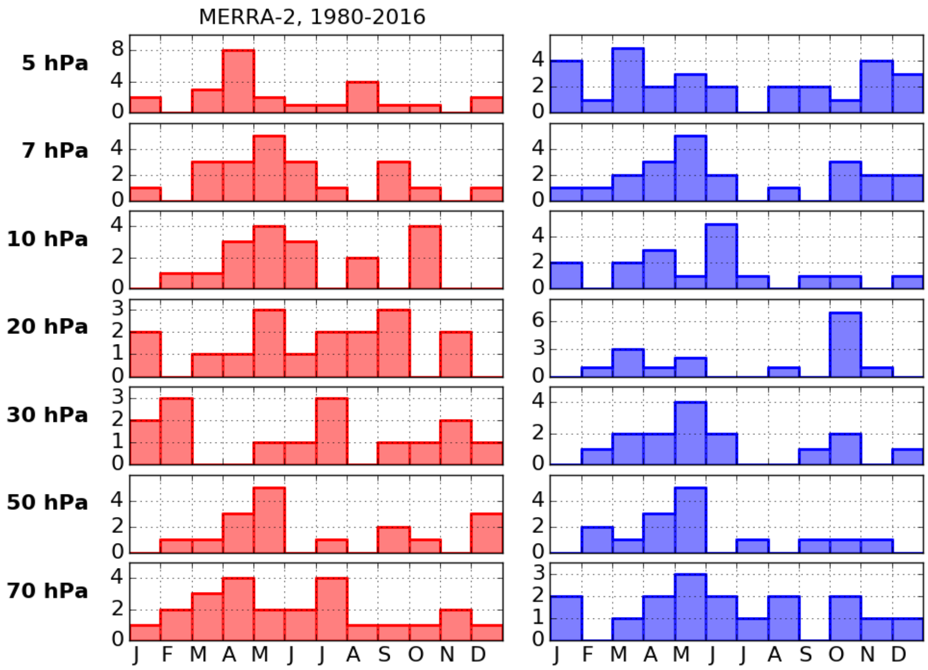


Figure AS9.11: *As* ***Figure AS9.10****, but for MERRA-2.*

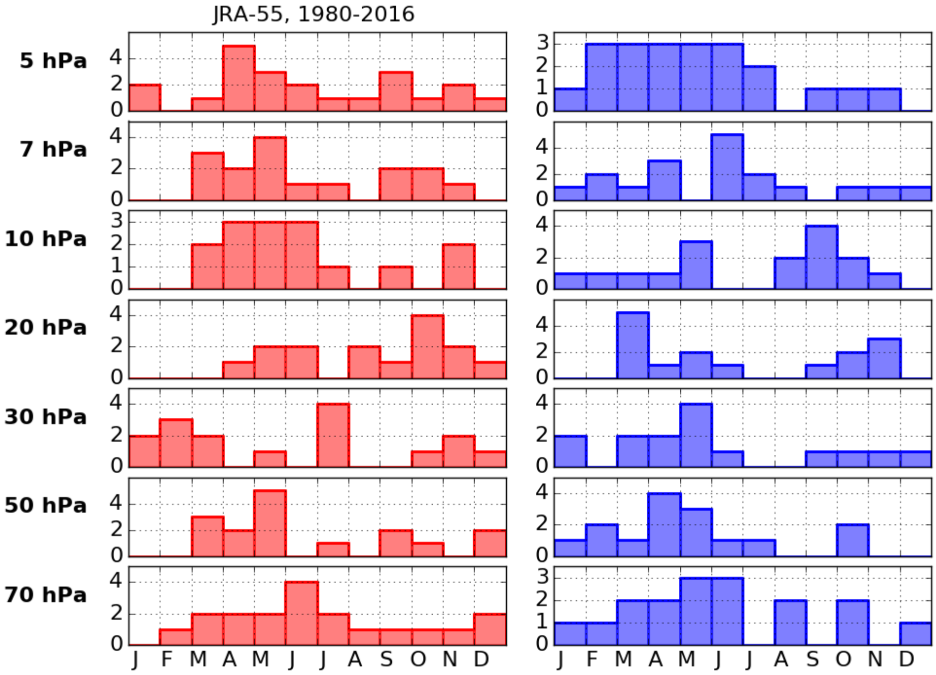


Figure AS9.12: *As **Figure AS9.10**, but for JRA-55.*

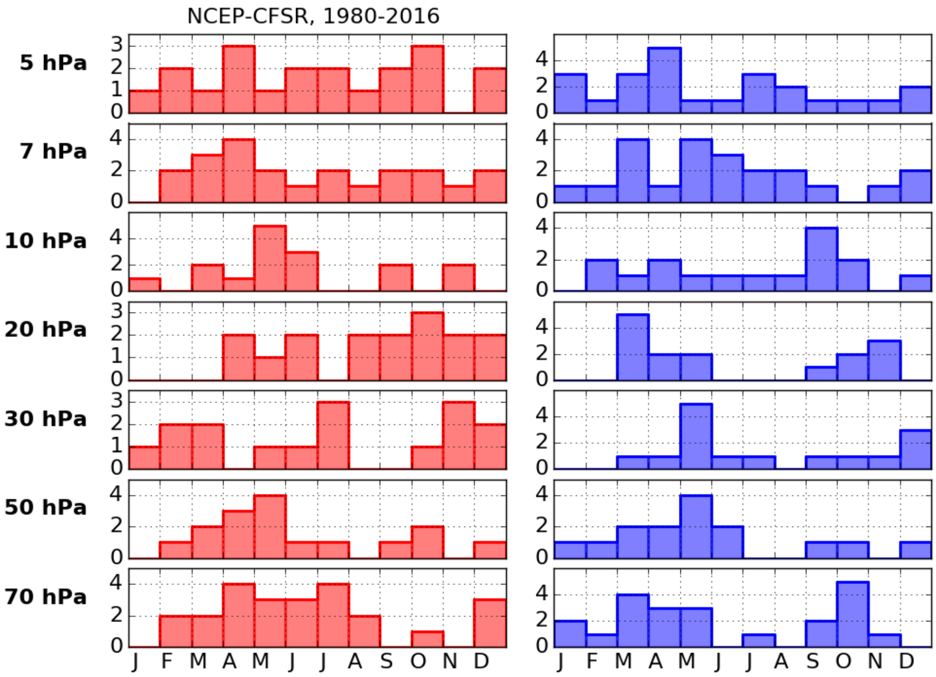


Figure AS9.13: *As **Figure AS9.10**, but for CFSR.*

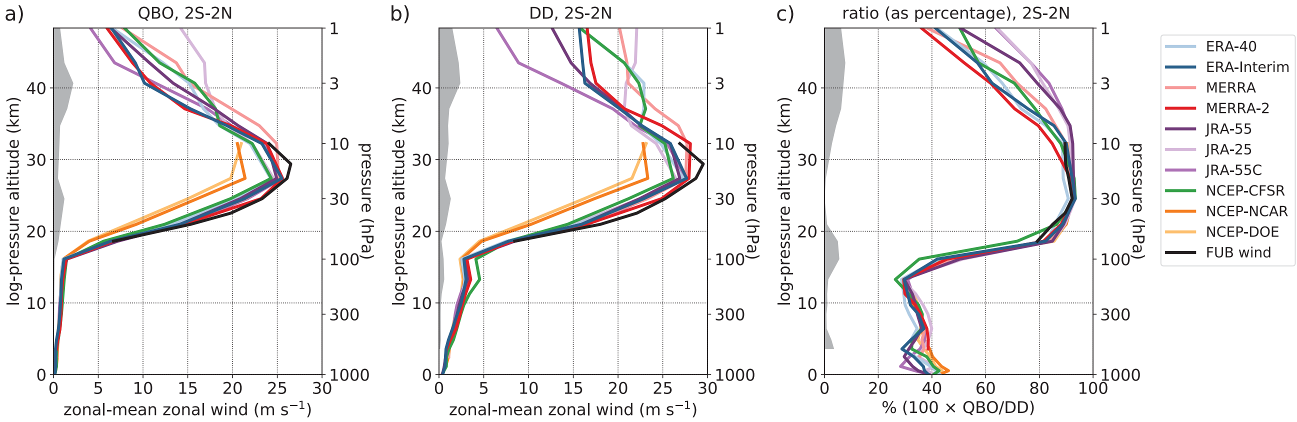


Figure AS9.14: *Comparison of QBO spectral and Dunkerton-Delisi amplitudes for monthly-mean 2° S - 2° N zonal-mean zonal wind, for reanalyses and FUB wind. The Dunkerton-Delisi amplitude (Dunkerton and Delisi, 1985) is the standard deviation of deseasonalized wind multiplied by $\sqrt{2}$. The spectral amplitude is also multiplied by $\sqrt{2}$ so that the two amplitudes are exactly comparable. (a) Spectral QBO amplitude, defined using 20 - 40 month period window (repeated from* ***Figure 9.19a****). (b) Dunkerton-Delisi amplitude. (c) Ratio of (a) to (b), given as a percentage. Grey shading indicates the inter-reanalysis standard deviation of the plotted quantity for the four most recent full-input reanalyses (ERA-Interim, MERRA-2, JRA-55 and CFSR).*

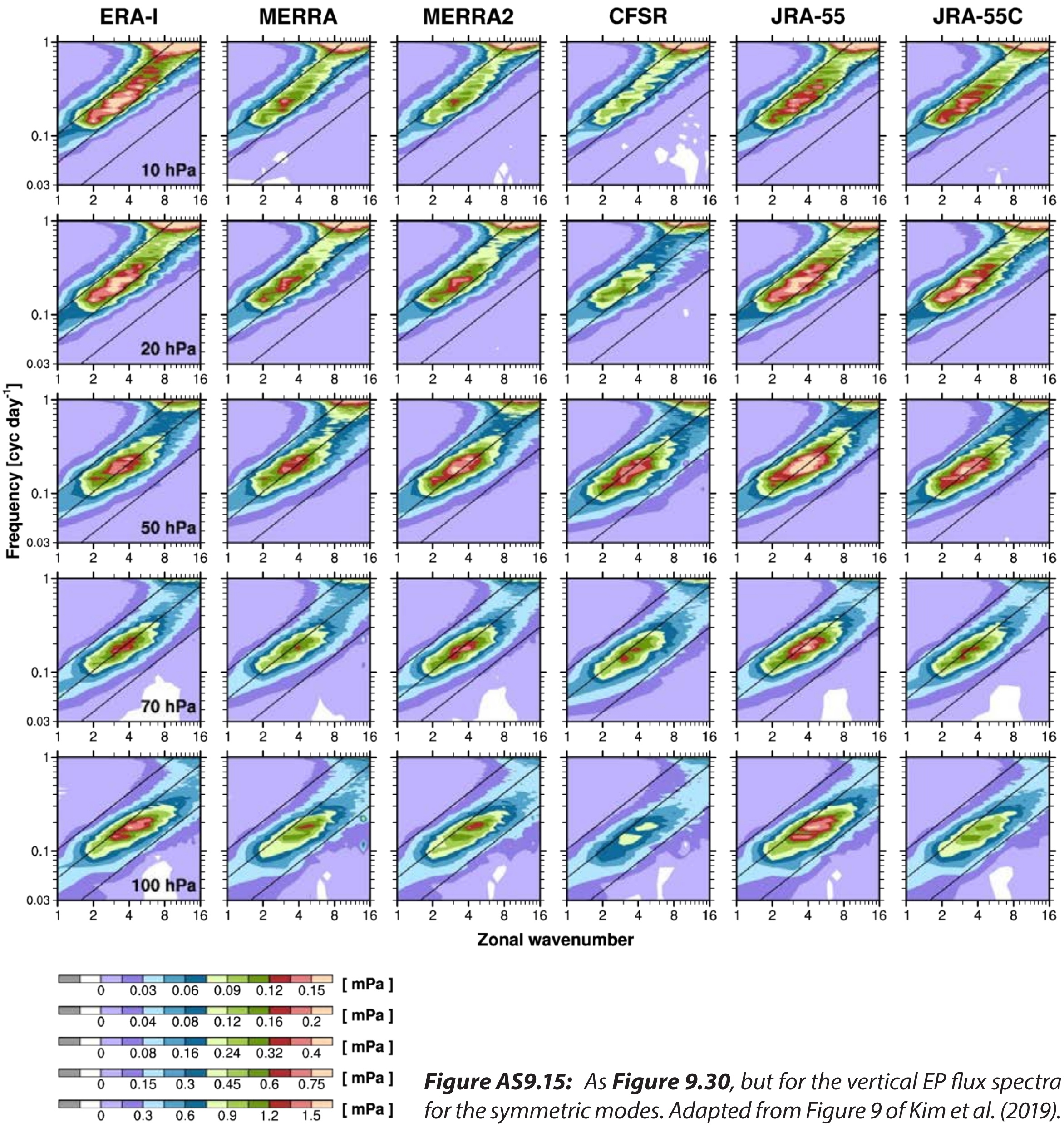


Figure AS9.15: *As* ***Figure 9.30****, but for the vertical EP flux spectra for the symmetric modes. Adapted from Figure 9 of Kim et al. (2019).*

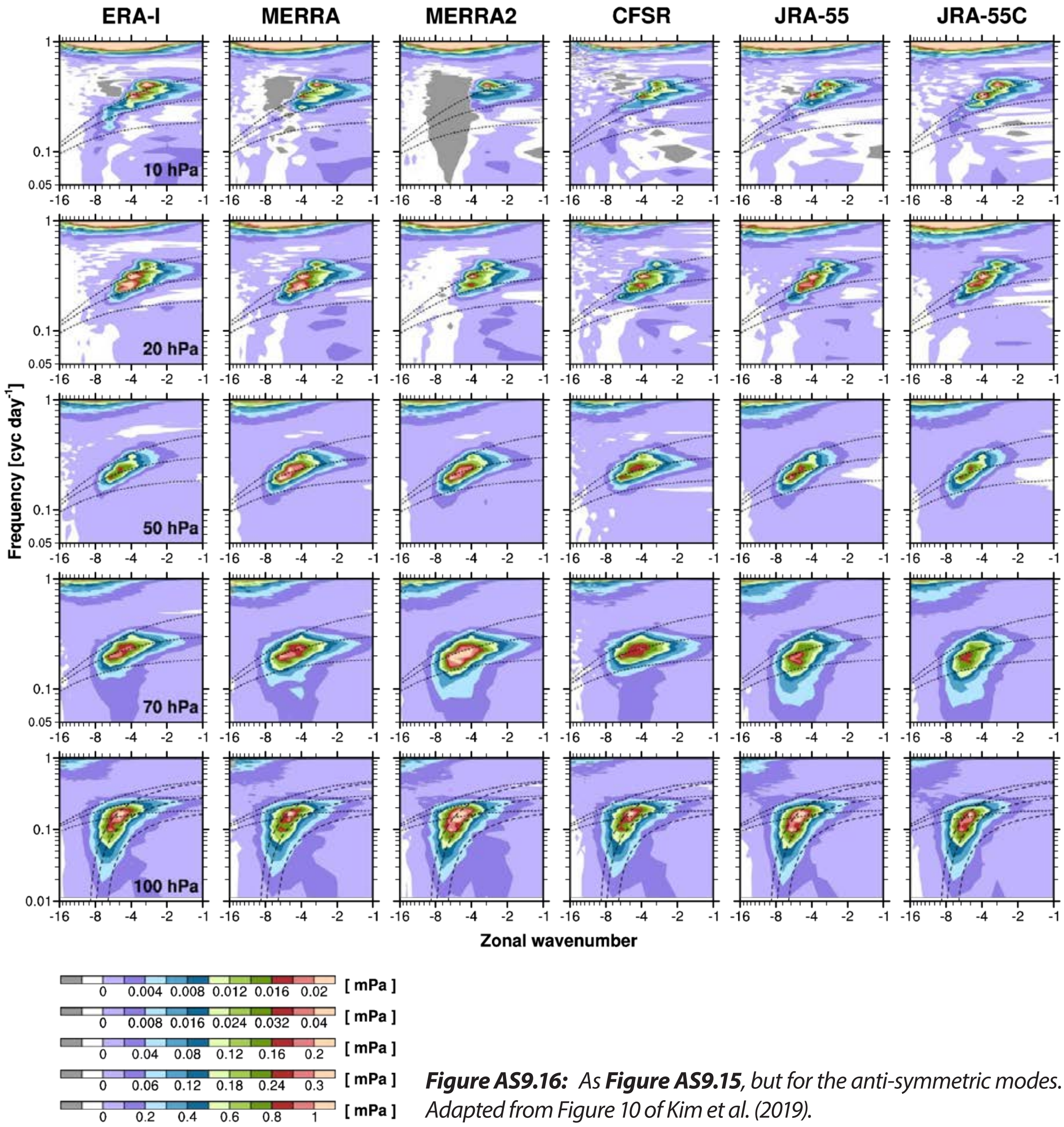


Figure AS9.16: *As **Figure AS9.15**, but for the anti-symmetric modes. Adapted from Figure 10 of Kim et al. (2019).*

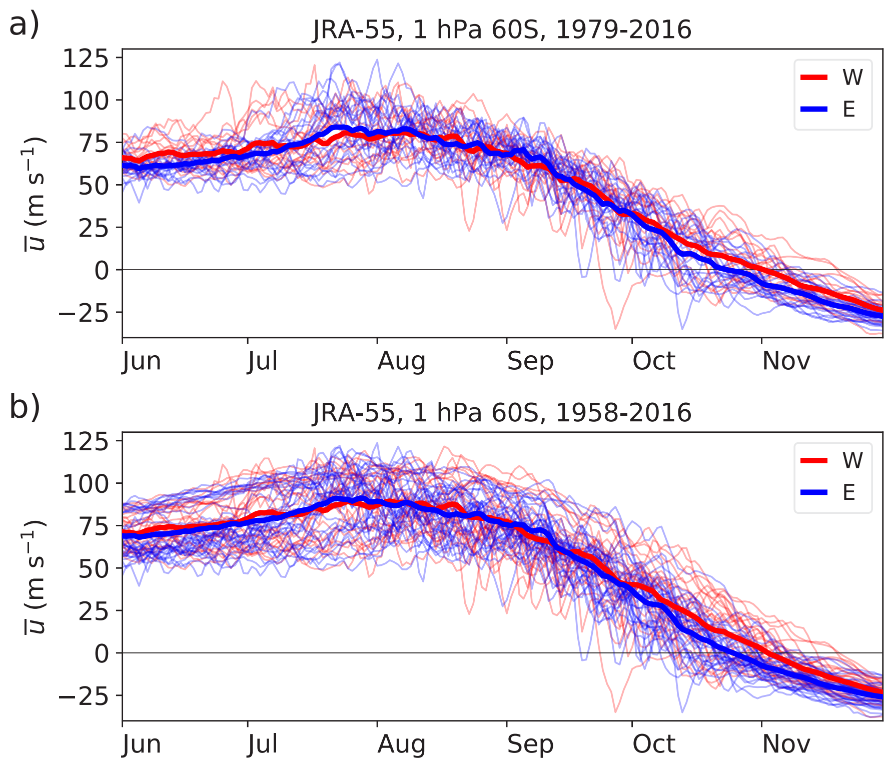


***Figure AS9.17:** As **Figure 9.44**, but for JRA-55 at 60° S, 1 hPa.*

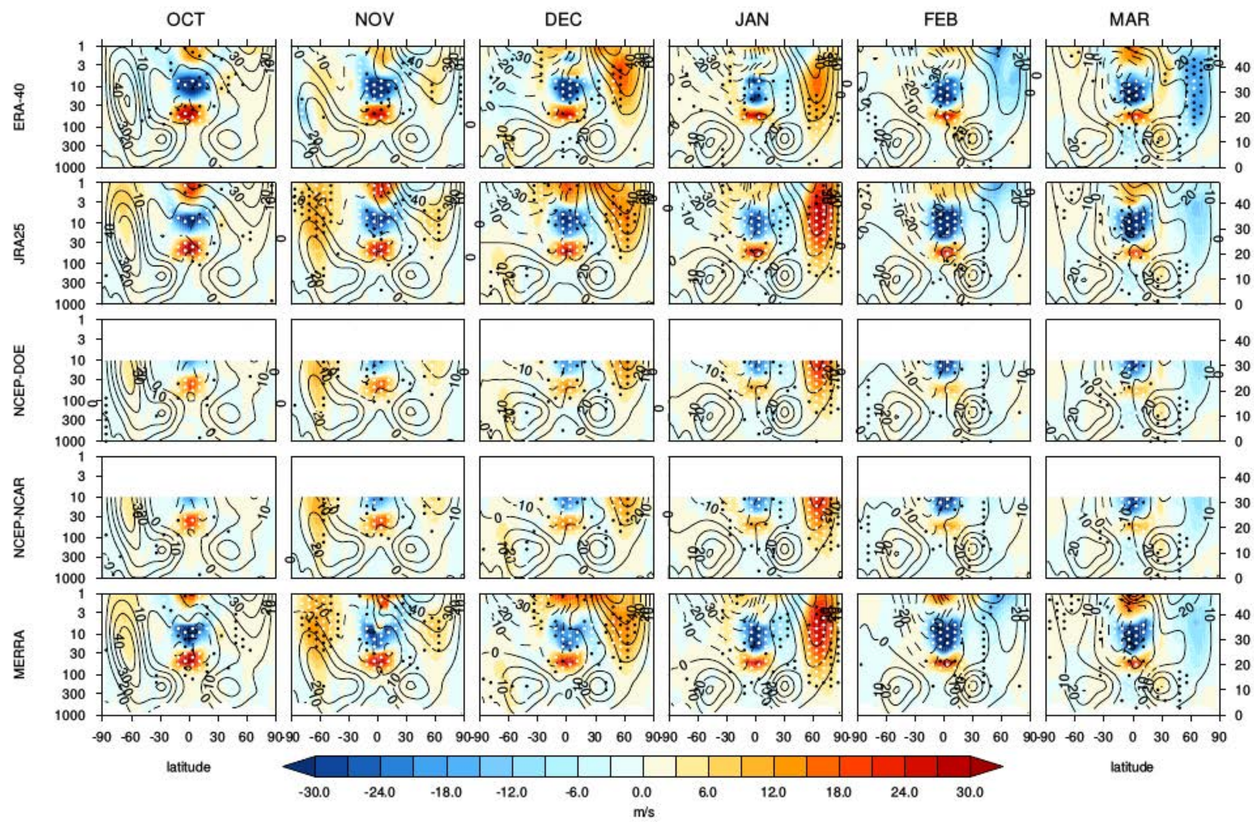


***Figure AS9.18:** As **Figure 9.47** but from the older reanalyses: ERA-40, JRA-25, NCEP-DOE, NCEP-NCAR and MERRA. The regression analysis was performed for 1980-2012, the period for which data were available from all datasets, apart from ERA-40 for which it was performed for 1980-2001.*

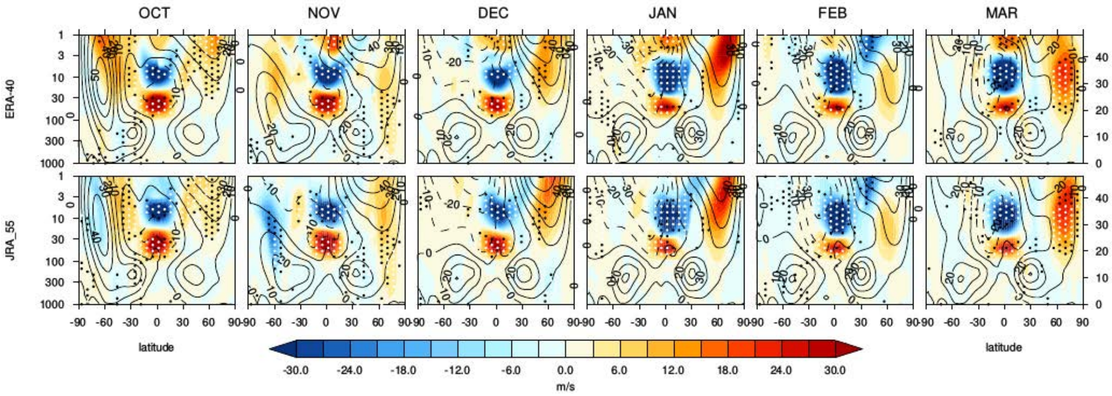


Figure AS9.19: *As **Figure 9.49** top row, but showing the separate analyses results.*

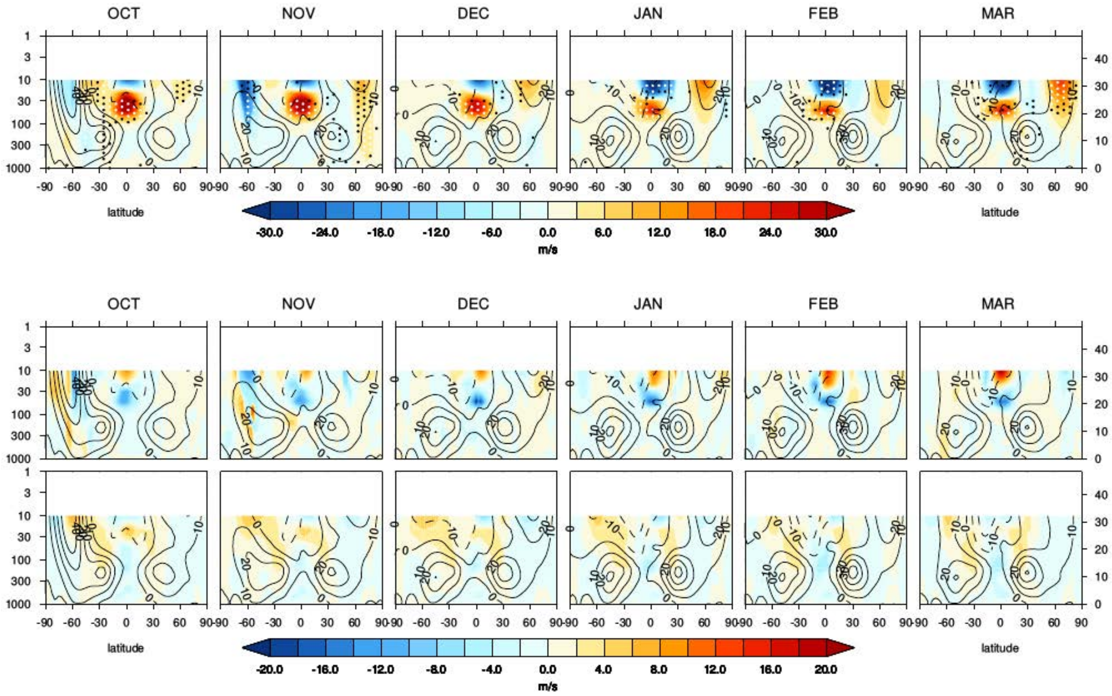


Figure AS9.20: *As **Figure 9.49** but comparing the NH winter QBO signal for NCEP-NCAR versus JRA-55 over the same period (1958 - 1979).*

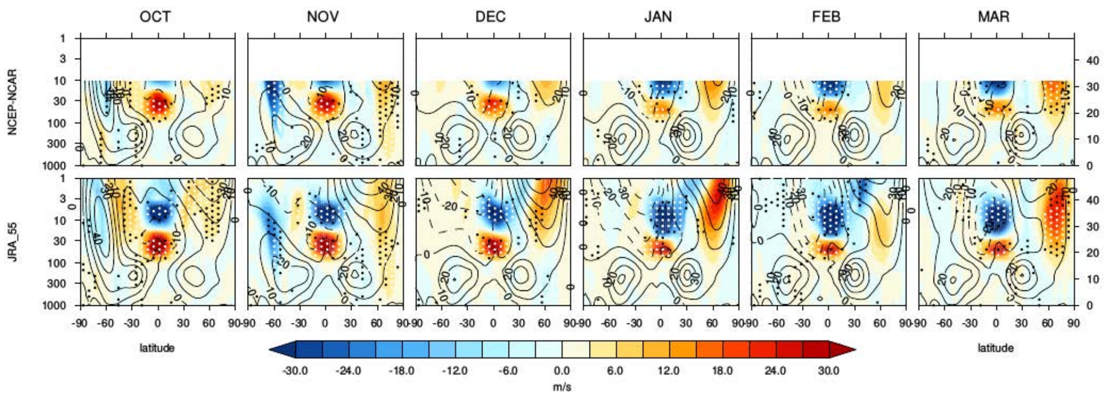


Figure AS9.21: *As **Figure AS9.20** top row, but showing the separate analyses results.*

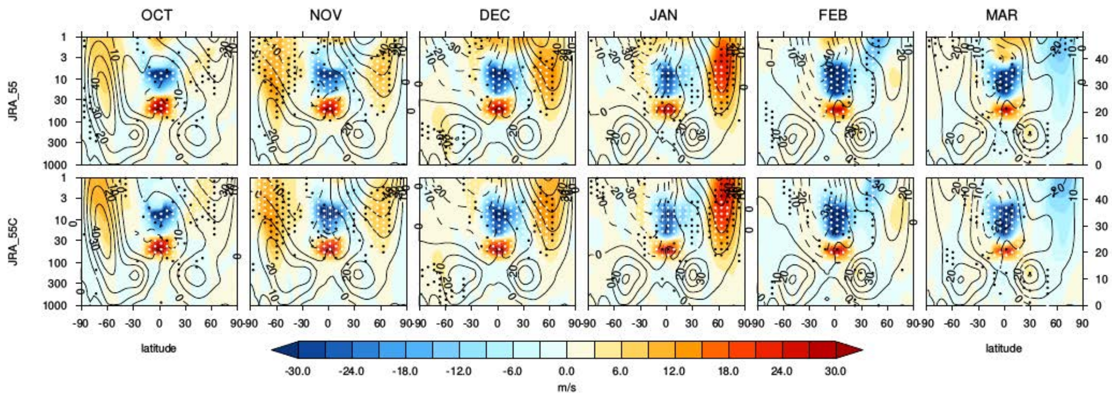


Figure AS9.22: *As **Figure 9.50** top row, but showing the QBO signal from each of the reanalyses separately.*

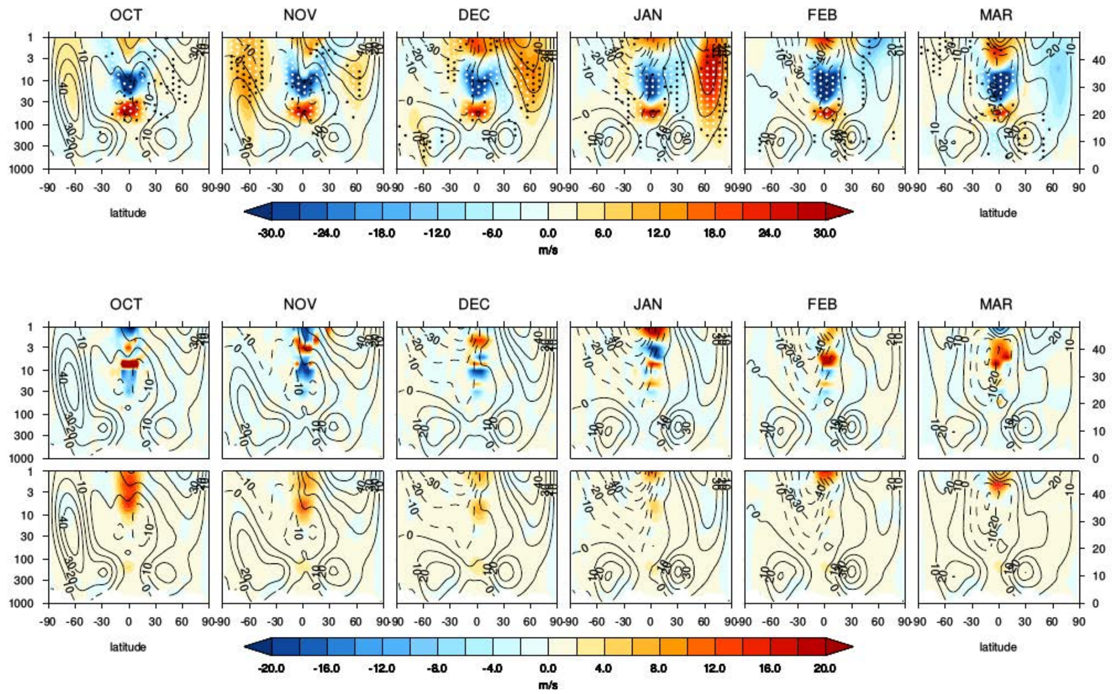


Figure AS9.23: *As **Figure 9.50** but showing the NH winter comparison between MERRA versus MERRA-2 (MERRA minus MERRA-2) for the period 1980 - 2012.*

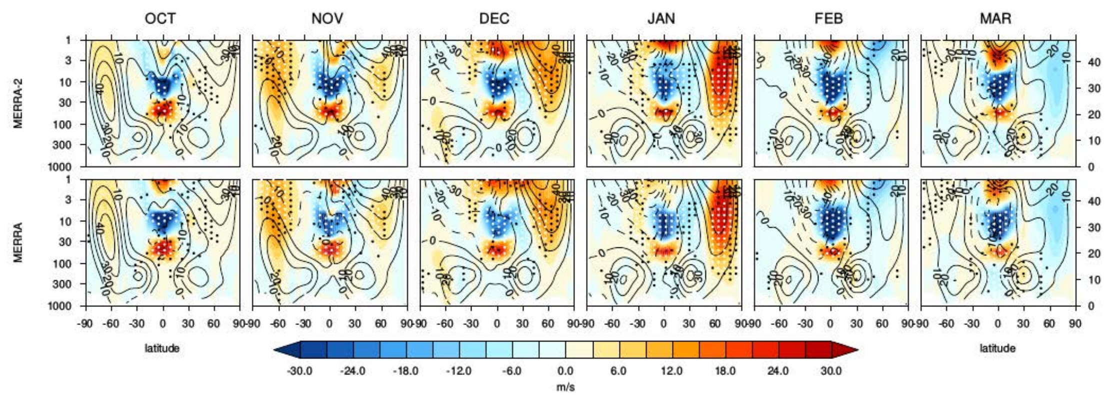


Figure AS9.24: *As **Figure AS9.23** but showing the NH winter QBO signal from each of the reanalyses separately.*

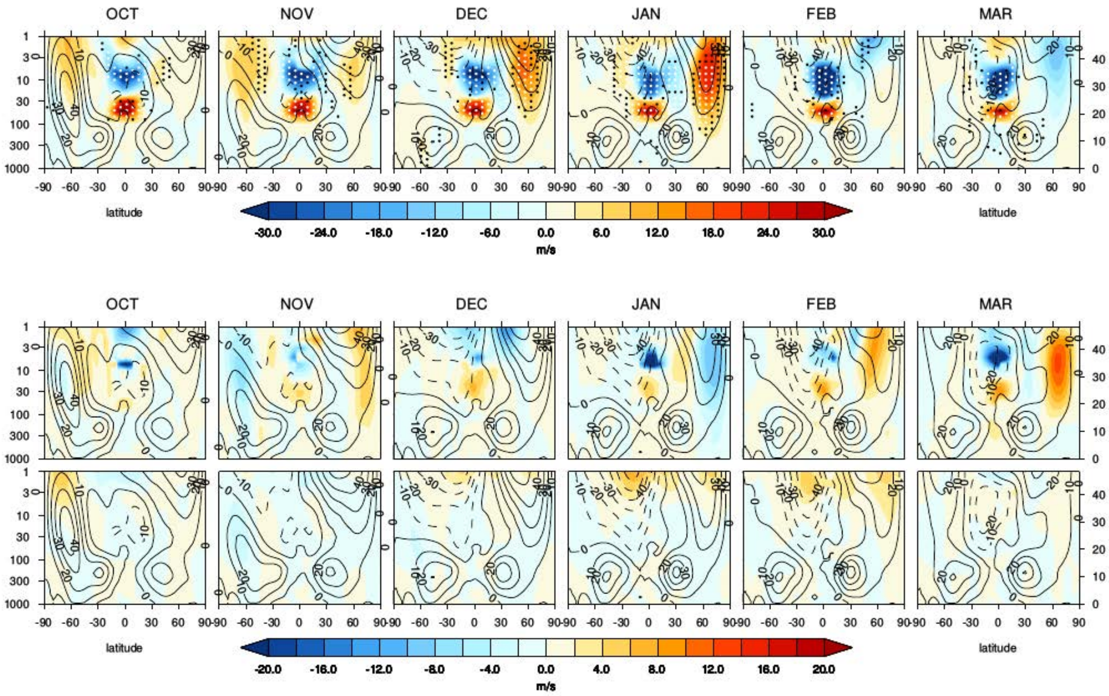


Figure AS9.25: *As* ***Figure 9.51****, but showing the QBO signal and climatology differences. Top row: the average QBO signal (m s*$^{-1}$*) from the two periods of JRA-55 (full record 1958 - 2016, and satellite era 1980 - 2016). Middle row: difference (m s*$^{-1}$*) in the two QBO signals (full minus satellite). Bottom row: difference (m s*$^{-1}$*) between the two climatological fields (full minus satellite). The corresponding average of the two climatological wind fields is overlaid on each plot (contour interval of 10 m s*$^{-1}$*).*

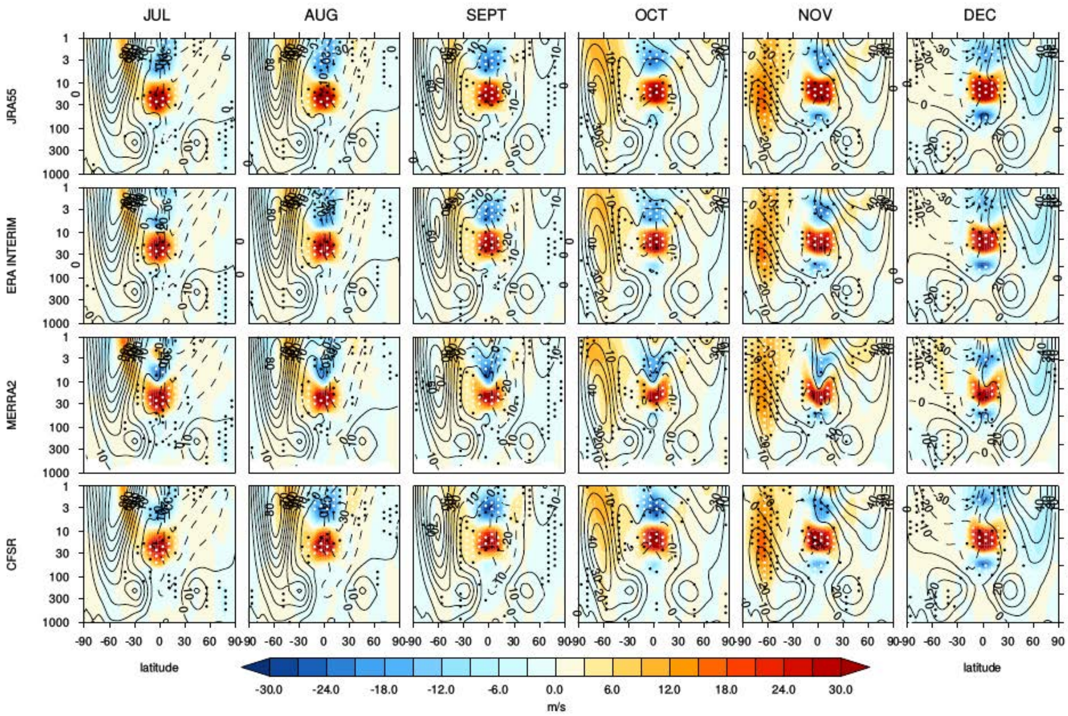


Figure AS9.26: *As* ***Figure 9.52*** *(top row), but showing the results from the individual regression analysis of the four recent reanalysis datasets (JRA-55, ERA-Interim, MERRA-2, CFSR), i.e., the individual contributions to the averaged signal shown in top row of* ***Figure 9.52****.*

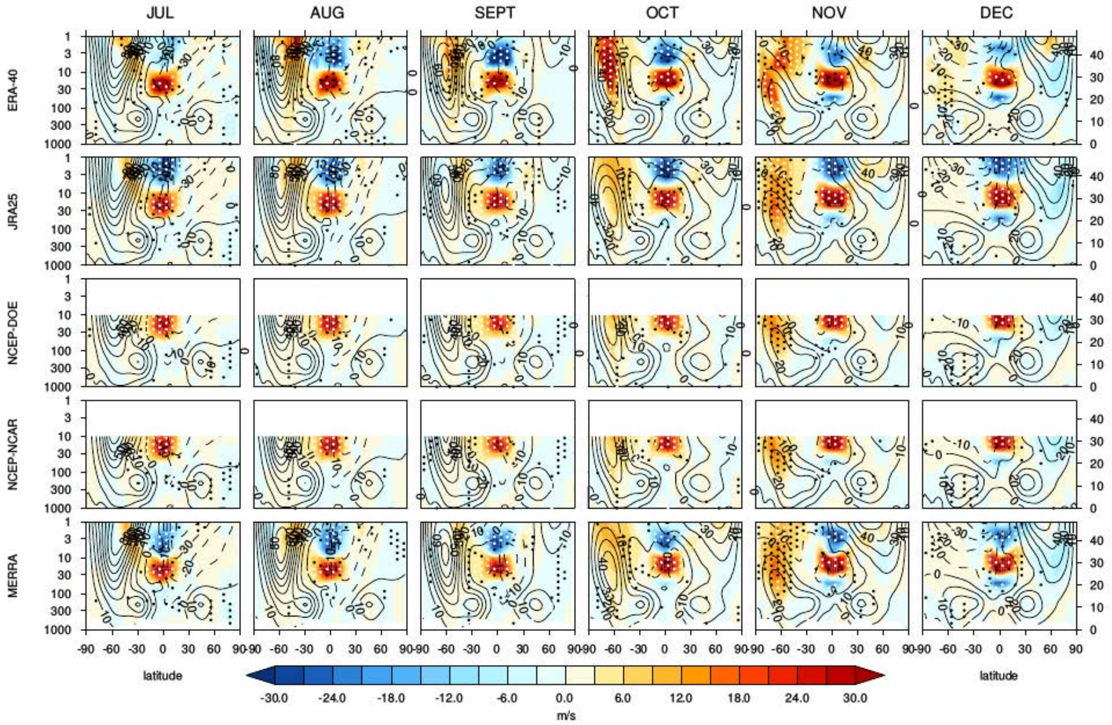


Figure AS9.27: *As **Figure AS9.26** but showing the results from the individual regression analysis of the older reanalyses: ERA-40, JRA-25, NCEP-DOE, NCEP-NCAR and MERRA. The regression analysis was performed for 1980 - 2012, the period for which data were available from all datasets, apart from ERA-40 for which it was performed for 1980 - 2001.*

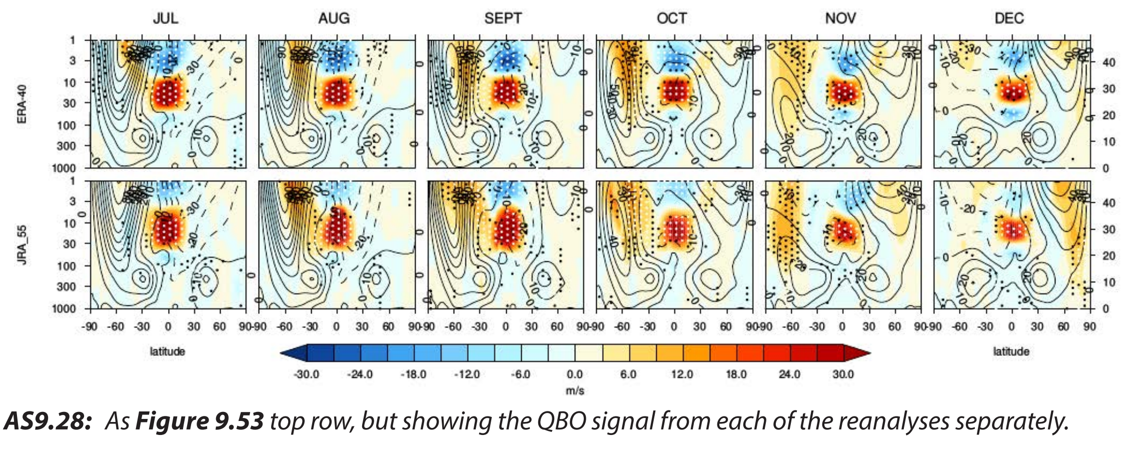


Figure AS9.28: *As **Figure 9.53** top row, but showing the QBO signal from each of the reanalyses separately.*

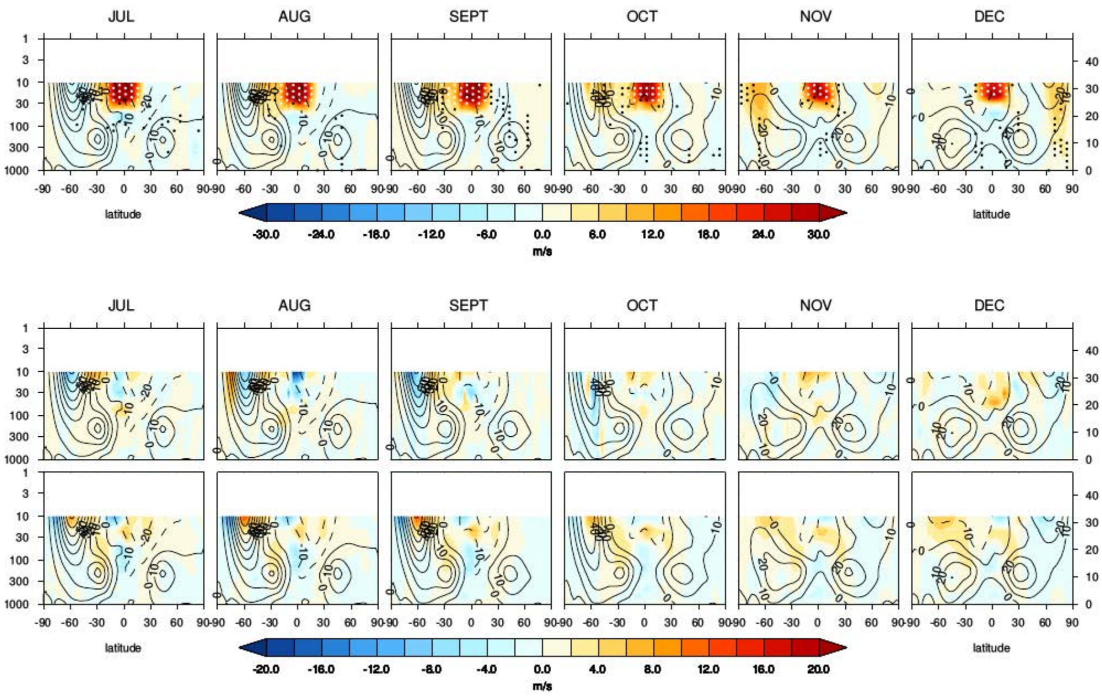


Figure AS9.29: *As* ***Figure 9.53*** *but comparing the NCEP-NCAR (R1) reanalysis dataset to JRA-55 for the same period (1958 - 1979).*

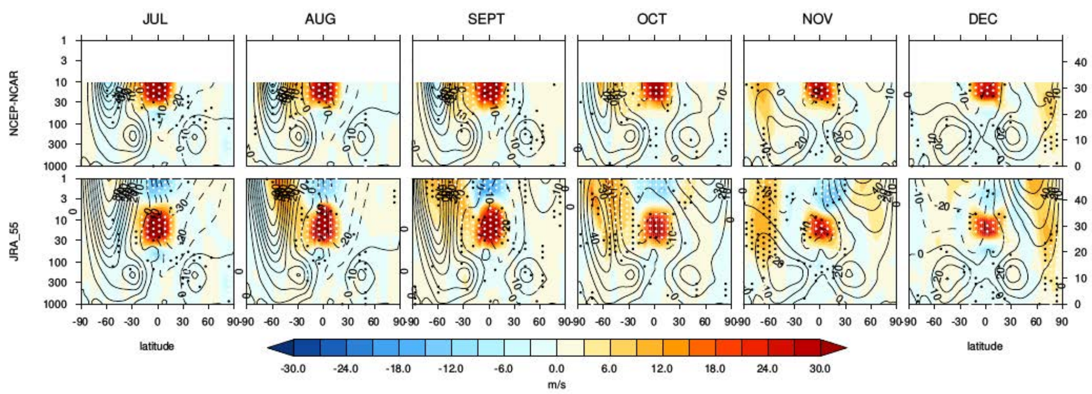


Figure AS9.30: *As* ***Figure AS9.29*** *top row, but showing the QBO signal from each of the reanalyses separately.*

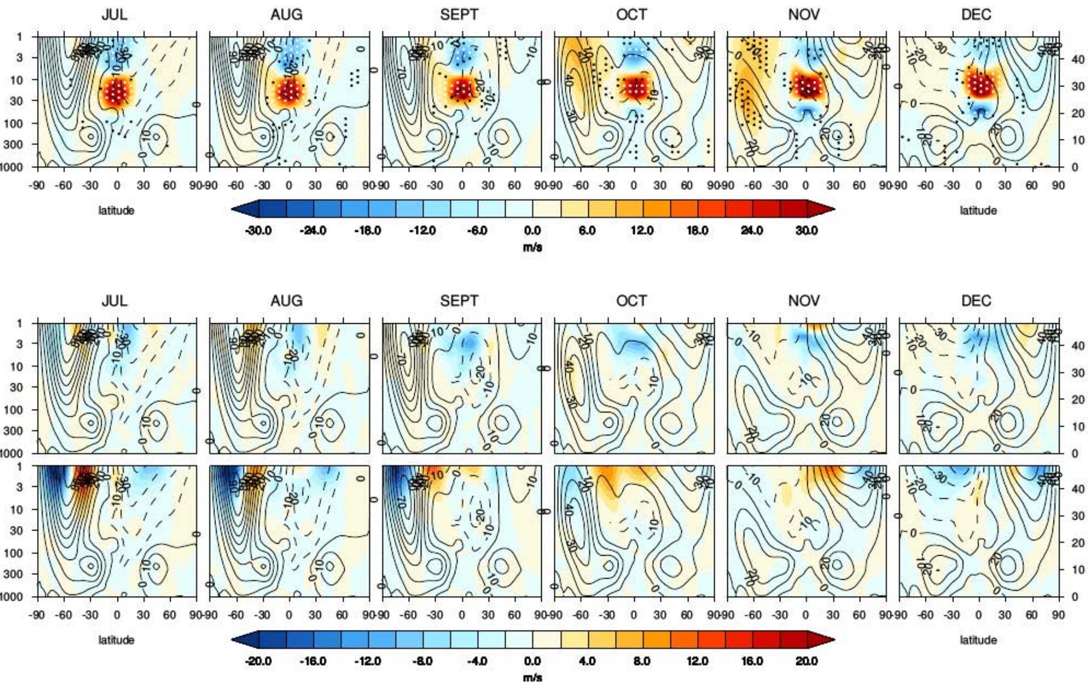


Figure AS9.31: *As **Figure 9.53** but comparing JRA-55 vs JRA-55C (1973 - 2012).*

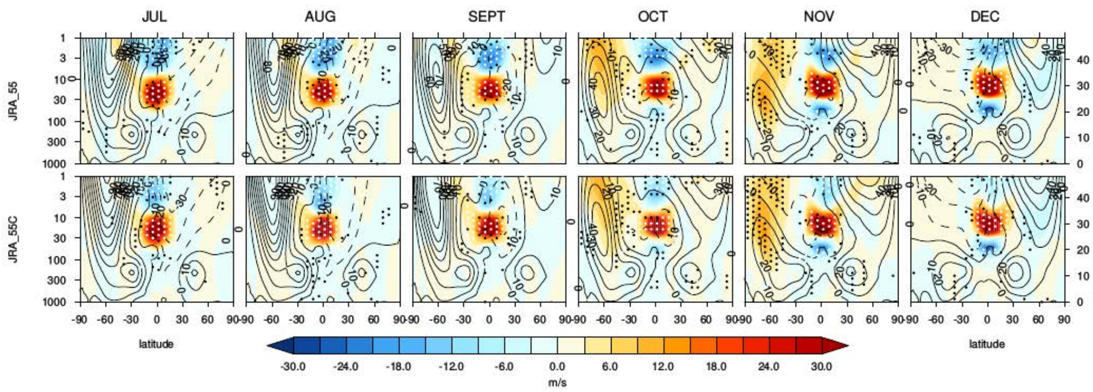


Figure AS9.32: *As **Figure AS9.31** top row, but showing the QBO signal from each of the reanalyses separately.*

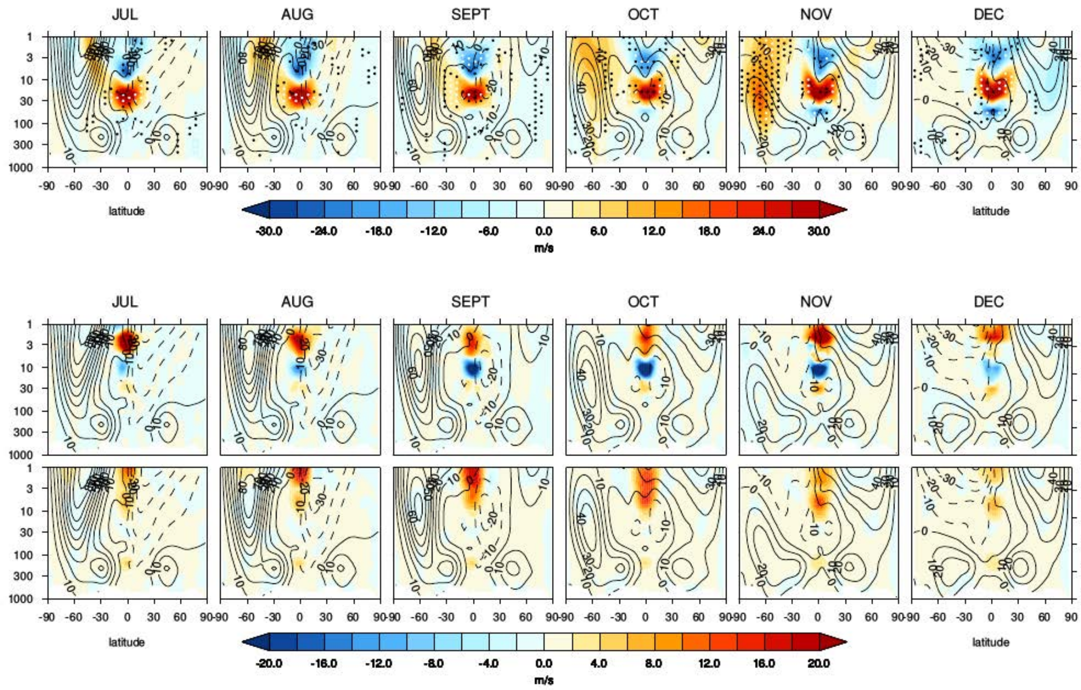


Figure AS9.33: *As* ***Figure 9.53*** *but comparing MERRA and MERRA-2 (MERRA minus MERRA-2) for the period 1980 - 2012.*

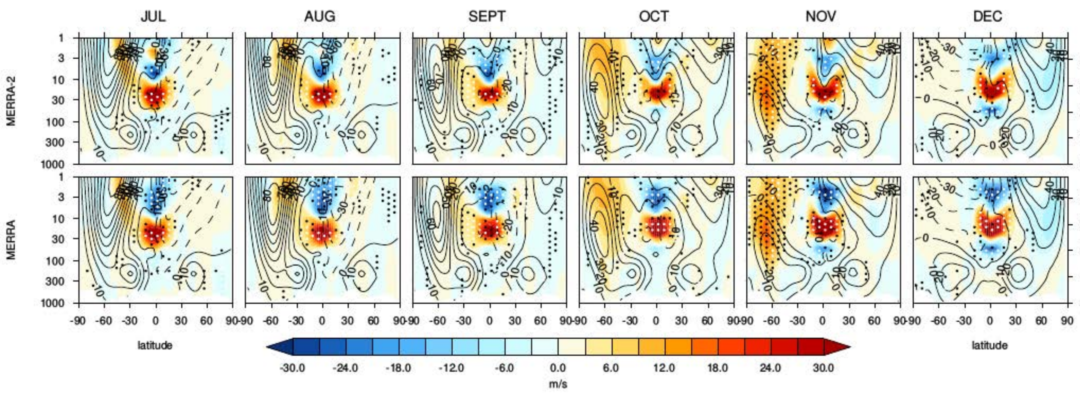


Figure AS9.34: *As* ***Figure AS9.33*** *top row, but showing the QBO signal from each of the reanalyses separately.*

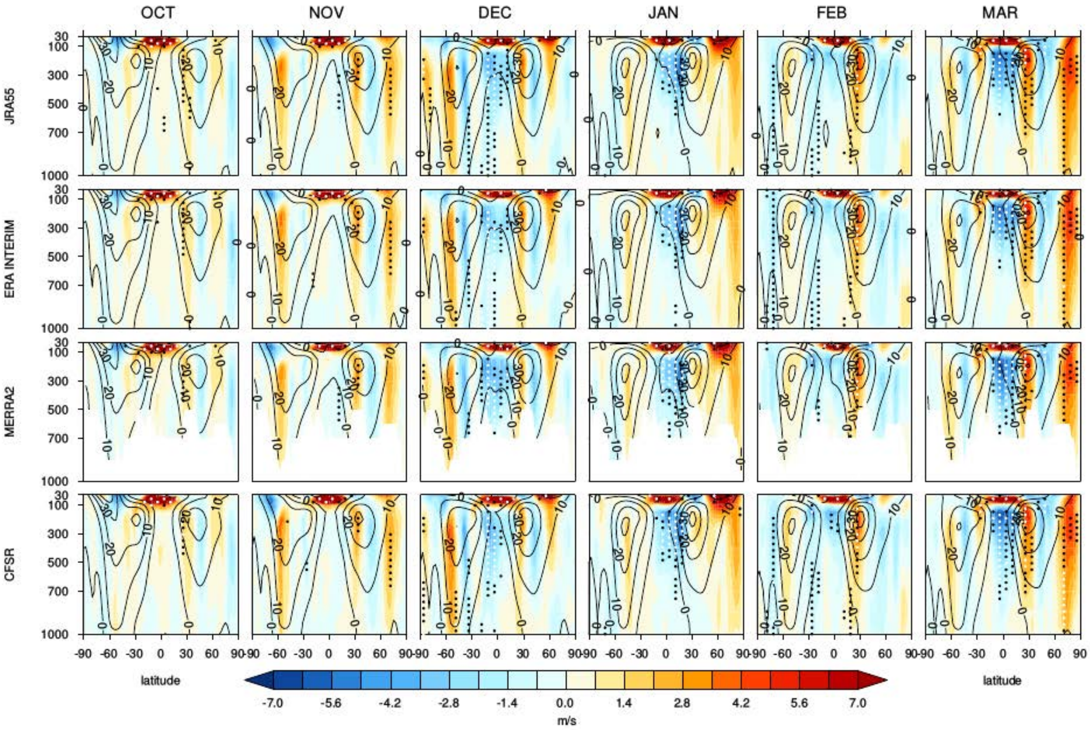


Figure AS9.35: *As **Figure 9.55** top row, but showing the QBO signal from each of the four recent reanalyses datasets separately.*

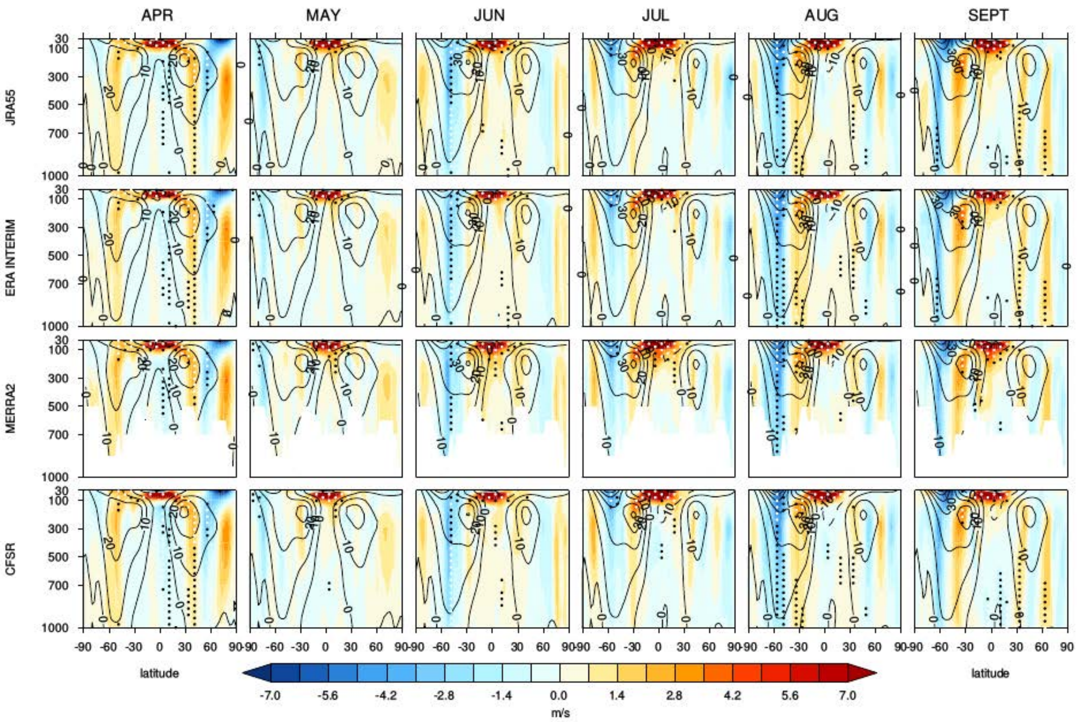


Figure AS9.36: *As **Figure 9.56** top row, but showing the QBO signal from each of the four recent reanalyses datasets separately.*

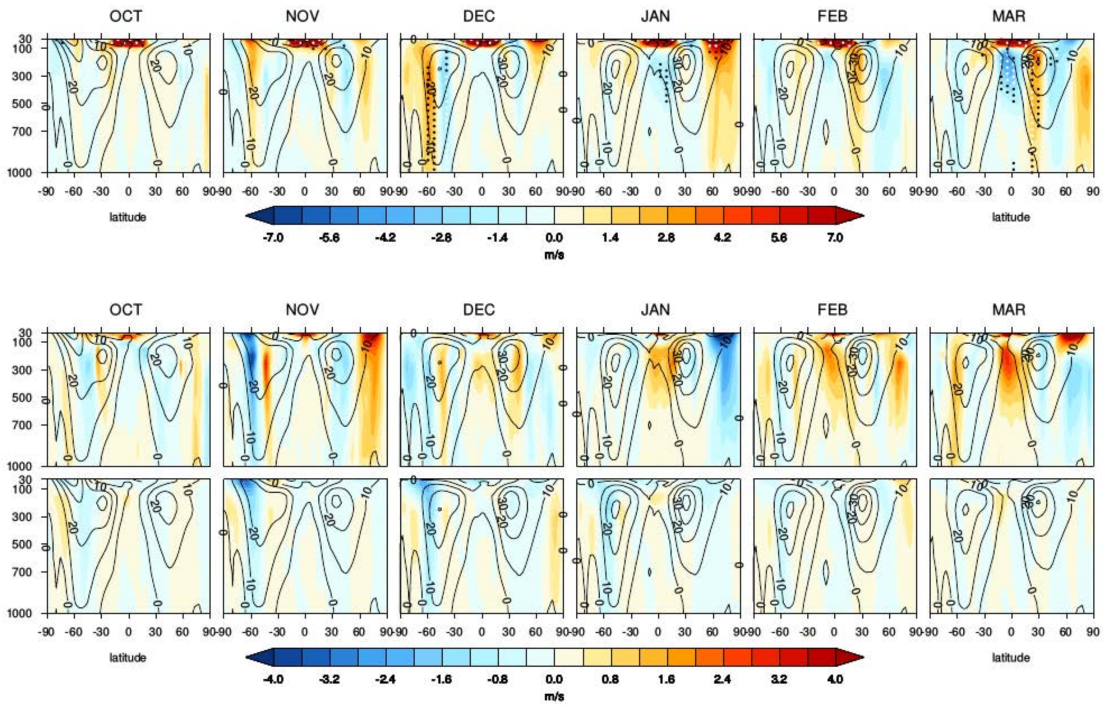


Figure AS9.37: *As* ***Figure 9.57*** *but showing the differences in QBO signals and climatologies. 1st row: the average QBO signal (together with the averaged climatologies and statistical significance levels). 2nd row: standard deviation (m s^{-1}) of the QBO signals. 3rd row: average of the interannual standard deviation (m s^{-1}) from the 2 reanalyses. Bottom row: standard deviation in the QBO signal as a percentage of the interannual variability (i.e., row 2 divided by row 3, multiplied by 100).*

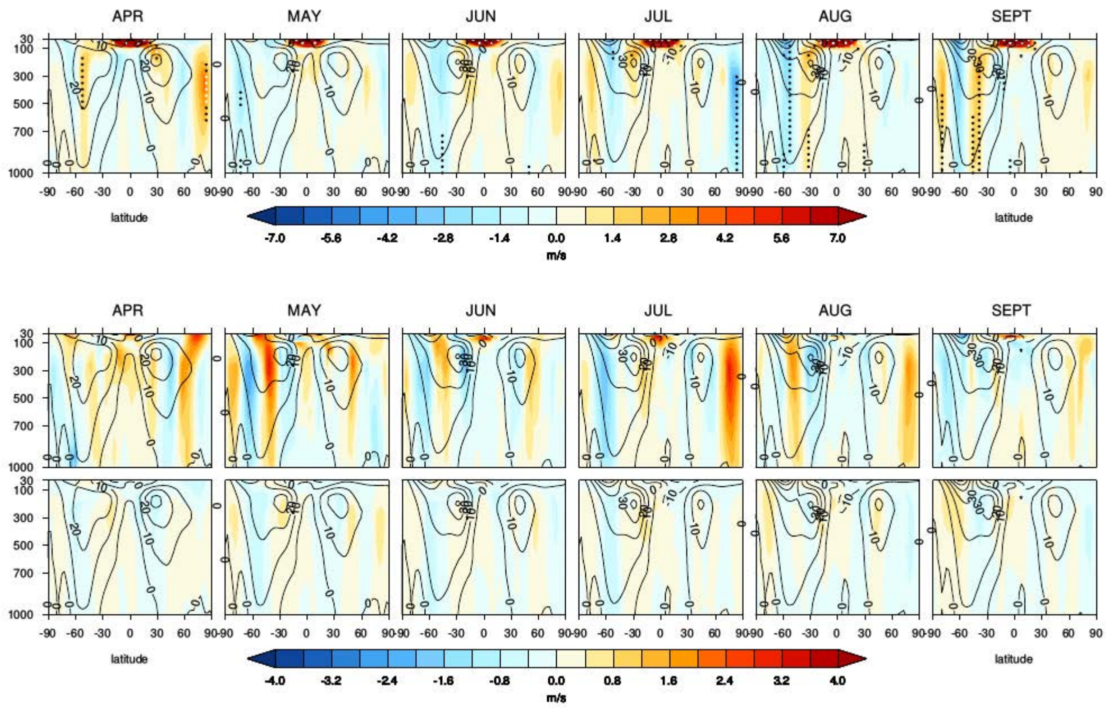


Figure AS9.38: *As* ***Figure AS9.37*** *but for April - September.*

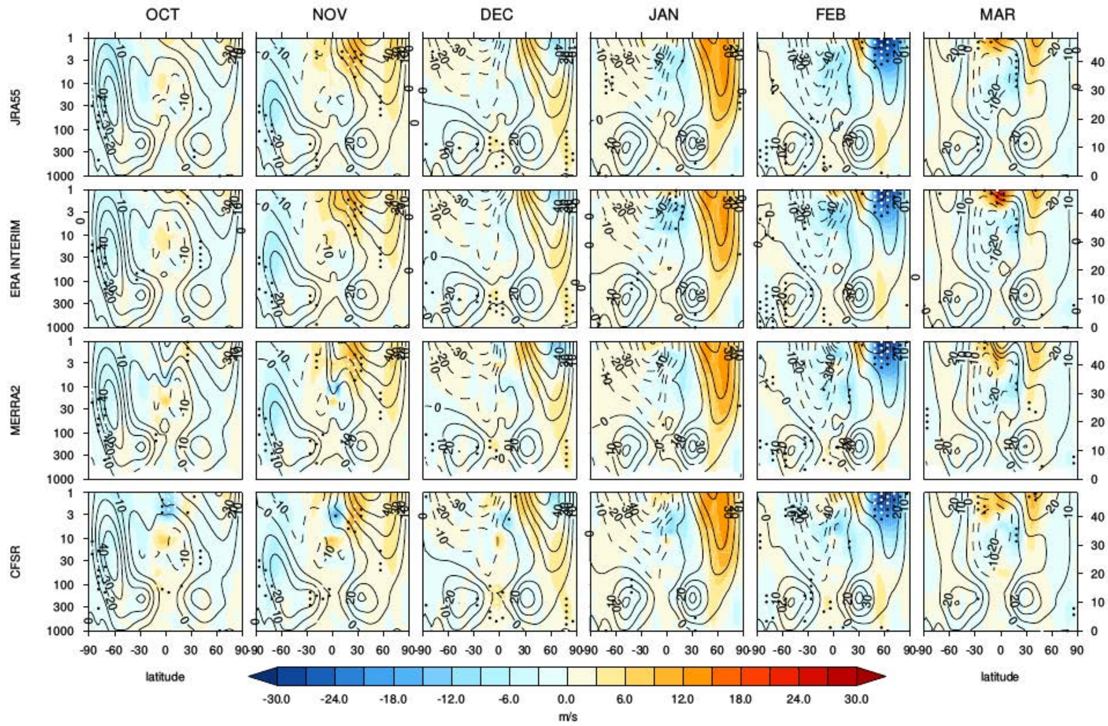


Figure AS9.39: *11-year solar cycle response for October - March for each of the four recent reanalysis datasets (JRA-55, ERA-Interim, MERRA-2, CFSR) for the period 1980 - 2016.*

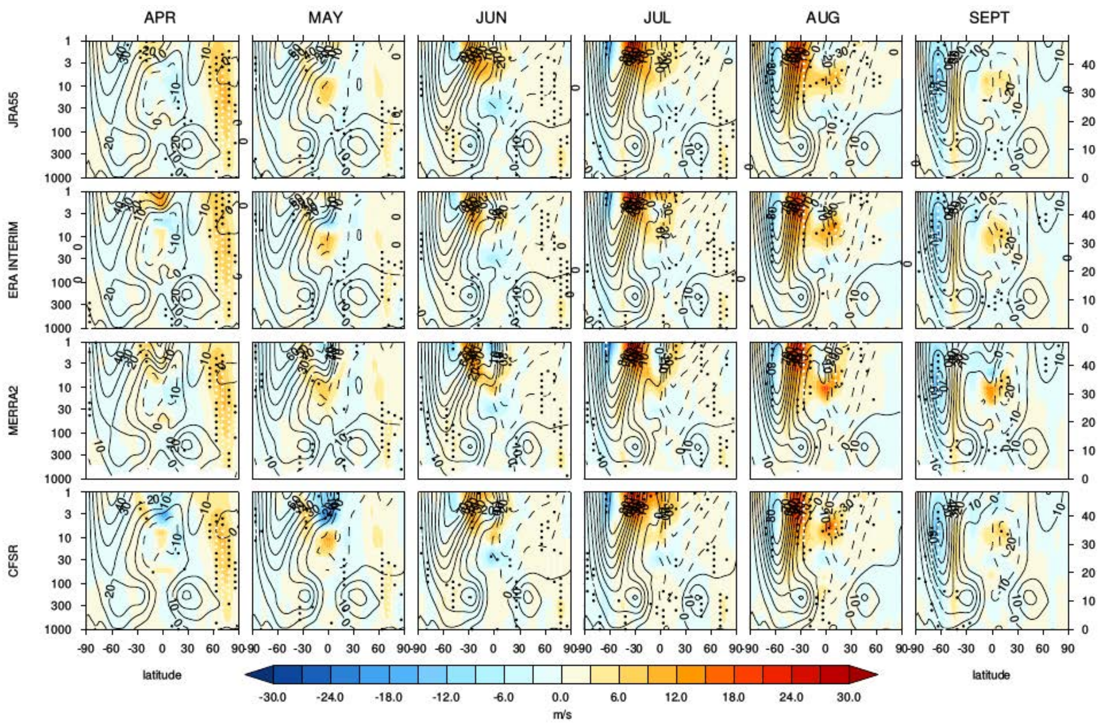


Figure AS9.40: *As **Figure AS9.39** but for April–September.*

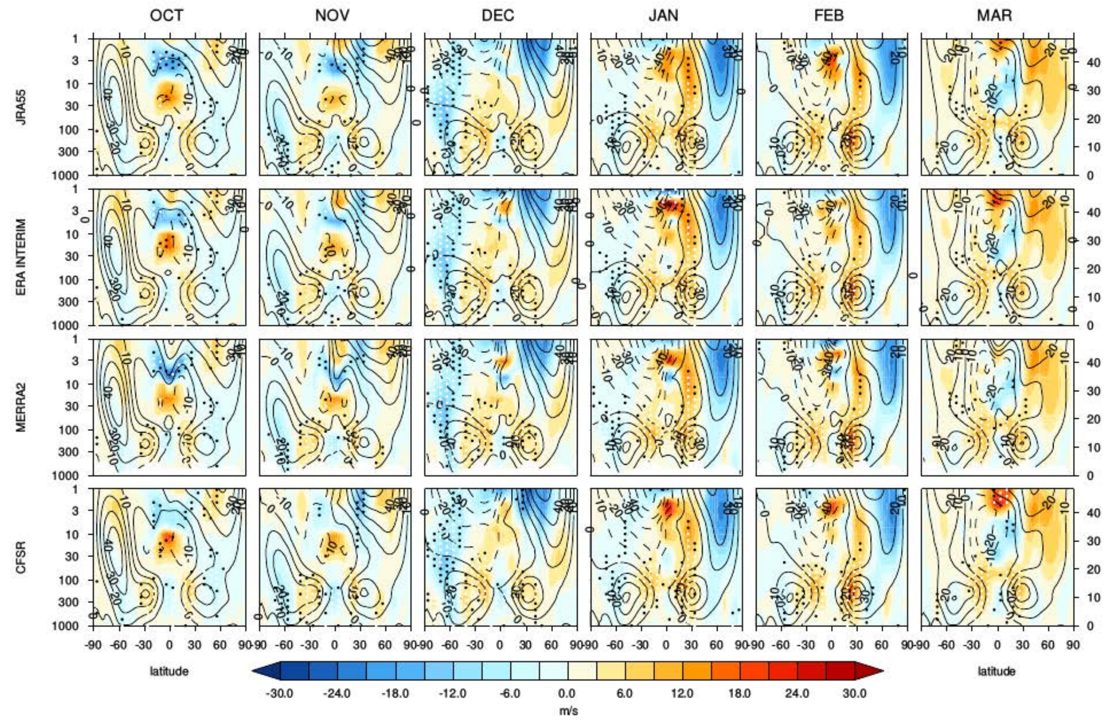


Figure AS9.41: *As **Figure AS9.39** but for the ENSO signal.*

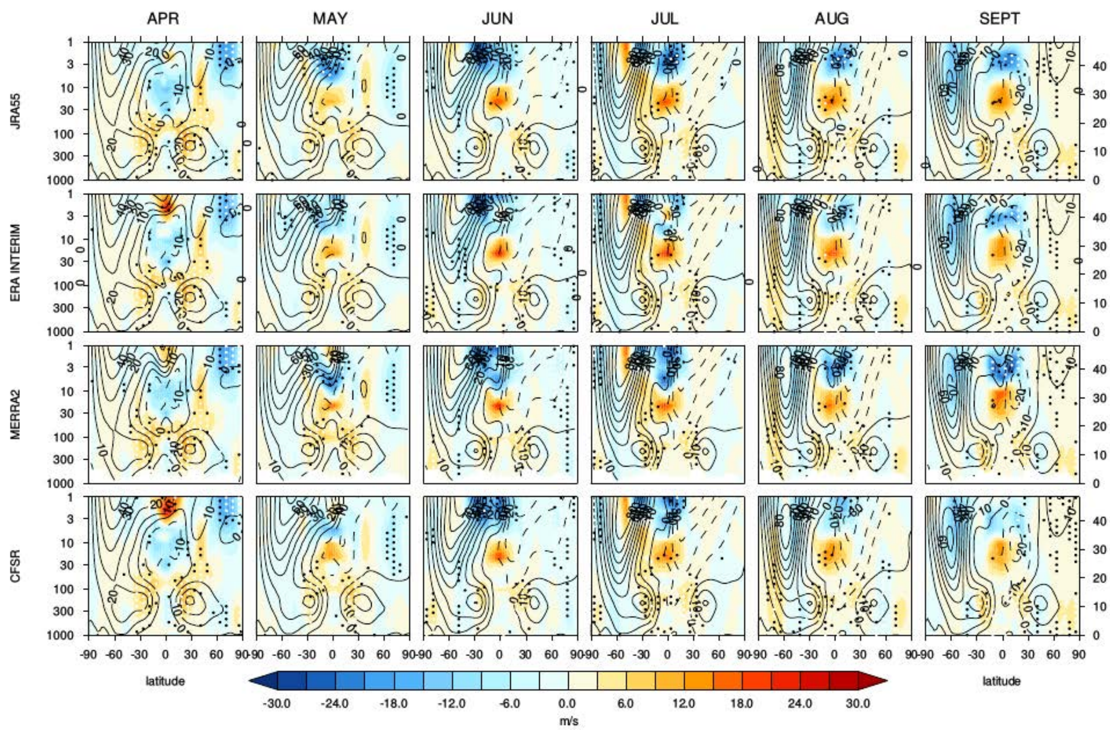


Figure AS9.42: *As **Figure AS9.41** but for April - September.*

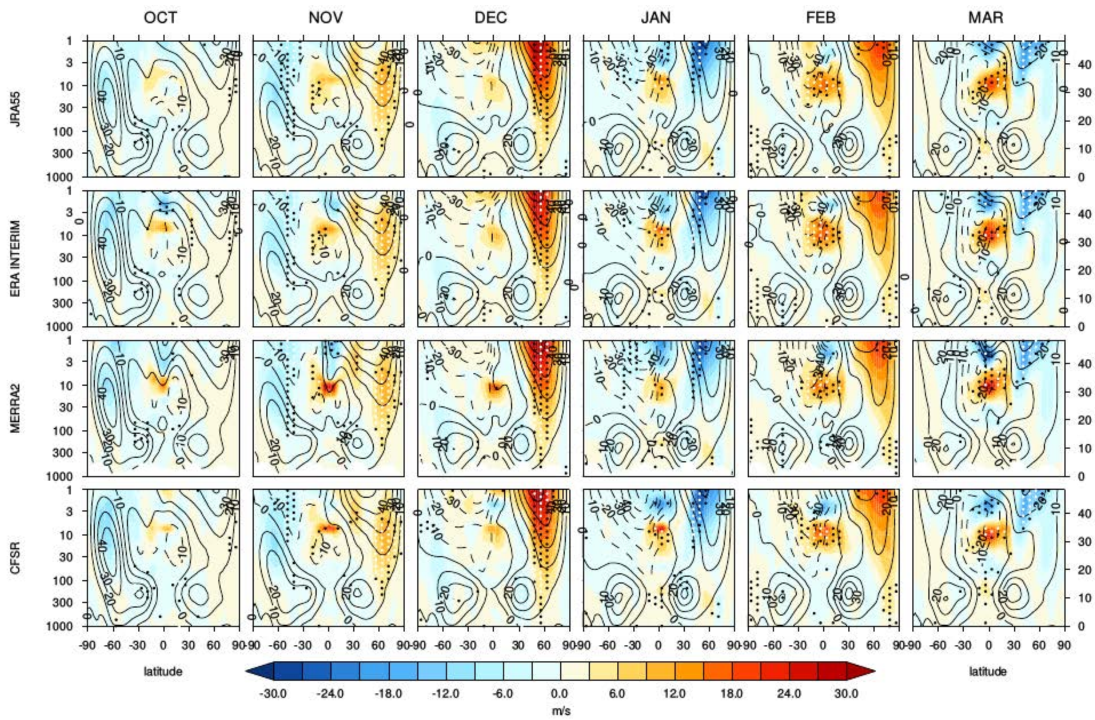


Figure AS9.43: *As **Figure AS9.39** but for the volcanic signal.*

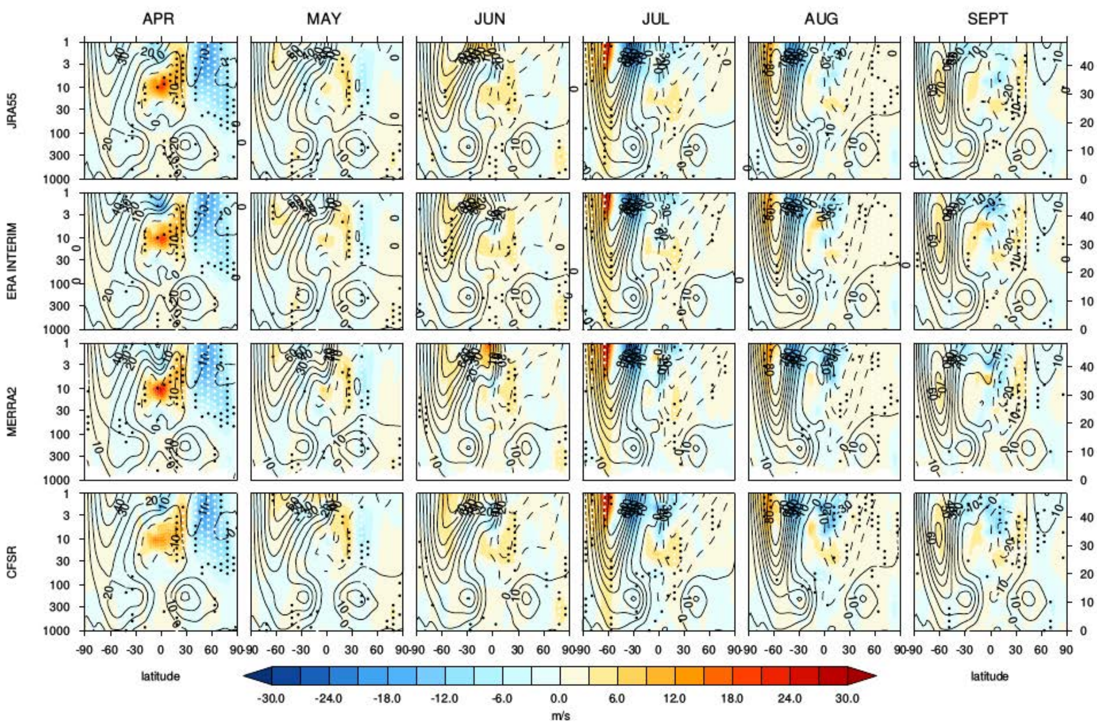


Figure AS9.44: *As **Figure AS9.43** but for April - September.*

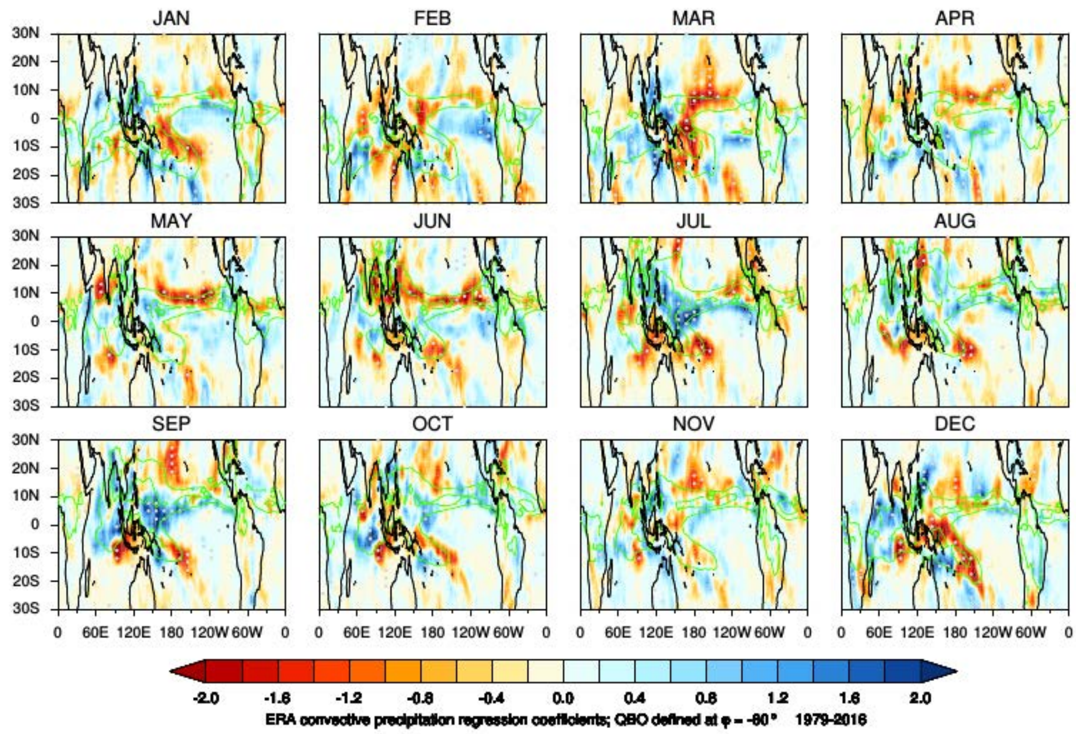


Figure AS9.45: *As* ***Figure 9.63****, but showing the response in the convective component of ERA-Interim total rainfall for 1979 - 2016. Filled contours show the regression coefficient for convective precipitation, consistent with Gray et al. (2018).*

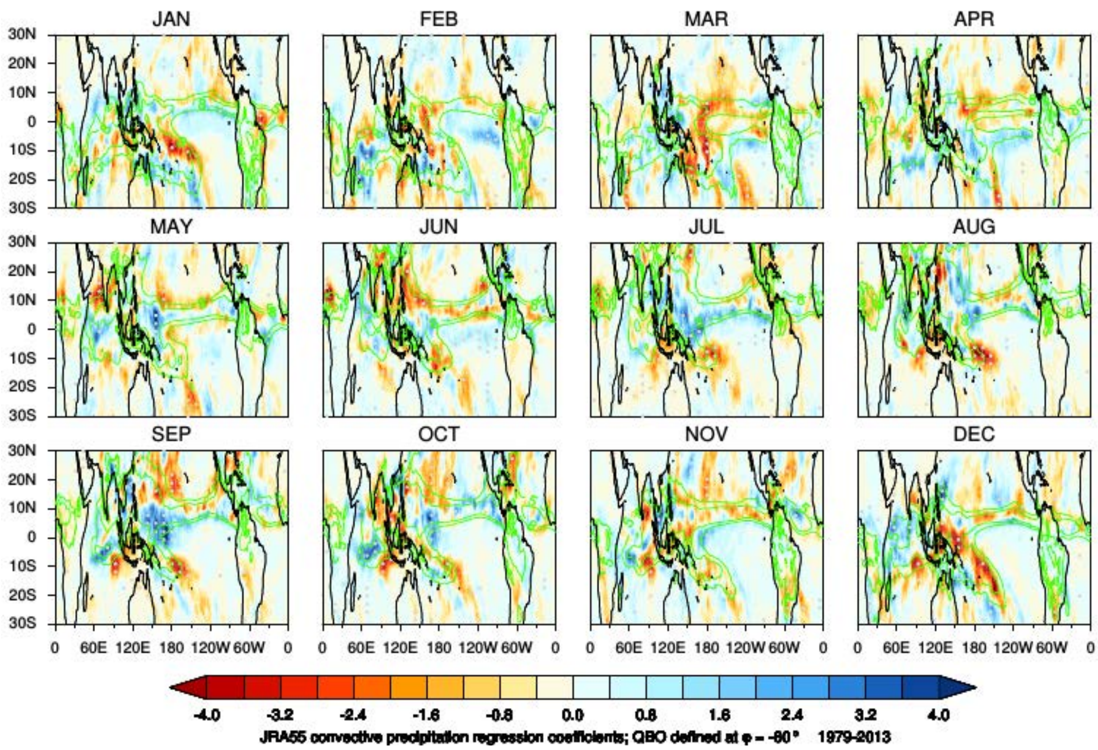


Figure AS9.46: *As* ***Figure 9.64****, but showing the response in the convective component of JRA-55 total rainfall for 1979 - 2013. Filled contours show the regression coefficient for convective precipitation, consistent with Gray et al. (2018).*

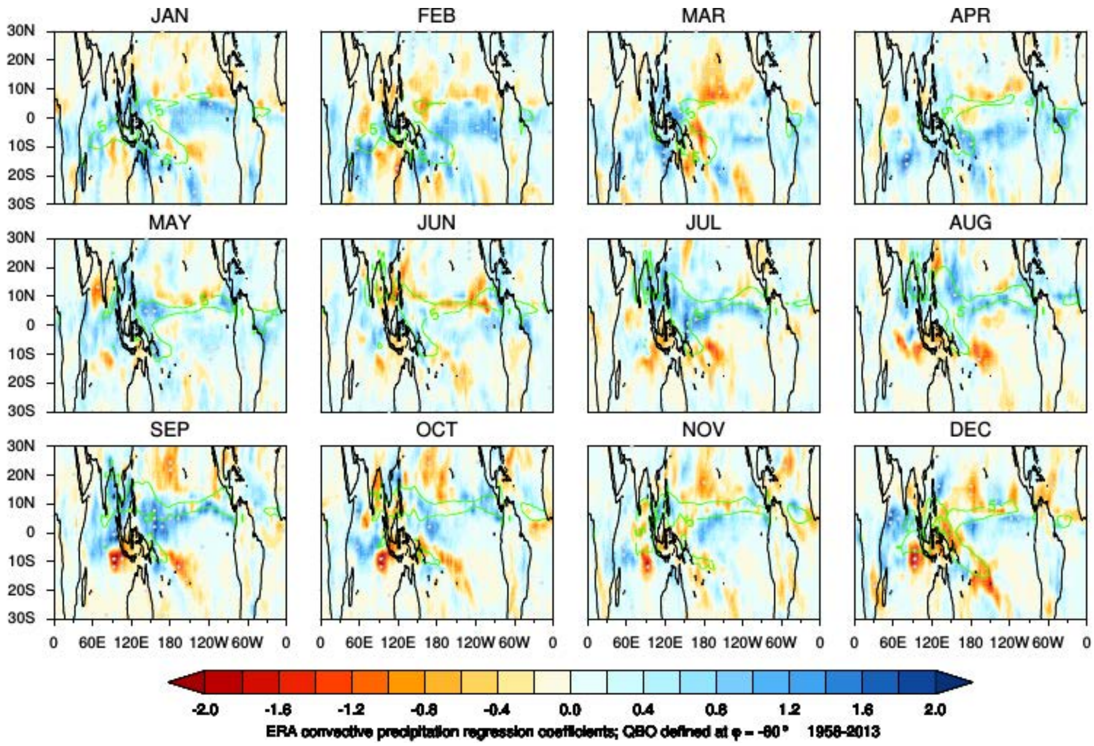


Figure AS9.47: *As* ***Figure 9.63****, but showing the response in the convective component of ERA reanalysis (concatenated ERA-40 and ERA-Interim, as described in the text) total rainfall for 1958 - 2013. Filled contours show the regression coefficient for convective precipitation, consistent with Gray et al. (2018).*

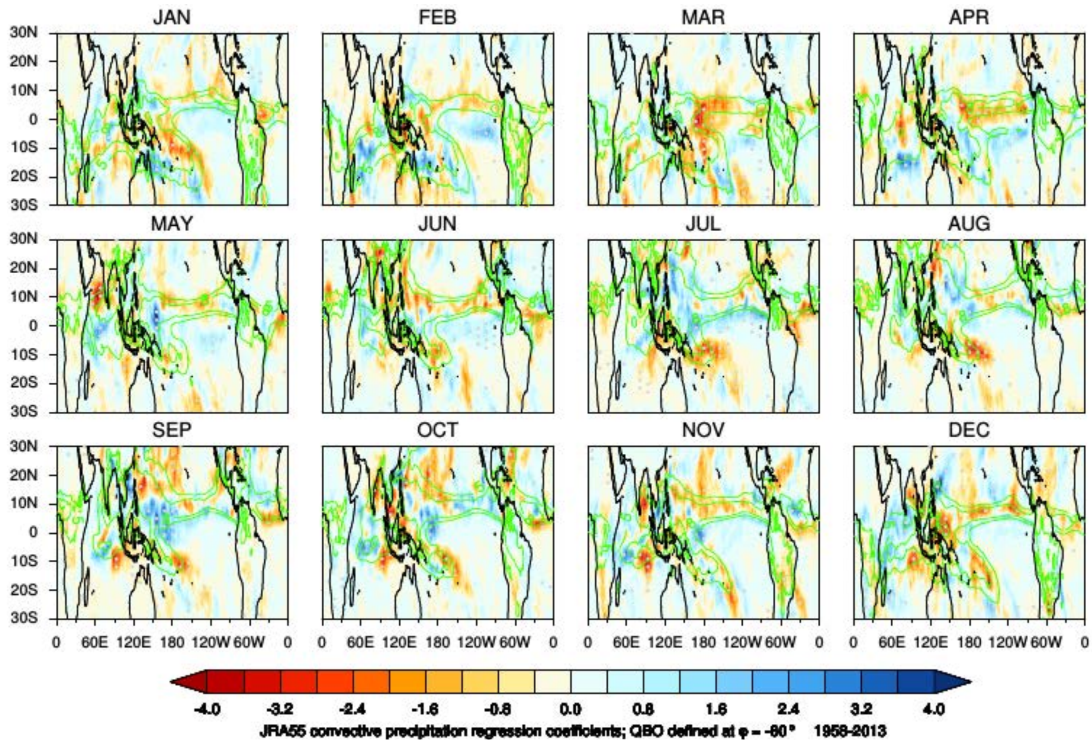


Figure AS9.48: *As* ***Figure 9.64****, but showing the response in the convective component of JRA-55 total rainfall for 1958 - 2013. Filled contours show the regression coefficient for convective precipitation, consistent with Gray et al. (2018).*

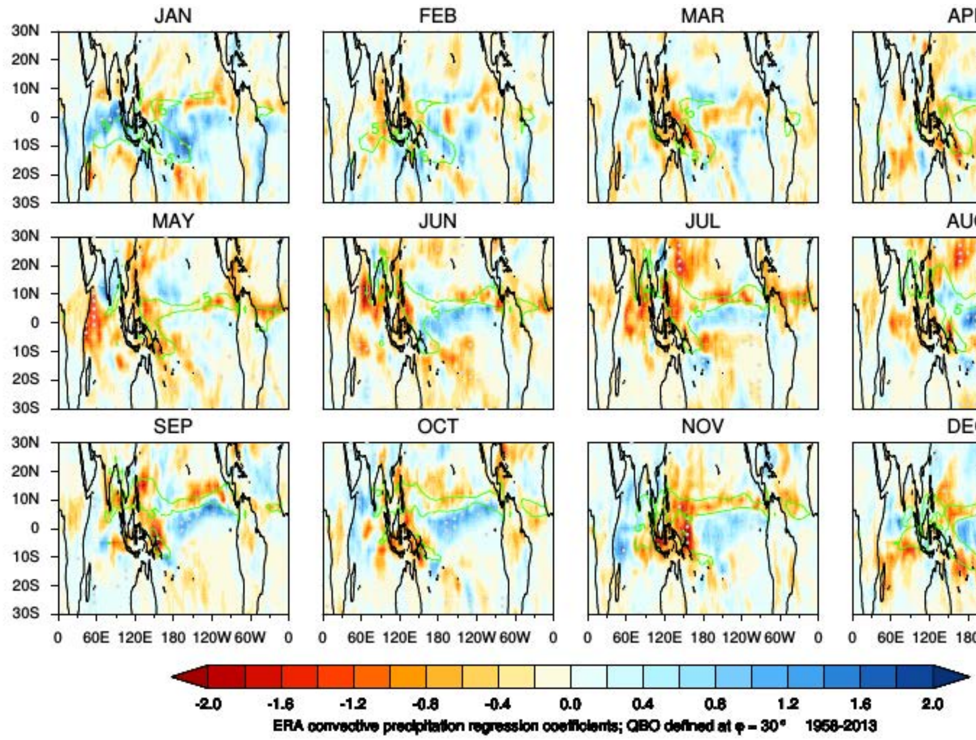


Figure AS9.49: *As **Figure 9.63**, but showing the response in the convective component of ERA reanalysis (concatenated ERA-40 and ERA-Interim, as described in the text) total rainfall for 1958 - 2013 using a QBO defined at phase angle of + 30° instead of -60°. Filled contours show the regression coefficient for convective precipitation, consistent with Gray et al. (2018).*

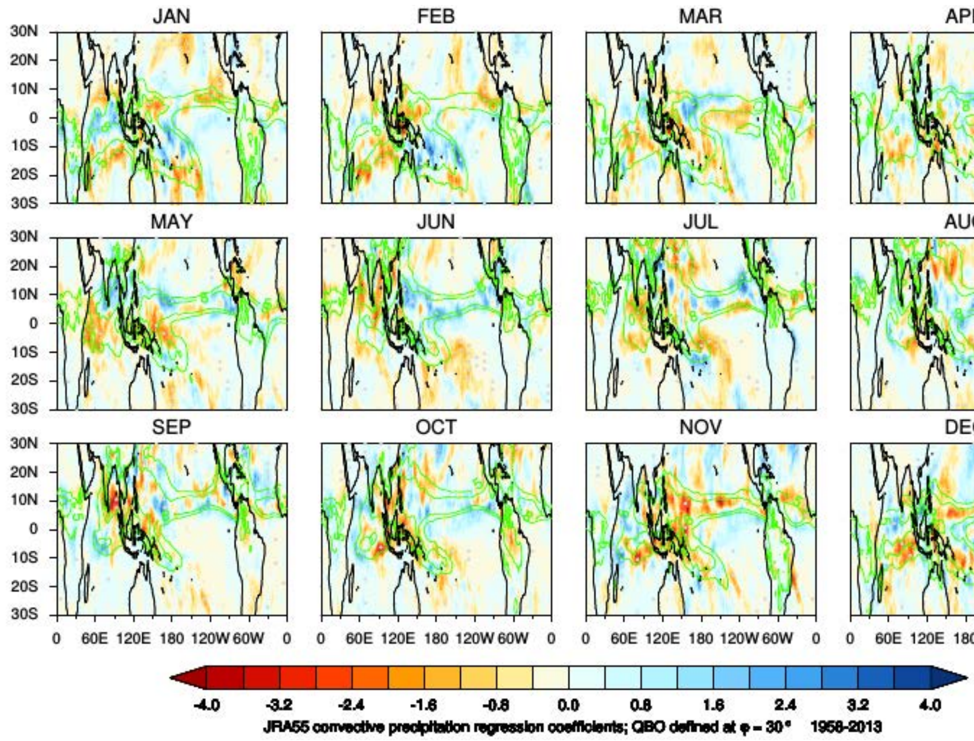


Figure AS9.50: *As **Figure 9.64**, but showing the response in the convective component of JRA-55 total rainfall for 1958 - 2013 using a QBO defined at phase angle of + 30° instead of - 60°. Filled contours show the regression coefficient for convective precipitation, consistent with Gray et al. (2018).*

A9.2 Supplement A: QBO winds in all reanalyses

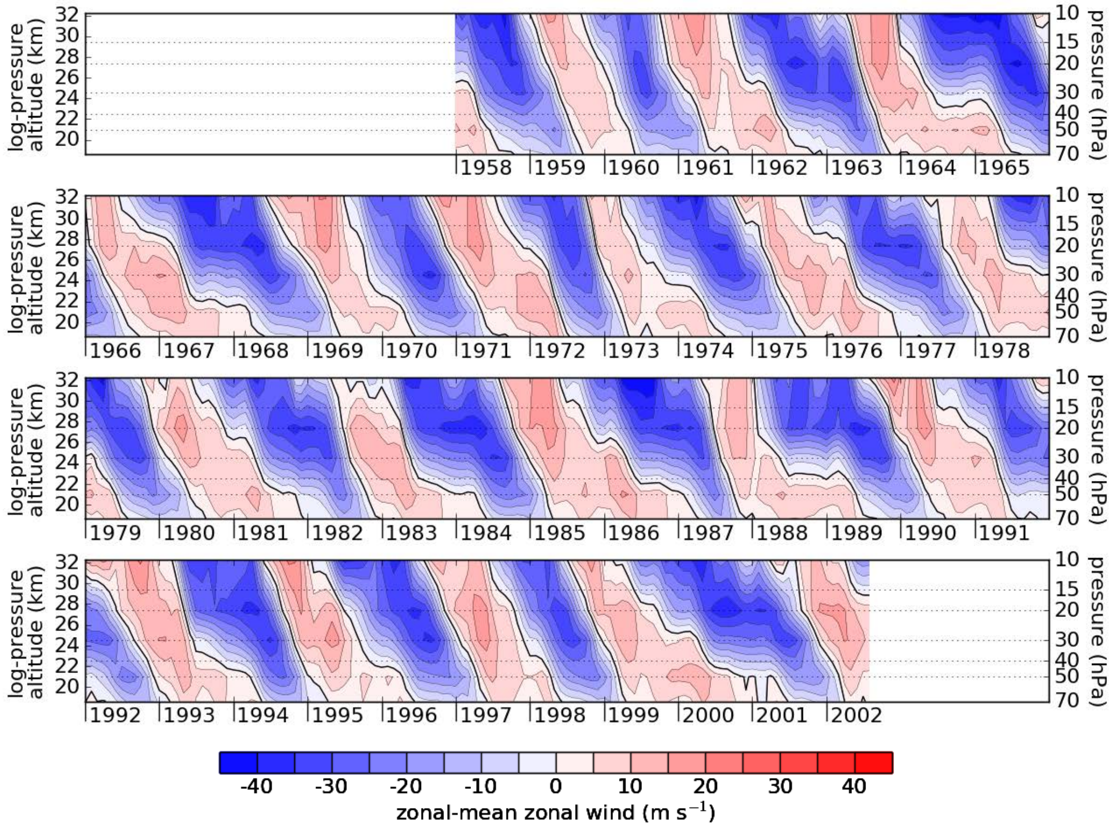


Figure AA9.1: *Time-series of monthly-mean equatorial (2° S - 2° N average) zonal-mean zonal wind (m s^{-1}) for ERA-40.*

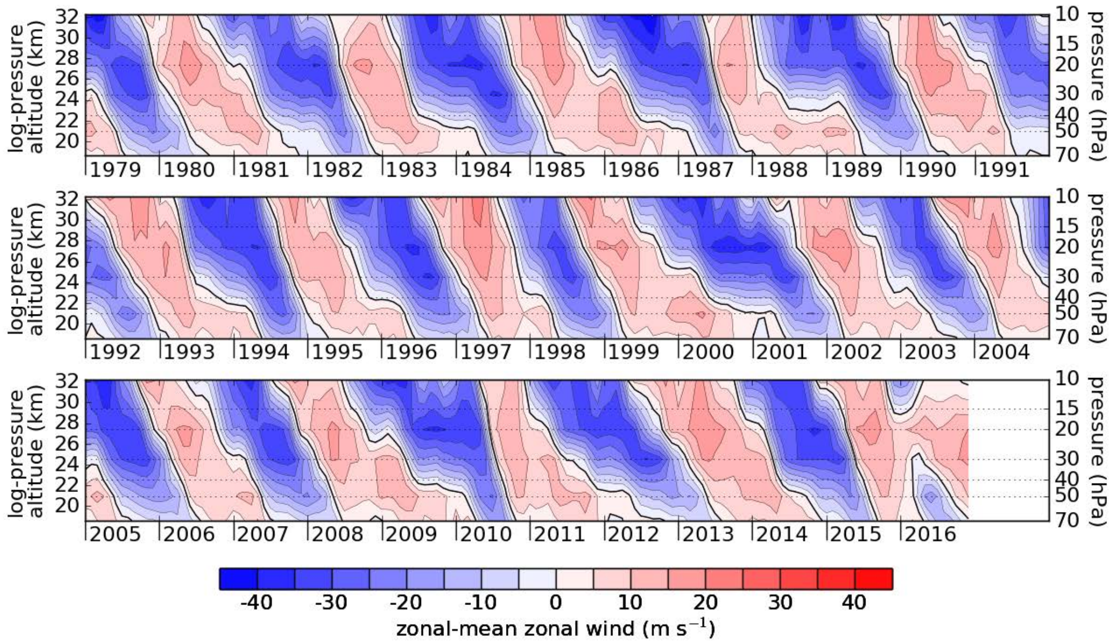


Figure AA9.2: *As **Figure AA9.1**, but for ERA-Interim.*

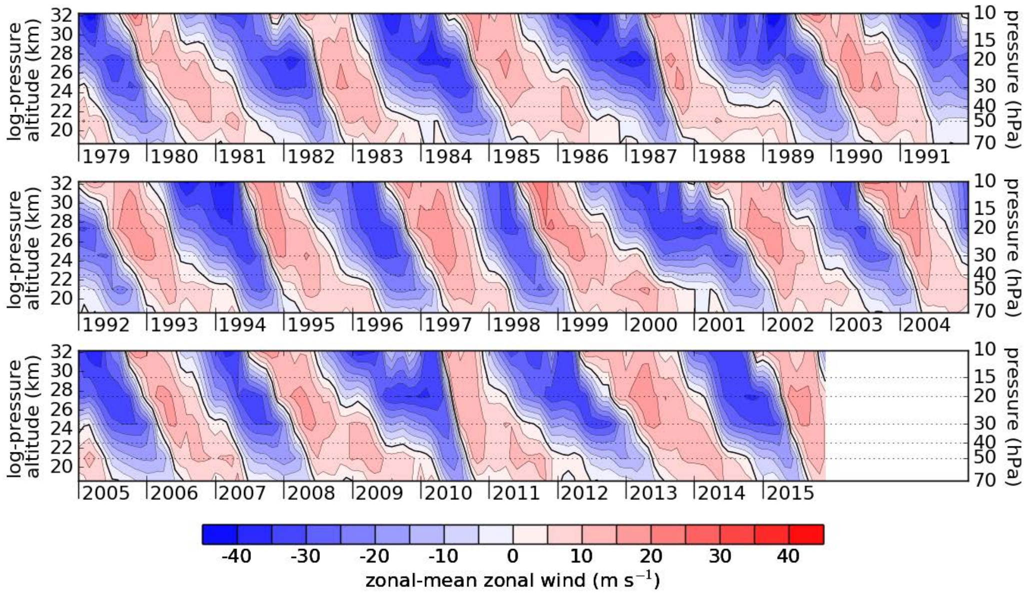


Figure AA9.3: *As **Figure AA9.1**, but for MERRA.*

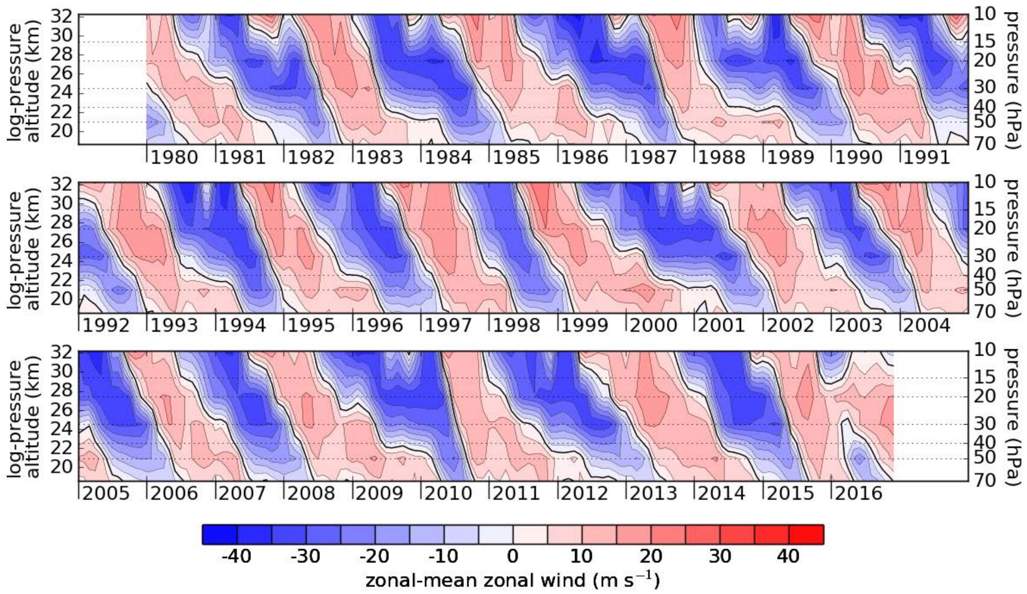


Figure AA9.4: *As **Figure AA9.1**, but for MERRA-2.*

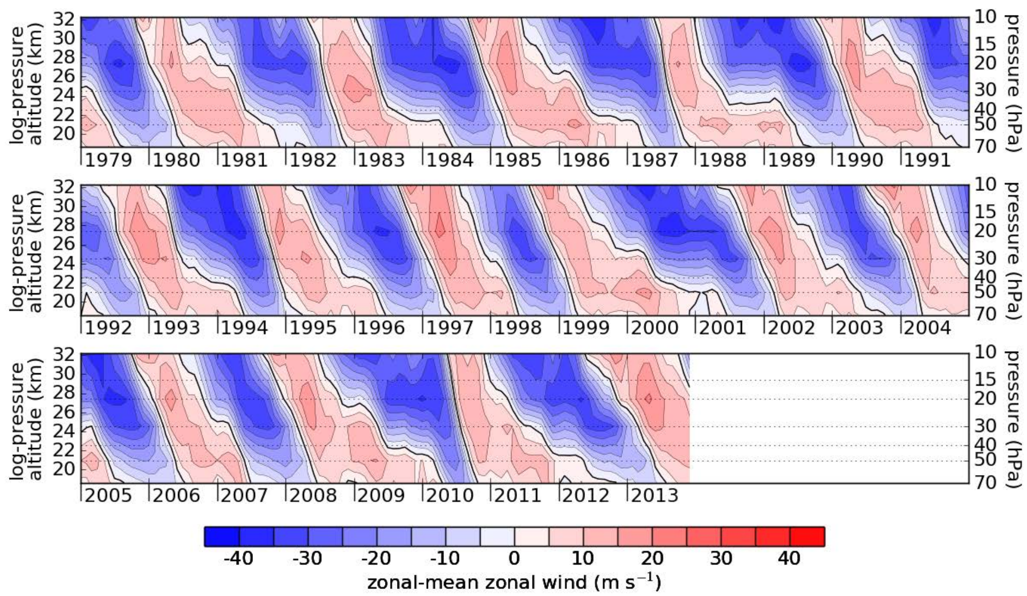


Figure AA9.5: *As **Figure AA9.1**, but for JRA-25.*

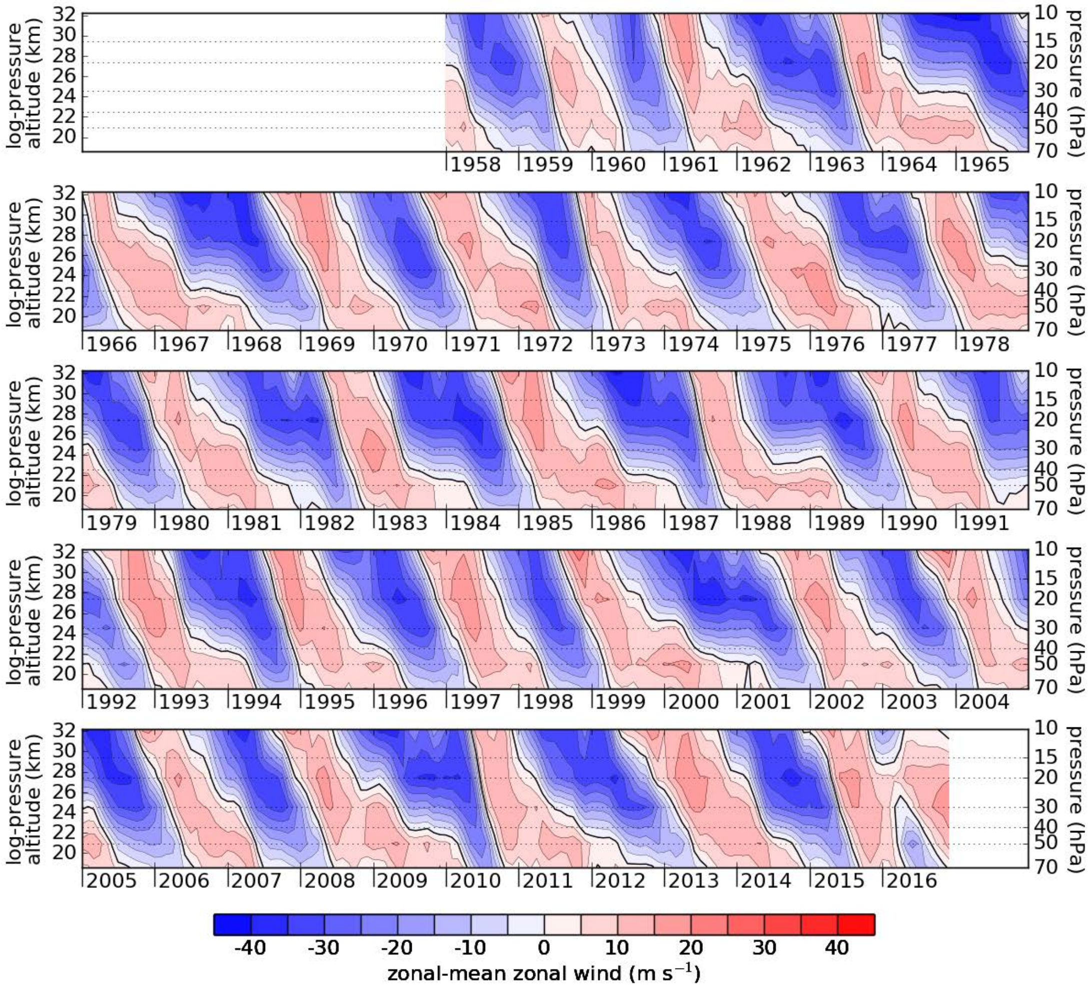


Figure AA9.6: *As **Figure AA9.1**, but for JRA-55.*

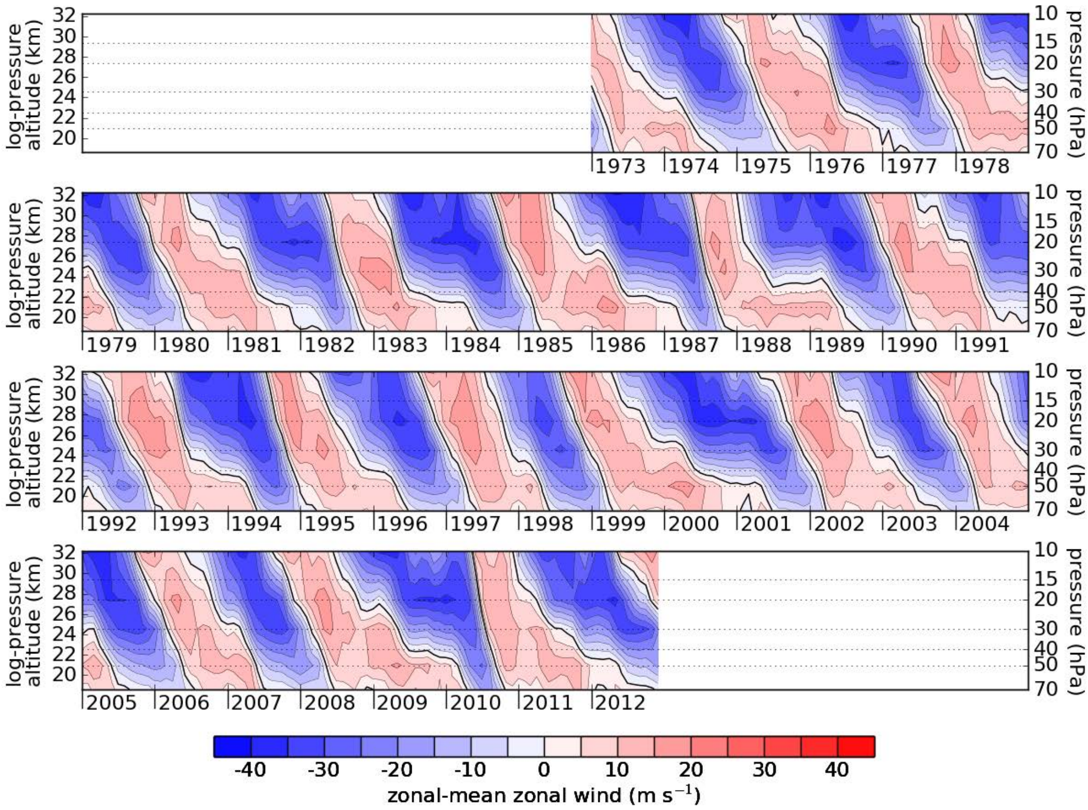


Figure AA9.7: *As **Figure AA9.1**, but for JRA-55C.*

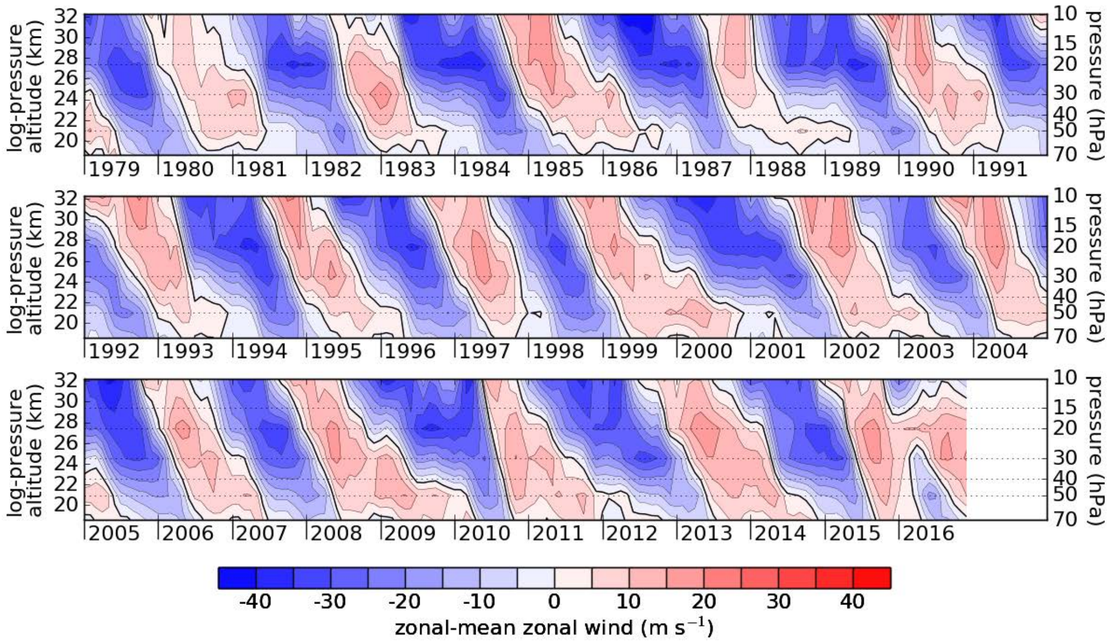


Figure AA9.8: *As **Figure AA9.1**, but for CFSR.*

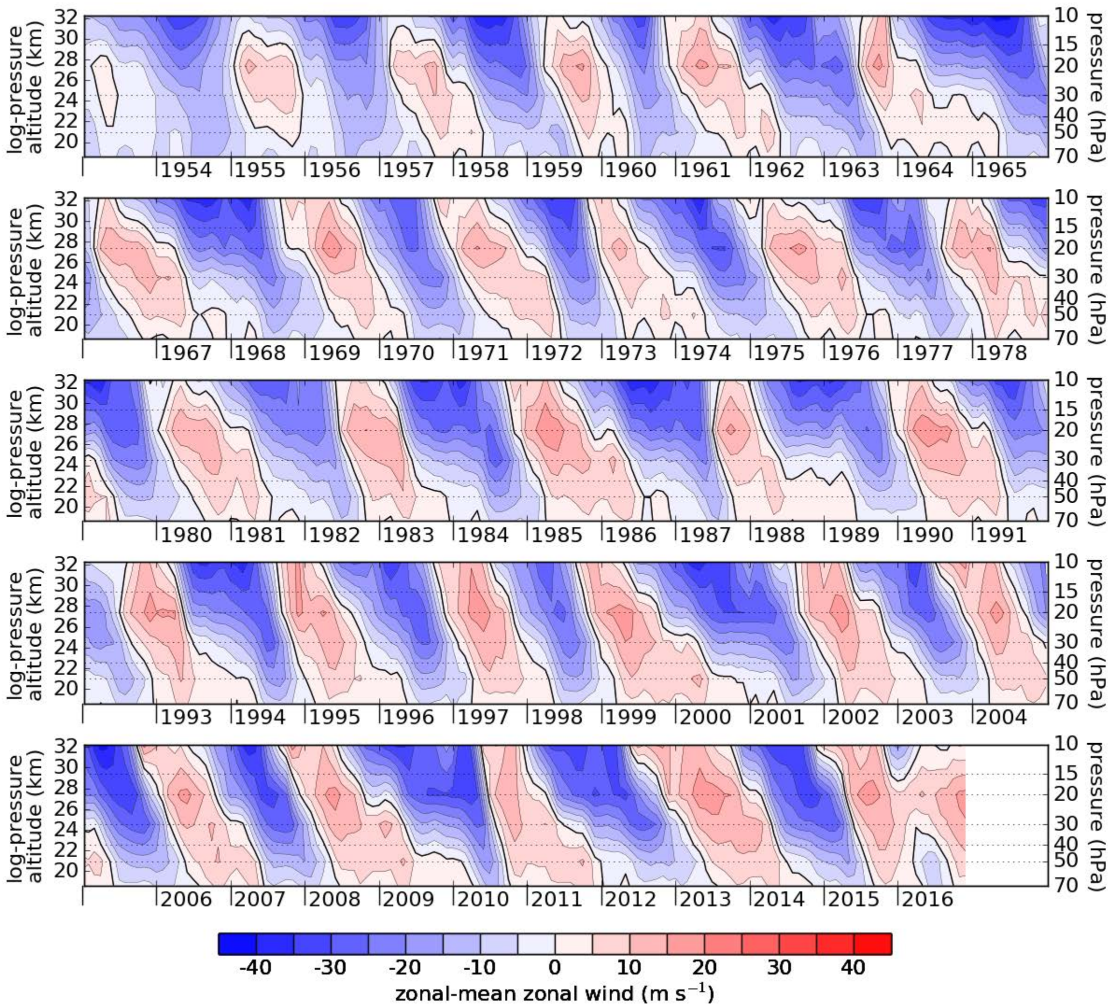


Figure AA9.9: *As **Figure AA9.1**, but for NCEP-NCAR (R1).*

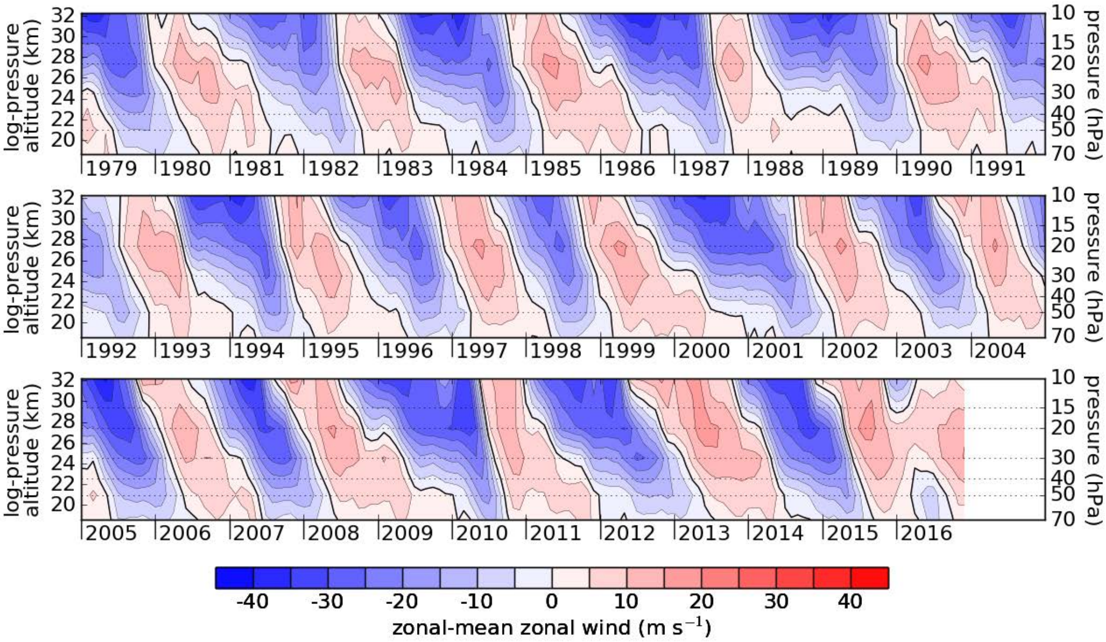


Figure AA9.10: *As **Figure AA9.1**, but for NCEP-DOE (R2).*

A9.3 Supplement B: QBO metrics for all reanalyses

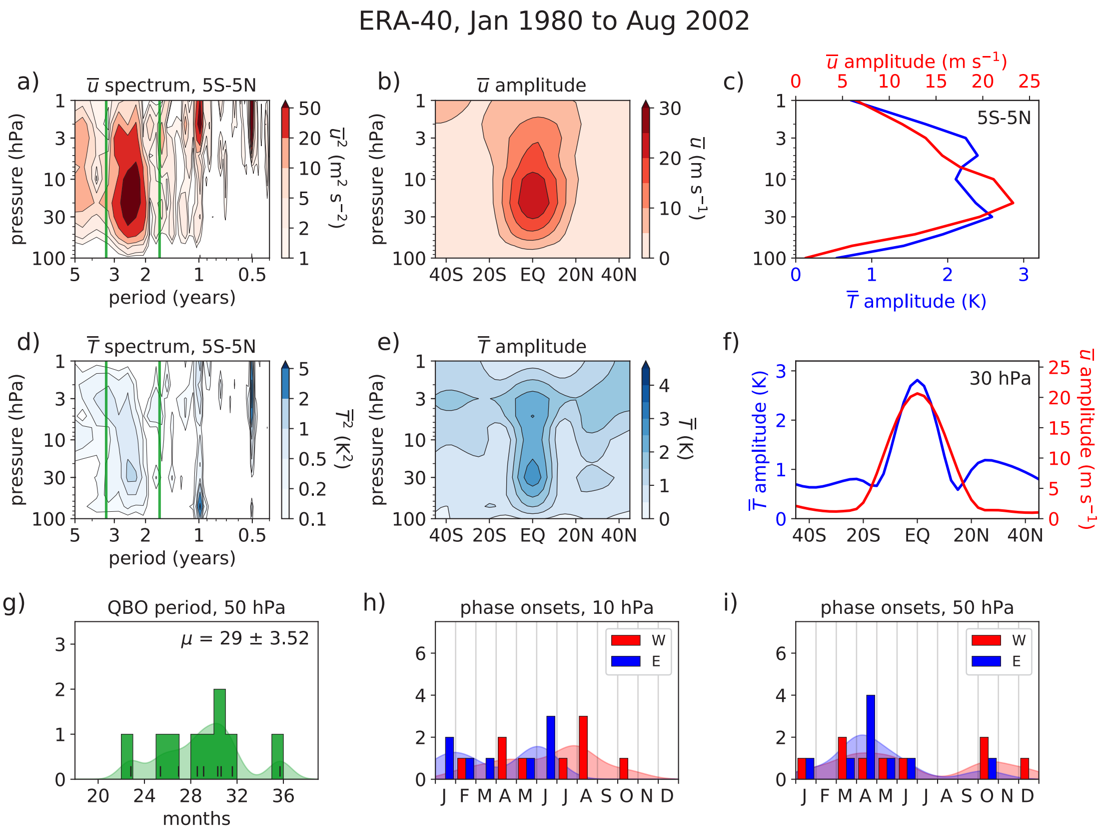


Figure AB9.1: *As* ***Figure 9.15****, but for ERA-40 for the 1980 - 2002 period.*

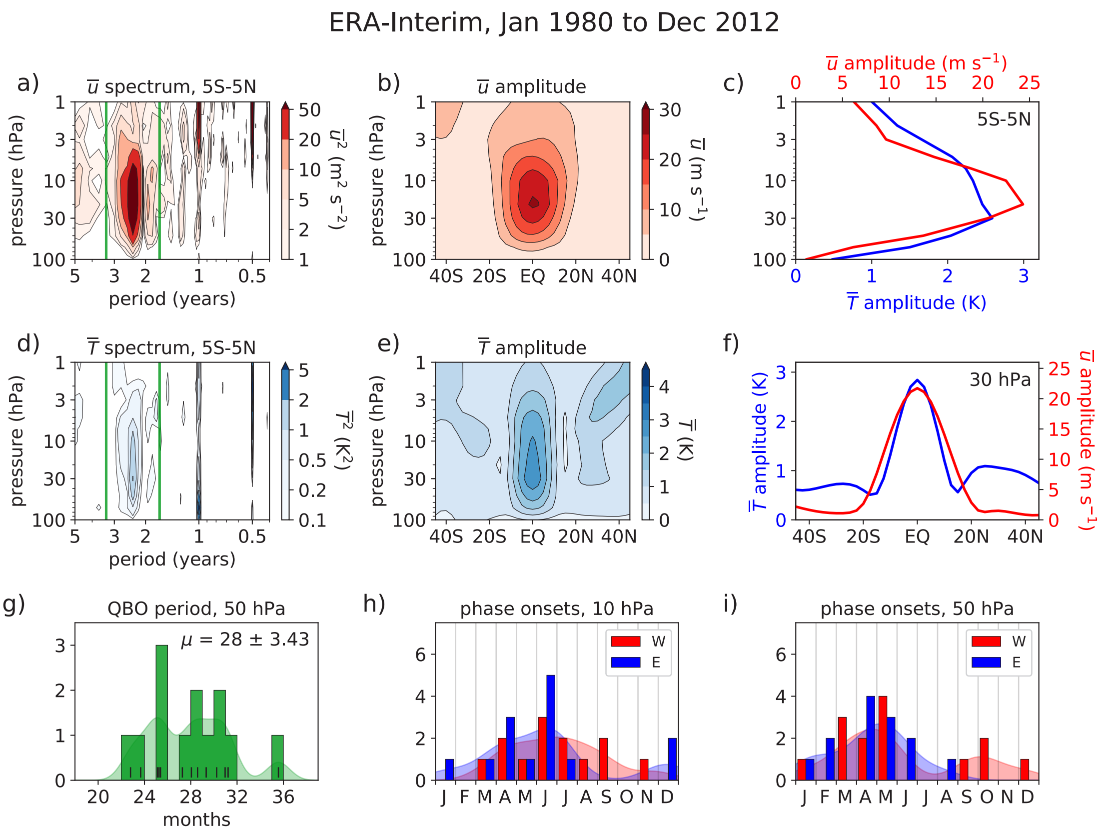


Figure AB9.2: *As* ***Figure 9.15****, but for ERA-Interim for the 1980 - 2012 period.*

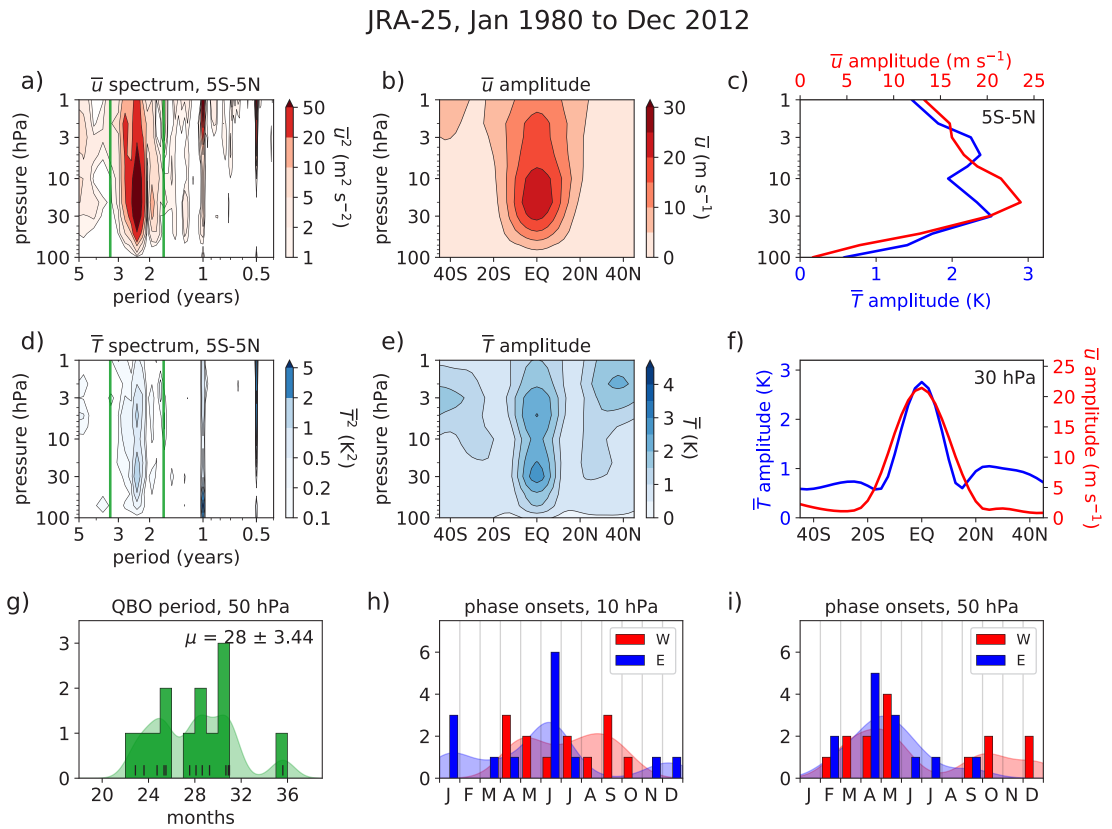


Figure AB9.3: *As* ***Figure 9.15****, but for JRA-25 for the 1980 - 2012 period.*

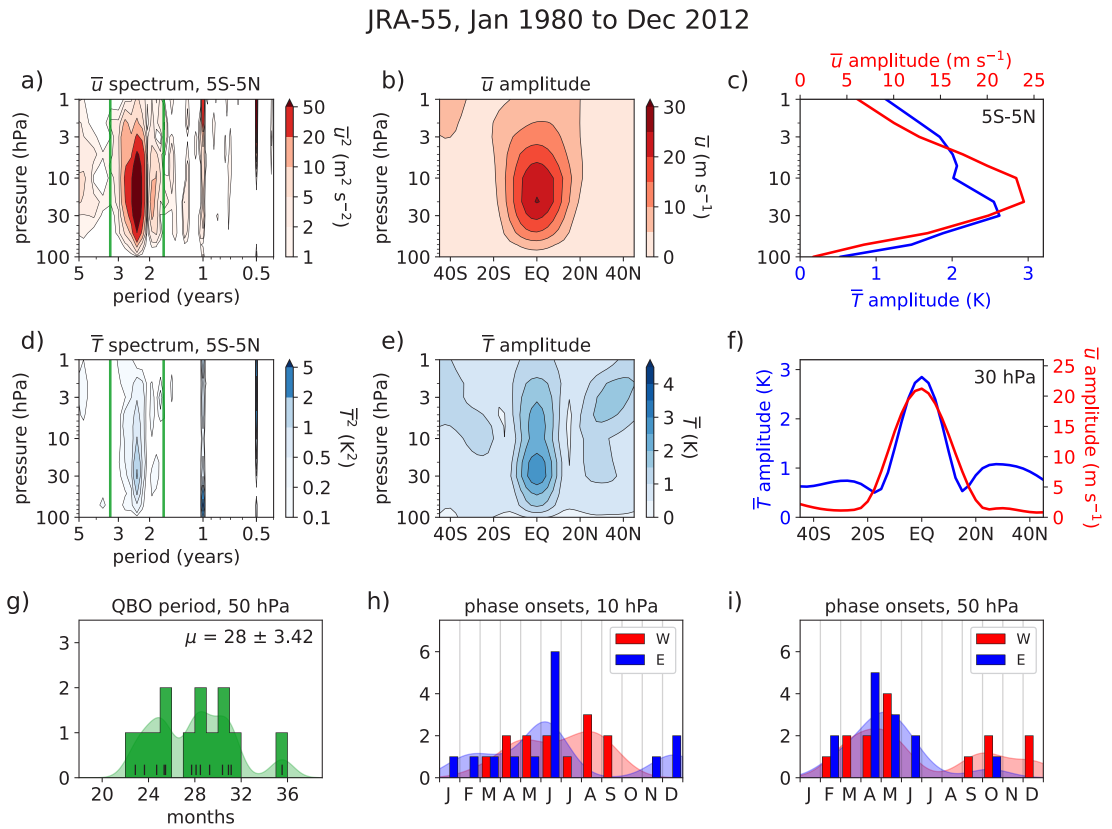


Figure AB9.4: *As* ***Figure 9.15****, but for JRA-55 for the 1980 - 2012 period.*

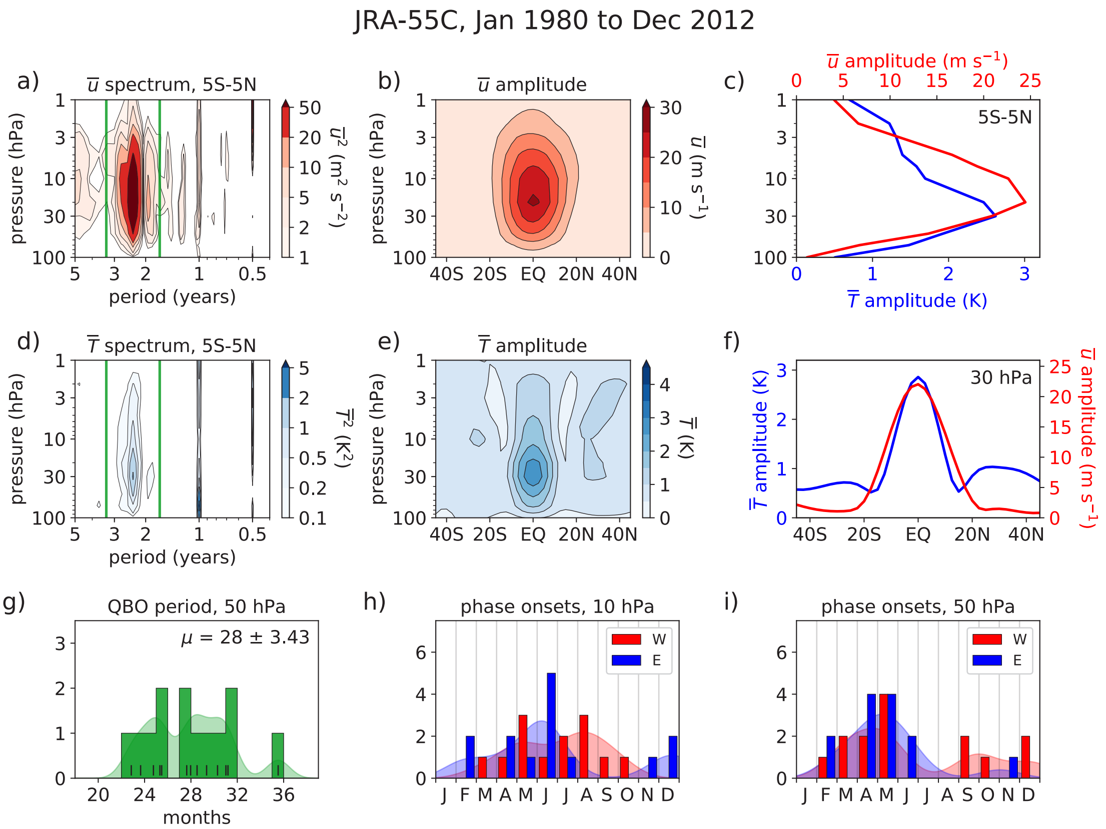


Figure AB9.5: *As* ***Figure 9.15****, but for JRA-55C for the 1980 - 2012 period.*

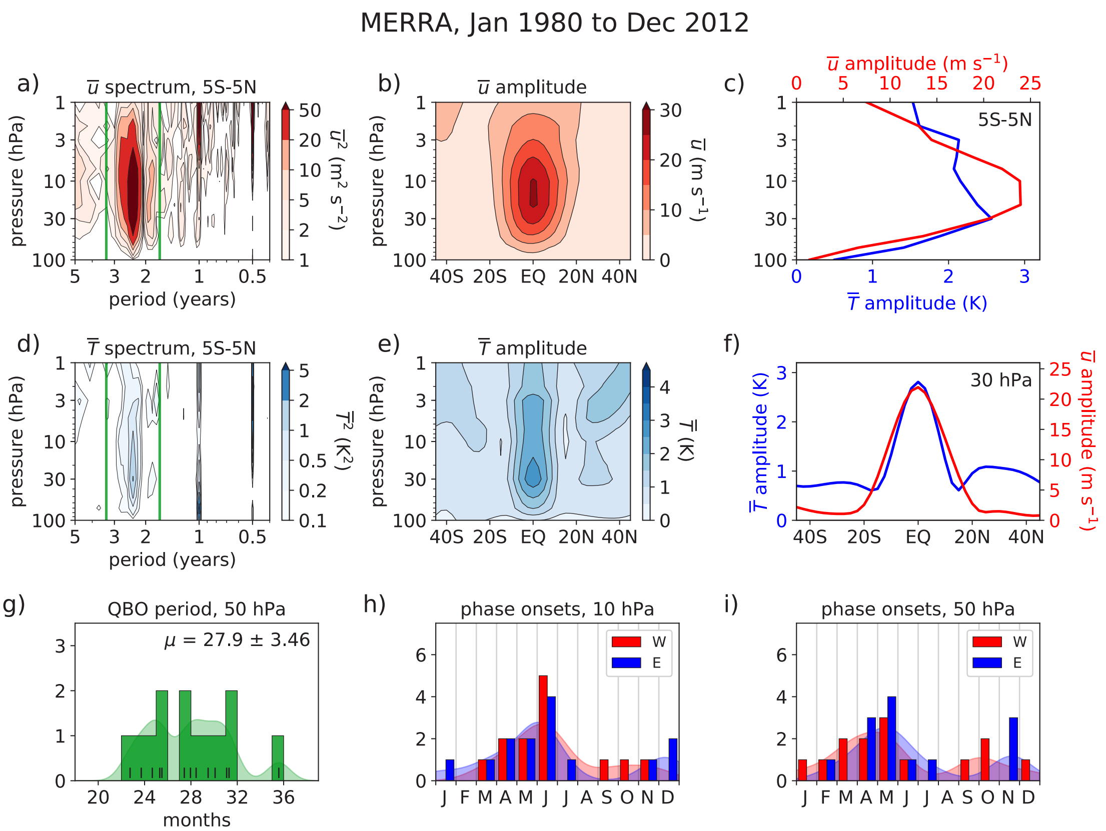


Figure AB9.6: *As* ***Figure 9.15****, but for MERRA for the 1980 - 2012 period.*

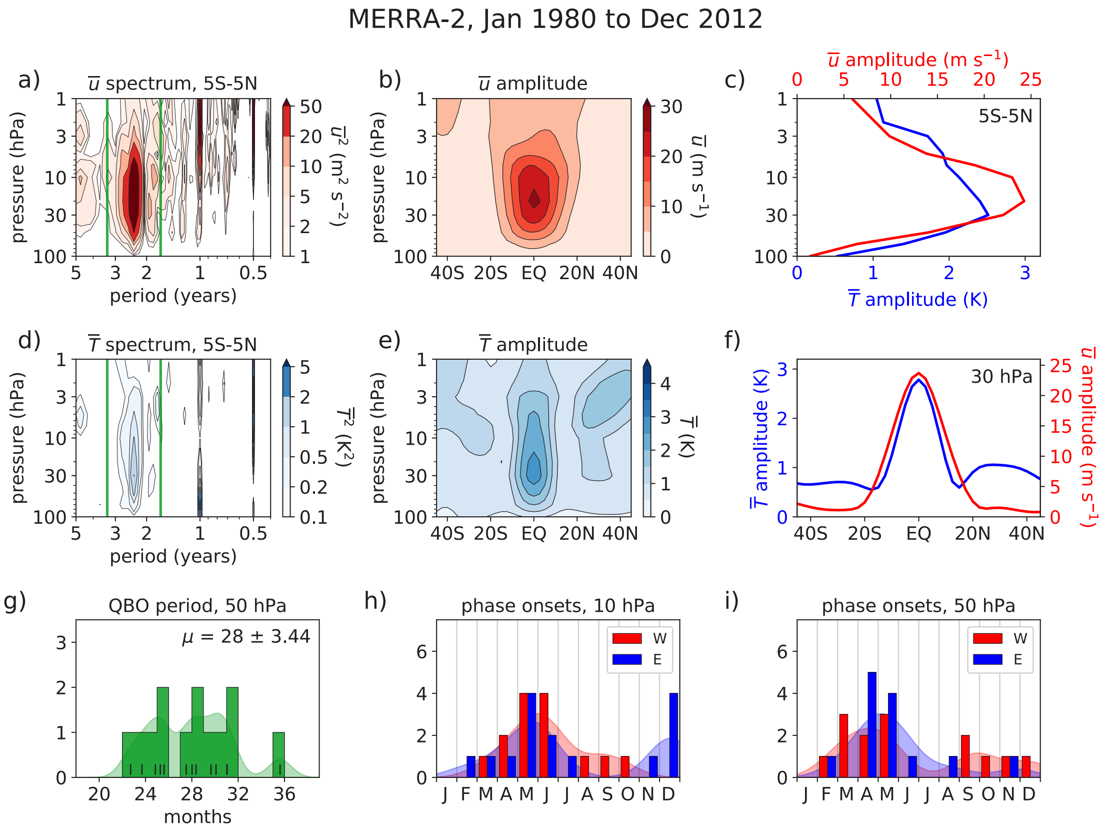


Figure AB9.7: *As* ***Figure 9.15****, but for MERRA-2 for the 1980 - 2012 period.*

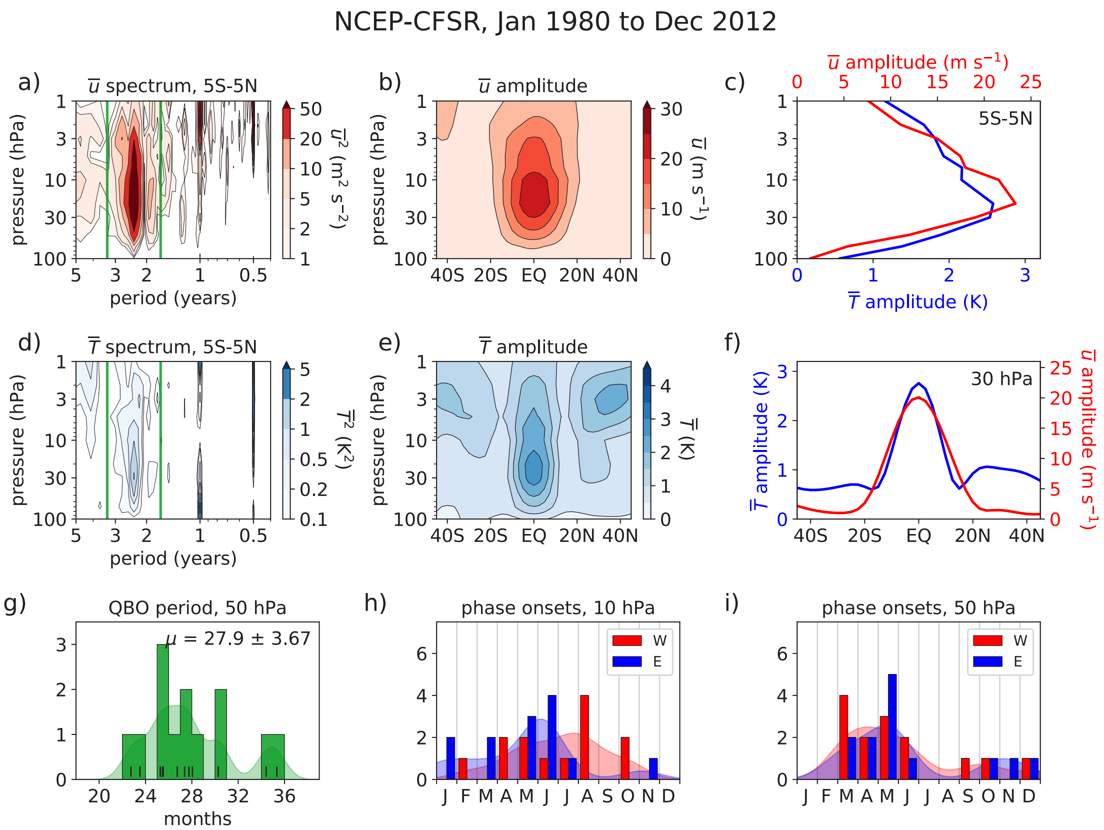


Figure AB9.8: *As* ***Figure 9.15****, but for CFSR for the 1980 - 2012 period.*

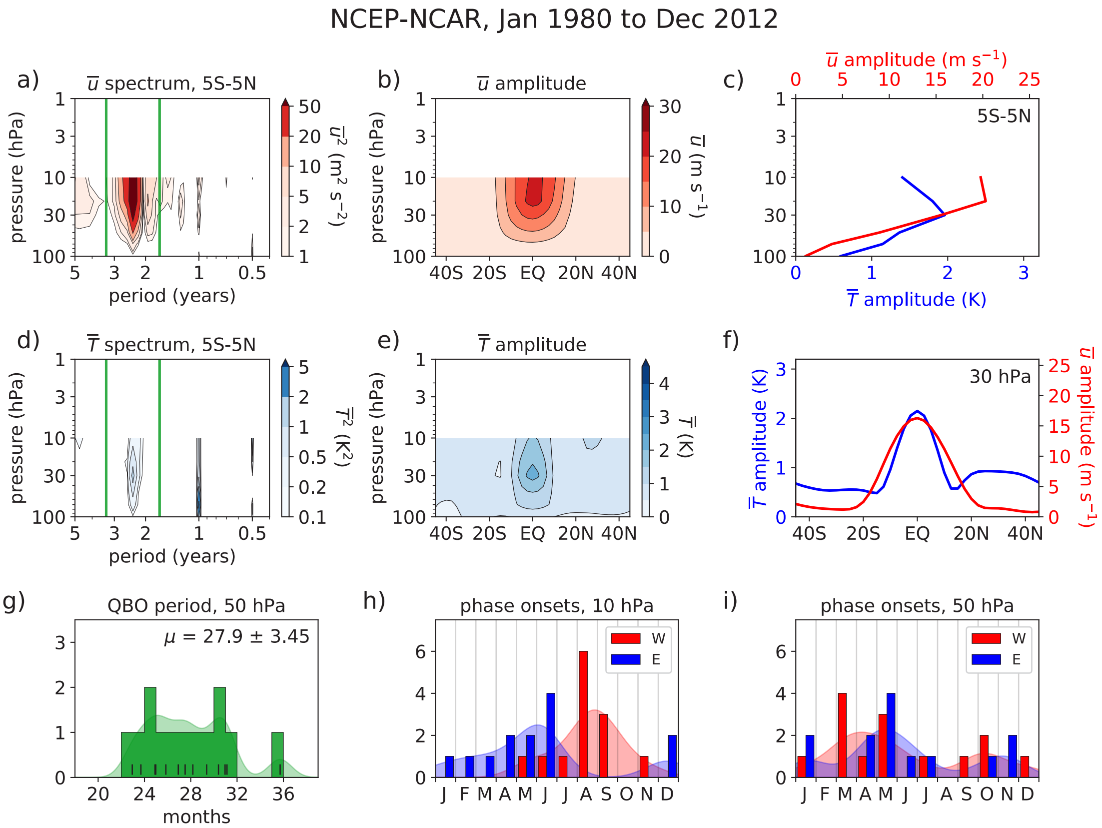


Figure AB9.9: *As* ***Figure 9.15****, but for NCEP-NCAR (R1) for the 1980 - 2012 period.*

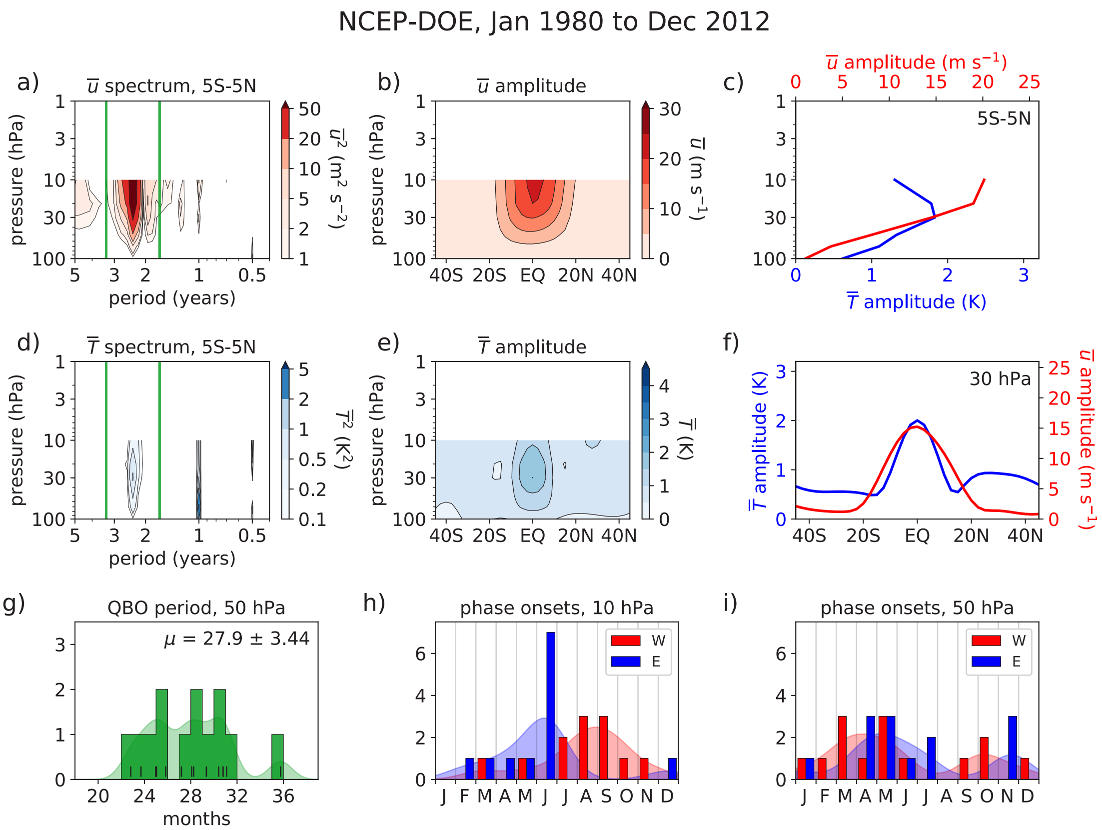


Figure AB9.10: *As* ***Figure 9.15****, but for NCEP-DOE (R2) for the 1980–2012 period.*

Major abbreviations and terms

ADV	Advection
AGCM	Atmospheric General Circulation Model
AMSU	Advanced Microwave Sounding Unit
ATOVS	Advanced TIROS Operational Vertical Sounder
COR	Coriolis force
COSMIC	Constellation Observing System for Meteorology, Ionosphere and Climate
CFSR	Climate Forecast System Reanalysis of the NCEP
CFSv2	Climate Forecast System version 2
DJF	December-January-February
DOE	Department of Energy
ECMWF	European Centre for Medium-range Weather Forecasts
ENSO	El Niño-Southern Oscillation
EOF	Empirical Orthogonal Function
EPD	Eliassen-Palm flux divergence
ERA-Interim	ECMWF interim reanalysis
ERA5	the fifth major global reanalysis produced by ECMWF
E-W	Easterly-to-westerly
FUB	Freie Universität Berlin
GNSS-RO	Global Navigation Satellite System Radio Occultation
GPCC	Global Precipitation Climatology Centre
GWD	Gravity wave drag
GWPE	Gravity wave potential energy
HIRDLS	High Resolution Dynamics Limb Sounder
IG	Inertio-gravity wave
IGRA	Integrated Global Radiosonde Archive
ITCZ	Intertropical Convergence Zone
JRA-25	Japanese 25-year Reanalysis
JRA-55	Japanese 55-year Reanalysis
JRA-55C	Japanese 55-year Reanalysis assimilating Conventional observations only
KW	Kelvin wave
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2
MF	Momentum flux
MLR	Multiple Linear Regression
MLS	Microwave Limb Sounder
MRG	Mixed Rossby-gravity wave
MSLP	Mean sea level pressure
NAO	North Atlantic Oscillation
NCEP	National Centers for Environmental Prediction of the NOAA
NCEP-DOE	Reanalysis of the NCEP and DOE
NCEP-NCAR	Reanalysis of the NCEP and NCAR
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
QBO E/W	Quasi-biennial oscillation easterly/westerly phase

REM	Reanalysis ensemble mean
RMS	Root mean square
RO	Radio occultation
RW	Rossby wave
SABER	Sounding of the Atmosphere using Broadband Emission Radiometry
SAO E/W	Semi-annual oscillation easterly/westerly phase
SD	Standard deviation
SH	Southern Hemisphere
SPARC	Stratosphere-troposphere Processes And their Role in Climate
S-RIP	SPARC Reanalysis Intercomparison Project
TIROS	Television Infrared Observation Satellite
TOVS	TIROS Operational Vertical Sounder
TTL	Tropical Tropopause Layer
W-E	Westerly-to-easterly

Chapter 10: Polar Processes

Chapter lead authors

Michelle L. Santee	NASA Jet Propulsion Laboratory, California Institute of Technology	USA
Alyn Lambert	NASA Jet Propulsion Laboratory, California Institute of Technology	USA
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA

Co-authors

Zachary D. Lawrence	(1) Cooperative Institute for Research in Environmental Sciences (CIRES), Univ. of Colorado (2) NOAA Physical Sciences Laboratory (3) NorthWest Research Associates	USA
Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Lars Hoffmann	Forschungszentrum Jülich GmbH	Germany
Sean P. Palmer	Department of Physics, New Mexico Institute of Mining and Technology	USA
Ken Minschwaner	Department of Physics, New Mexico Institute of Mining and Technology	USA

Abstract. This chapter focuses on microphysical and chemical processes in the winter polar lower stratosphere, such as polar stratospheric cloud (PSC) formation; denitrification and dehydration; heterogeneous chlorine activation and deactivation; and chemical ozone loss. These are "threshold" phenomena that depend critically on meteorological conditions. A range of diagnostics is examined to quantify differences between reanalyses and their impact on polar processing studies, including minimum lower stratospheric temperatures; area and volume of stratospheric air cold enough to support PSC formation; maximum latitudinal gradients in potential vorticity (a measure of the strength of the winter polar vortex); area of the vortex exposed to sunlight each day; vortex break-up dates; and polar cap average diabatic heating rates. For such diagnostics, the degree of agreement between reanalyses is an important direct indicator of the systems' inherent uncertainties, and comparisons to independent measurements are frequently not feasible. For other diagnostics, however, comparisons with atmospheric observations are very valuable. The representation of small-scale temperature and horizontal wind fluctuations and the fidelity of Lagrangian trajectory calculations are evaluated using observations obtained during long-duration superpressure balloon flights launched from Antarctica. Comparisons with satellite measurements of various trace gases and PSCs are made to assess the thermodynamic consistency between reanalysis temperatures and theoretical PSC equilibrium curves. Finally, to explore how the spatially and temporally varying differences between reanalyses interact to affect the conclusions of typical polar processing studies, simulated fields of nitric acid, water vapour, several chlorine species, nitrous oxide, and ozone from a chemistry-transport model driven by the different reanalyses for specific Arctic and Antarctic winters are compared to satellite measurements.

Contents

10.1 Introduction 493
10.2 Description of atmospheric measurements 495
10.2.1 Aura Microwave Limb Sounder 495
10.2.2 Envisat MIPAS 496
10.2.3 CALIPSO CALIOP 496
10.2.4 COSMIC GNSS-RO 496
10.2.5 Concordiasi superpressure balloon measurements 497
10.3 Overview of reanalysis polar temperature differences 498
10.4 Polar temperature and vortex diagnostics 498
10.5 Polar diabatic heating rates 502
10.6 Concordiasi superpressure balloon comparisons 505
10.6.1 Temperatures and winds 505
10.6.2 Trajectories 506
10.7 PSC Thermodynamic-consistency diagnostics 507
10.7.1 Mountain waves 507
10.7.2 Thermodynamic equilibrium 508
10.7.3 Temperature difference profiles 509
10.7.4 Summary of the temperature differences 510
10.8 Chemical modeling diagnostics 510
10.8.1 Details of the BASCOE system and experimental setup 511
10.8.2 Analysis approach 512
10.8.3 Case study: 2009 Antarctic winter 512
10.8.4 Case study: 2009/2010 Arctic winter 519
10.8.5 Comparisons of chemical ozone loss 520
10.8.6 Discussion and implications 521
10.9 Summary, key findings, and recommendations 522
References 526
Major abbreviations and terms 531

10.1 Introduction

One of the main research themes in atmospheric science over the past three decades has been the investigation of the chemical and dynamical processes involved in stratospheric ozone depletion, the most severe manifestation of which is the Antarctic ozone hole. In general, the processes controlling polar stratospheric ozone are now well understood (*e.g.*, *WMO*, 2018). In the very cold conditions that prevail inside the lower stratospheric winter polar vortices, water vapour (H_2O) and nitric acid (HNO_3) condense to form polar stratospheric clouds (PSCs). PSC particles and cold sulphate aerosols provide surfaces on which heterogeneous reactions can take place very rapidly, converting chlorine from relatively benign reservoir species such as hydrogen chloride (HCl) and chlorine nitrate ($ClONO_2$) into highly reactive ozone-destroying forms such as chlorine monoxide (ClO). Moreover, sequestration in PSCs substantially reduces gas-phase HNO_3 concentrations, and if solid HNO_3-containing PSC particles grow large enough to undergo appreciable gravitational sedimentation, then HNO_3 can be irreversibly removed from the stratosphere in a process known as denitrification. Similarly, sedimentation of water ice particles leads to dehydration. Severe denitrification and dehydration routinely occur in the cold, isolated Antarctic vortex. Compared to the Antarctic, the Arctic vortex is usually substantially warmer, more dynamically disturbed, smaller, and shorter lived, and thus in a typical year it experiences little or no denitrification or dehydration. Chlorine activation is also typically less intense, extensive, and prolonged in the Arctic than in the Antarctic. Consequently, although the same fundamental processes are at work in the lower stratosphere in both polar regions, in most years chlorine-catalyzed ozone loss is considerably weaker in the Arctic than in the Antarctic.

Lower stratospheric polar processes and chemical ozone loss are "threshold" phenomena that depend critically on stratospheric temperatures and other meteorological and dynamical factors (*e.g.*, winter polar vortex characteristics, breakup dates, *etc.*). Several studies over the years have explored the temperature sensitivity of these processes; for example, *Wegner et al.* (2012) showed that heterogeneous reaction rates on liquid aerosols are doubled for every 1 K in cooling and increase tenfold over a 2-K range around 192 K, and *Solomon et al.* (2015, see also references therein) showed that a 2-K perturbation in temperature applied to heterogeneous chemical reactivities and PSC surface area in a specified-dynamics chemistry climate model induces a change in simulated Arctic column ozone loss of ~ 40 DU.

As in many other Earth system science specialties, atmospheric polar processing studies often rely heavily on global meteorological data sets. Thus it is essential to understand the accuracy and reliability of reanalysis fields in a polar processing context. Differences between reanalyses are likely to have the largest impact on such studies when conditions are marginal, *i.e.*, in the Arctic (in most years) and in the autumn and spring in the Antarctic. As noted, for example, by *Hoffmann et al.* (2017a) and *Lambert and Santee* (2018) and discussed further below, in addition to discrepancies in physical parameters (*e.g.*, temperature, winds) between the various reanalyses, differences in their temporal and/or spatial resolution may also play a role in detailed quantitative studies.

Given the importance of stratospheric temperatures, transport, and mixing for ozone chemistry, a number of studies over the last twenty years have assessed the representativeness of meteorological analyses and reanalyses. We briefly summarize here several studies that carried out comparisons of two or more analyses/reanalyses specifically in a stratospheric polar processing framework. In one of the earliest such studies, *Manney et al.* (1996) examined temperatures, geopotential heights, winds, and potential vorticity (PV) calculated from stratospheric analyses provided by the (then) UK Meteorological Office (UKMO) and the US National Meteorological Center (NMC) in both hemispheres during dynamically active periods, when substantial discrepancies between analyses were likely to be seen. Although both analyses captured the qualitative features and evolution of the large-scale winter stratospheric circulation, differences in their temperatures and polar vortex characteristics implied significant effects on quantitative process studies, especially for the Southern Hemisphere. *Knudsen* (1996) also found substantial biases between observed lower stratospheric temperatures and analyses from UKMO and the European Centre for Medium-range Weather Forecasts (ECMWF). *Knudsen et al.* (2001) assessed the accuracy of analyzed winds from ECMWF, UKMO, and the US National Centers for Environmental Prediction (NCEP) Climate Prediction Center (CPC) by comparing calculated air parcel trajectories based on those analyses with long-duration balloon flights in the Arctic stratospheric vortex. Similarly, *Knudsen et al.* (2002) used independent meteorological measurements from long-duration balloon flights in the Arctic stratospheric vortex to quantify errors in five sets of analyzed temperatures: ECMWF, Met Office (formerly UKMO), the Goddard Space Flight Center Data Assimilation Office (DAO), NCEP/CPC, and NCEP/National Center for Atmospheric Research reanalysis (NCEP-NCAR R1); although some of the analyses showed larger scatter around the balloon values than others, occasional large differences occurred in all of them, particularly during a major sudden stratospheric warming. *Manney et al.* (2003b) compared commonly used meteorological analyses (Met Office, NCEP/CPC, NCEP-NCAR R1, ECMWF, DAO) during two cold Arctic winters, examining not only temperatures (average and minimum values, number of cold days, *etc.*) but also temperature histories along trajectories to assess simulated PSC lifetimes and the overall potential for chlorine activation. They found that discrepancies between analyses arise from differences in both the magnitude and the morphology of wind and temperature fields, such that dissimilarities in dynamical conditions in comparably cold winters may strongly influence the degree of agreement between meteorological data sets.

Following on from that study, *Manney et al.* (2005) investigated an extensive set of diagnostics related to lower stratospheric chemistry, transport, and mixing during the 2002 Antarctic winter, when unusual dynamical activity may have exacerbated the disagreement between meteorological data sets. Comparing four operational products (Met Office, ECMWF, NCEP/CPC, and the NASA Global Modeling and Assimilation Office (GMAO) Goddard Earth Observing System (GEOS-4)), as well as the 40-year reanalysis from ECMWF (ERA-40), NCEP-NCAR R1, and a second NCEP/Department of Energy reanalysis (NCEP-DOE R2), they again found considerable differences; such large disparities undermine confidence in the results from scientific studies based on any of those analyses/reanalyses. In particular, NCEP-NCAR R1, NCEP-DOE R2, and ERA-40 were shown to suffer from substantial deficiencies in their depiction of the magnitude, structure, or evolution of temperatures and/or winds that rendered them unsuitable for detailed studies of lower stratospheric polar processing. *Labitzke and Kunze* (2005) compared stratospheric temperatures over the Arctic from NCEP-NCAR R1 and ERA-40 with an independent data set (historical daily analyses of Northern Hemisphere temperature fields over 100 - 10 hPa produced by hand at FU Berlin); although agreement in the long-term mean temperatures and the trends (1957 - 2001) was generally good, they also found unrealistic behavior in ERA-40, which displayed larger biases in the October to January interval after 1979. *Tilmes et al.* (2006) focused specifically on the volume of air below the temperature threshold for PSC existence (V_{PSC}); they found that, although the general patterns of V_{PSC} evolution were similar for Met Office and ERA-40 reanalyses as well as ECMWF operational analyses and data from FU-Berlin, differences between the two reanalyses were as large as 10 % during their period of overlap (1991 - 1999). *Rieder and Polvani* (2013) also touched on comparisons of V_{PSC}, computing it using temperatures from MERRA, ECMWF Interim Reanalysis (ERA-Interim), and NCEP-NCAR R1. Although the depiction of year-to-year variability was seen to be strongly correlated among the three reanalyses, the magnitude of V_{PSC} varied considerably, with MERRA and ERA-Interim indicating V_{PSC} values roughly 30 % larger than those from NCEP-NCAR R1. Rieder and Polvani reiterated the cautions raised earlier about using NCEP-NCAR R1 in detailed polar processing studies.

A few studies have looked at the impact of differences in meteorological fields on results from chemical transport models (CTMs). *Davies et al.* (2003) investigated the effects of denitrification on ozone depletion in a cold Arctic winter by forcing a 3D CTM incorporating different PSC schemes with both UKMO and ECMWF analyses, finding that the two meteorological data sets led to disparate patterns of modeled PSC formation and denitrification, and consequently also chlorine activation and ozone loss. Similarly, *Feng et al.* (2005) applied the same CTM and analyses to examine the evolution of the ozone hole during the highly disturbed Antarctic winter of 2002; although runs driven by both analyses reproduced the anomalous conditions in that winter, differences in the structure and magnitude of simulated total ozone were seen.

Lawrence et al. (2015) revisited the use of polar processing diagnostics to evaluate reanalyses, performing a comprehensive intercomparison of NASA's Modern Era Retrospective-analysis for Research and Applications (MERRA) and ERA-Interim over the period 1979 to 2013. Agreement between the two meteorological data sets changed substantially during this interval, with many stratospheric temperature and vortex characteristics converging to greater consistency over time as more high-quality observations were assimilated. *Lawrence et al.* (2015) concluded that for the years since 2002 the MERRA and ERA-Interim reanalyses are equally appropriate choices and either can be used with confidence in polar processing studies in both hemispheres. In a follow-up study, *Lawrence et al.* (2018) extended the application of polar processing diagnostics to encompass other current full-input reanalyses, including MERRA-2, the Japanese 55-year Reanalysis (JRA-55), and the NCEP Climate Forecast System Reanalysis / Climate Forecast System, version 2 (CFSR/CFSv2). Results from this later study are described in detail in *Section 10.4*.

In summary, several previous studies have found considerable discrepancies between meteorological analyses/reanalyses in various parameters of relevance for polar processing, revealing that significant quantitative and qualitative differences may arise from the choice of which meteorological products are used in a given study. The recent work of *Lawrence et al.* (2015, 2018) indicates that agreement among various modern reanalyses improved substantially for some polar processing diagnostics in the post-2001 timeframe, following the introduction of new data streams. Nevertheless, previous studies have not examined all reanalyses of interest for S-RIP; moreover, a comparison of metrics not explored in earlier papers would be informative. Thus a comprehensive reassessment is warranted.

In this chapter we intercompare recent full-input reanalyses using an extensive set of polar processing diagnostics. The specific reanalyses considered here are: MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2. *Fujiwara et al.* (2017) provide an overview of these reanalysis systems, and they are also described in detail in *Chapter 2* of this Report. We note that ECMWF stopped producing ERA-Interim in August 2019 and replaced it with ERA5. Because the bulk of the analysis for this chapter had already been completed by the time ERA5 became available, its performance has not been assessed here. Although we expect that ERA5 will prove to be at least as reliable for polar processing studies as other modern reanalyses, we can make no conclusive judgments about its suitability for such studies at this time.

The primary focus here is on reanalyses; however, in some cases where comparisons with atmospheric observations are made we also examine the ECMWF operational analysis (OA) and the NASA GMAO Goddard Earth Observing System Version 5.9.1 (GEOS-591) assimilation product. The ECMWF OA evaluated by *Hoffmann et al.* (2017a), whose results are summarized in *Section 10.6*, is characterized by 3-hr temporal resolution, 0.125°×0.125° horizontal resolution, and 91 vertical levels with an upper lid at 0.01 hPa. The GEOS-591 near-real-time analysis, which was produced by the GEOS-5 data assimilation system (*Molod et al.*, 2015; *Rienecker et al.*, 2011), was characterized by 3-hr temporal resolution, 0.625°×0.5° horizontal resolution, and 72 vertical levels with an upper lid at 0.01 hPa. This stable system, used by NASA Earth Observing System satellite instrument teams in their data processing, provided consistent meteorological fields over much of the Aura record and was thus somewhat akin to a reanalysis. It was assessed by *Lambert and Santee* (2018), whose results are summarized in *Section 10.7.*

Much of this chapter focuses on process-oriented and case studies. For many diagnostics, the degree of agreement between reanalyses is an important direct indicator of the systems' inherent uncertainties, for which comparisons to independent measurements are not required. In addition, some diagnostics are based on PV or other dynamical quantities that cannot be provided directly by any measurement system. These situations pertain to many of the polar temperature and vortex diagnostics presented in *Section 10.4*, including minimum lower stratospheric temperature, area and volume of stratospheric air with temperatures below PSC existence thresholds, maximum latitudinal gradients in PV (a measure of the strength of the winter polar vortex), area of the vortex exposed to sunlight each day, and vortex breakup dates, as well as the polar cap average diabatic heating rates discussed in *Section 10.5*. On the other hand, comparisons with atmospheric measurements can be made for some diagnostics, especially the more derived ones. Such comparisons typically demand fairly broad spatial coverage on a daily basis, which is best afforded by satellite measurements. For the most part, comparisons between reanalysis fields and independent observations are left to *Chapter 3*; *Long et al.* (2017) also presented comparisons of reanalysis temperatures against satellite observations. However, analyses/reanalyses are evaluated through comparisons with long-duration superpressure balloon temperature and wind measurements in *Section 10.6*. In addition, the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) global navigation satellite system (GNSS) radio occultation (RO) temperatures are examined in connection with PSC thermodynamic-consistency diagnostics in *Section 10.7*. The latter section also relies on vertical profiles of gas-phase HNO_3 and H_2O measured by the Aura Microwave Limb Sounder (MLS), as well as PSC characteristics determined from Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) lidar aerosol and cloud backscatter. Finally, because results from chemical models synthesize the interplay among the spatially and temporally varying differences between reanalyses and exemplify how their net effects impact the bottom-line conclusions of typical real-life studies, in *Section 10.8* we compare simulated sequestration of HNO_3 and H_2O in PSCs, chlorine activation, and ozone fields with those observed by Aura MLS and the Envisat Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) for one winter in each hemisphere. Model-based estimates of Antarctic chemical ozone loss in the stratospheric partial column are also compared with those derived from MLS data. Other commonly used ozone loss metrics such as ozone hole area, ozone mass deficit, *etc*., are not included here, nor are other processes that affect polar ozone but that are covered extensively elsewhere in this Report (*e.g.*, sudden stratospheric warmings are discussed in *Chapter 6*). Further direct comparisons between observations and the ozone fields from the reanalyses can be found in *Chapter 4*.

10.2 Description of atmospheric measurements

10.2.1 Aura Microwave Limb Sounder

MLS measures millimeter- and submillimeter-wavelength thermal emission from the limb of Earth's atmosphere (*Waters et al*., 2006). The Aura MLS field-of-view (FOV) points in the direction of orbital motion and vertically scans the limb in the orbit plane, providing data coverage from 82°S to 82°N latitude on every orbit. Because the Aura orbit is sun-synchronous (with a 13:45 local time ascending equator-crossing time), MLS observations at a given latitude on either the ascending (mainly day) or descending (mainly night) portions of the orbit have the same local solar time. Northern high latitudes are sampled by ascending measurements near midday local time, whereas southern high latitudes are sampled by ascending measurements in the late afternoon. Vertical profiles are measured every ~165 km along the suborbital track, yielding a total of ~3500 profiles per day.

Here, we use the MLS version 4.2 (v4.2) data (*Livesey et al.*, 2020). Detailed information on the quality of a previous version of MLS data, v2.2, can be found in dedicated validation papers by *Lambert et al*. (2007) for stratospheric H_2O, *Santee et al.* (2007) for HNO_3, *Santee et al.* (2008) for ClO, *Froidevaux et al.* (2008a) for HCl, *Froidevaux et al.* (2008b) for stratospheric O_3, and *Schwartz et al.* (2008) for temperature. The precision, resolution, and useful vertical range of the v4.2 measurements, as well as assessments of their accuracy through systematic error quantification (and, in some cases, validation comparisons with correlative data sets), are reported for each species by *Livesey et al.* (2020). Briefly, MLS measurements have single-profile precisions (accuracies) of 4 - 15% (4 - 20%) for H_2O, 0.6 ppbv (1 - 2 ppbv) for HNO_3, 0.1 ppbv (0.05 - 0.25 ppbv) for ClO, 0.2 - 0.3 ppbv (0.2 ppbv) for HCl, 0.05 - 0.1 ppmv (0.1 - 0.25 ppmv) for O_3, and 0.6 - 1.2 K (0 - 5 K) for temperature in the stratosphere.

We note that MERRA-2 assimilates MLS temperatures, but only at pressures less than 5 hPa and not within the pressure range investigated here (*Gelaro et al.*, 2017).

Errors in the MLS H_2O contribute a few tenths of a kelvin to the error in calculated frost point temperatures and are substantially smaller than the errors in the temperature limb sounding retrievals obtained from MLS. From August 2004 until December 2013, mean differences between NOAA frost point hygrometer and MLS H_2O data showed no statistically significant differences (agreement to better than < 1 %) from 68 - 26 hPa, although significant biases at 100 hPa and 83 hPa were found to be 10 % and 2 %, respectively (*Hurst et al.*, 2014). However, increasing the time frame to mid-2015 revealed a long-term drift in MLS H_2O of up to 1.5 % per year starting around 2010 (*Hurst et al.*, 2016). Although changes to the MLS data processing system have substantially mitigated this drift in the version 5 MLS H_2O measurements (*Livesey et al.*, 2021), for the v4 data used here the effect on the calculated supercooled ternary solution (STS) reference and frost point temperatures is less than 0.1 K per year.

To aid in the analysis of MLS measurements, particularly in *Section 10.7,* we make use of MLS Derived Meteorological Products (DMPs). These files contain meteorological data (*e.g.*, temperature) and derived parameters (*e.g.*, equivalent latitude) interpolated from gridded reanalysis fields to the along-track geolocations of the MLS measurements. The original version of the MLS DMPs was described in detail by *Manney et al.* (2007). Here we use updated files (version 2, the DMP version of record for the MLS v4.2 data; see the MLS web page, **http://mls.jpl.nasa.gov**, for more details); the v2 DMPs are from the software described by *Manney et al.* (2011a). DMP files containing associated meteorological information at the MLS measurement locations have been produced for all five full-input reanalyses considered here.

10.2.2 Envisat MIPAS

The MIPAS instrument (*Fischer et al.*, 2008) was launched in March 2002 on the ESA Environment Satellite (Envisat) and was operational until April 2012. MIPAS was an infrared Fourier transform spectrometer for measuring limb emission spectra between 685 cm^{-1} and 2410 cm^{-1} (14.6 - 4.15 μm). Through azimuth scanning it provided global coverage from 87.5 °S to 89.3 °N. The instrument FOV was 30 km across-track and 3 km in the vertical, and the horizontal along-track sampling distance for nominal-mode observations was ~ 530 km from 2002 to 2004 and ~ 400 km from 2005 onward. Several retrieval algorithms have been developed for the MIPAS spectra; here we use profiles of temperature and atmospheric constituents generated by the KIT-IMF-ASF (Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Atmospheric Trace Gases and Remote Sensing) group in cooperation with the Instituto de Astrofísica de Andalucía (*von Clarmann et al.*, 2009). For $ClONO_2$ below 40 km, precision is 8 - 14 %, with vertical resolution 2.5 - 9 km (*Höpfner et al.*, 2007; *von Clarmann et al.*, 2009). Retrieval of $ClONO_2$ is hindered by the presence of optically thick PSCs along the MIPAS line of sight.

10.2.3 CALIPSO CALIOP

The CALIOP dual-wavelength elastic backscatter lidar (*Winker et al.*, 2009) flies on the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite launched in April 2006. We use the CALIOP Level 2 operational data set L2PSCMask (v1 Polar Stratospheric Cloud Mask Product) produced by the CALIPSO science team. The Level 2 operational data consist of nighttime-only data and contain profiles of PSC presence, composition, optical properties, and meteorological information along the CALIPSO orbit tracks at a horizontal resolution of 5 km and a vertical resolution of 180 m. We have applied post-processing to generate coarser horizontal/vertical bins for a better comparison at the scale of the MLS along-track and vertical resolution (see *Section 10.7* for details). Each averaging bin is the size of the MLS along-track vertical profile separation (165 km) and the height between the mid-points of the retrieval pressure levels (2.16 km) for the MLS HNO3 data product. This we refer to as the MLS geometric FOV. There are approximately four hundred 5 km × 180 m CALIOP “pixels” within the MLS geometric FOV.

The CALIOP PSC classification scheme used here is described by *Pitts et al.* (2009), with modifications discussed by *Pitts et al.* (2013), and consists of four main PSC types. MIX1 and MIX2 denote detections of nitric acid trihydrate (NAT) particles, with the MIX1/MIX2 boundary marking a transition between lower (MIX1) and higher (MIX2) NAT number/volume densities. The STS type indicates supercooled liquid ternary solution ($H_2SO_4/HNO_3/H_2O$) particles, and ICE indicates water-ice particles.

10.2.4 COSMIC GNSS-RO

We use the US/Taiwan Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) network data obtained from the Universities for Cooperative Atmospheric Research (UCAR) COSMIC Data Analysis and Archive Center (CDAAC). Global navigation satellite system radio occultation (GNSS-RO) data have provided high accuracy (bias < 0.2 K and precision > 0.7 K, *Gobiet et al.*, 2007), global (day and night) coverage, coupled with excellent long term stability, for nearly two decades (*Anthes*, 2011). The vertical resolution is better than about 0.6 km over the 15 - 30 km vertical range considered here. The introduction of GNSS-RO has been documented to improve numerical weather prediction (NWP) forecast skill in the ECMWF Integrated Forecast System (IFS) (*Bonavita*, 2014) and to reduce tropopause and lower stratospheric temperature biases in ERA-Interim (*Poli et al.*, 2010). The direct assimilation of bending angles or refractivity is now the common practice for many global reanalyses; however, for many other purposes the production of vertical atmospheric temperature profiles from GNSS-RO data is required. The retrieval of vertical atmospheric geophysical profiles from RO requires a number of assumptions because of the long ray path through a non-uniform atmosphere (*Ho et al.*, 2012). Therefore, corrections are required for ionospheric effects, variations in water vapour, and gradients in temperature along the ray path (*Anthes*, 2011; *Poli and Joiner*, 2004). Many other studies have intercompared GNSS-RO with independent operational analyses, *e.g.,* with forecast versions that have not assimilated the GNSS-RO data. The near real time COSMIC data (in the form of bending angles or refractivity) are ingested by most of the data assimilation procedures considered here (except for MERRA), and therefore these reanalyses are not strictly independent of the postprocessed COSMIC temperatures. We have chosen to use the COSMIC temperatures as a common reference to evaluate the reanalysis departures, rather than using the reanalysis ensemble mean.

10.2.5 Concordiasi superpressure balloon measurements

Superpressure balloons are aerostatic balloons, which are filled with a fixed amount of lifting gas, and for which the maximum volume of the balloon is kept constant by means of a closed, inextensible, spherical envelope. After launch, the balloons ascend and expand until they reach a float level where the atmospheric density matches the balloon density. On this isopycnic surface a balloon is free to float horizontally with the motion of the wind. Hence, superpressure balloons behave as quasi-Lagrangian tracers in the atmosphere. The Concordiasi field campaign in Antarctica in September 2010 to January 2011 was aimed at making innovative atmospheric observations to study the circulation and chemical species in the polar lower stratosphere and to reduce uncertainties in diverse fields in Antarctic science (*Rabier et al.*, 2010). During the field campaign, 19 superpressure balloons with 12 m diameter were launched from McMurdo Station (78 ° S, 166 ° E), Antarctica, by the French space agency, Centre National d'Etudes Spatiales (CNES). Balloons of this size typically drift at pressure levels of ~ 60 hPa and altitudes of ~ 18 km. The balloons were launched between 8 September and 26 October 2010, and each balloon flew in the mid- and high-latitude lower stratosphere for a typical period of 2 to 3 months.

The positions of the balloons were tracked every 60 s by means of global positioning satellite (GPS) receivers. At each observation time the components of the horizontal wind are computed by finite differences between the GPS positions. The uncertainty is about 1 m for the GPS horizontal position and 0.1 m s^{-1} for the derived winds (*Podglajen et al.*, 2014). Each balloon launched during Concordiasi was equipped with a meteorological payload called the Thermodynamical SENsor (TSEN). TSEN makes in situ measurements of atmospheric pressure and temperature every 30 s during the whole flight. The pressure is measured with an accuracy of 1 Pa and a precision of 0.1 Pa. The air temperature is measured via two thermistors. During daytime, the thermistors are heated by the sun, leading to daytime temperature measurements being warmer than the real air temperature. An empirical correction has been used to correct for this effect, which is described in detail by *Hertzog et al.* (2004). The precision of the corrected temperature observations is about 0.25 K during daytime and 0.1 K during nighttime.

Most of the measurements (*i.e.*, more than 90 %) took place between 25 September and 22 December 2010, at an altitude range of 17.0 - 18.5 km, and within a latitude range of 59 ° - 84 ° S. The pressure measurements are mostly within a range of 58.2 - 69.1 hPa and the temperature measurements within 189 - 227 K. The density of air, calculated from pressure and temperature, varies between 0.099 kg m^{-3} and 0.120 kg m^{-3}. The zonal winds are predominately westerly and mostly within a range of 1 - 44 m s^{-1}. The meridional wind distributions are nearly symmetric, with meridional winds being in the range of ± 17 m s^{-1}. Horizontal wind speeds are mostly within 5 - 47 m s^{-1}.

The Concordiasi balloon observations have been assimilated into the ECMWF, MERRA, and MERRA-2 data sets, but they were not considered for NCEP-NCAR R1. The observations therefore provide an independent data source only for the validation of the NCEP-NCAR R1 data set. However, as meteorological analyses are a result of combining various satellite and in situ observations, a forecast model, and a data assimilation procedure, a comparison of the meteorological data with the Concordiasi observations still provides information on the performance of the overall system, even for the reanalyses that assimilate those observations. As the observational data have been subject to downsampling and data thinning before they were assimilated, an assessment of the representation of small-scale structures due to gravity waves also remains meaningful.

Trajectory calculations for the Concordiasi balloon observations have been analyzed using the Lagrangian particle dispersion model Massive-Parallel Trajectory Calculations (MPTRAC) (*Hoffmann et al.*, 2016). Transport is simulated by calculating trajectories for large numbers of air parcels based on given wind fields from global meteorological reanalyses. The numerical accuracy and efficiency of trajectory calculations with MPTRAC was assessed by *Rößler et al.* (2018). Turbulent diffusion and subgrid-scale wind fluctuations are simulated based on the Langevin equation, closely following the approach implemented in the Flexible Particle

(FLEXPART) model (*Stohl et al.*, 2005).

10.3 Overview of reanalysis polar temperature differences

To provide context for later results derived from more complex analysis techniques, we show in **Figure 10.1** a basic overview of reanalysis temperatures in the polar lower stratosphere. We have chosen as a suitable metric the daily (12 UT) mean 60° polar cap temperature differences at 46 hPa. Time series of the temperature differences calculated for MERRA, MERRA-2, JRA-55, and CFSR/CFSv2 relative to ERA-Interim over 2008 - 2013 have been smoothed using a 10-day boxcar average. The daily mean standard error of the temperature differences is less than 0.1 K. In both hemispheres, temperature differences display annual cycles, with positive deviations mainly in summer and negative deviations mainly in winter. In the Antarctic the largest deviations are ~ 1 K in MERRA – ERA-Interim, whereas in the Arctic the largest deviations are in JRA-55 – ERA-Interim, but they only reach ~ 0.5 K.

The grey-shaded regions in **Figure 10.1** mark the useful wintertime measurement periods, chosen to capture the bulk of the PSC activity needed for the evaluation of the thermodynamic temperature comparisons that are discussed in *Section 10.7*. These periods also happen to largely coincide with times of smaller variability in the temperature differences, when biases of the other reanalyses with respect to ERA-Interim are predominantly negative. Therefore, we caution that intercomparisons of reanalyses undertaken using other time periods, especially for summertime, could even obtain temperature deviations of opposite sign whilst maintaining about the same magnitude. Indeed, whereas *Hoffmann et al.* (2017a) find MERRA to be the warmest and ERA-Interim the coldest compared to superpressure balloon temperature measurements made during the Antarctic Concordiasi campaign in September 2010 to January 2011, *Lambert and Santee* (2018) find the opposite order compared to the COSMIC and thermodynamic temperature references for May to August during 2008 to 2013. To reconcile this apparent discrepancy, in **Figure 10.2** daily mean temperature differences (at 12 UT) for MERRA and MERRA-2 relative to ERA-Interim are used to highlight the non-overlapping intervals of the PSC analysis window (green line) and the balloon flights (red line). Measurements in the later time period of the Concordiasi balloon flights (September - December) clearly sample different atmospheric conditions than those prevailing in the earlier time period (May - August). Moreover, differences between reanalysis temperatures along individual balloon trajectories are likely to be amplified compared to the differences in mean polar cap temperatures. We note that MERRA does not assimilate COSMIC data, whereas MERRA-2 and the other reanalyses investigated here do; hence some of the reduction in the bias of MERRA-2 compared to MERRA seen in **Figure 10.2** is likely attributable to the former's use of GNSS-RO data.

10.4 Polar temperature and vortex diagnostics

Lawrence et al. (2018) expanded on the diagnostics in *Lawrence et al.* (2015) and applied them to CFSR/CFSv2,

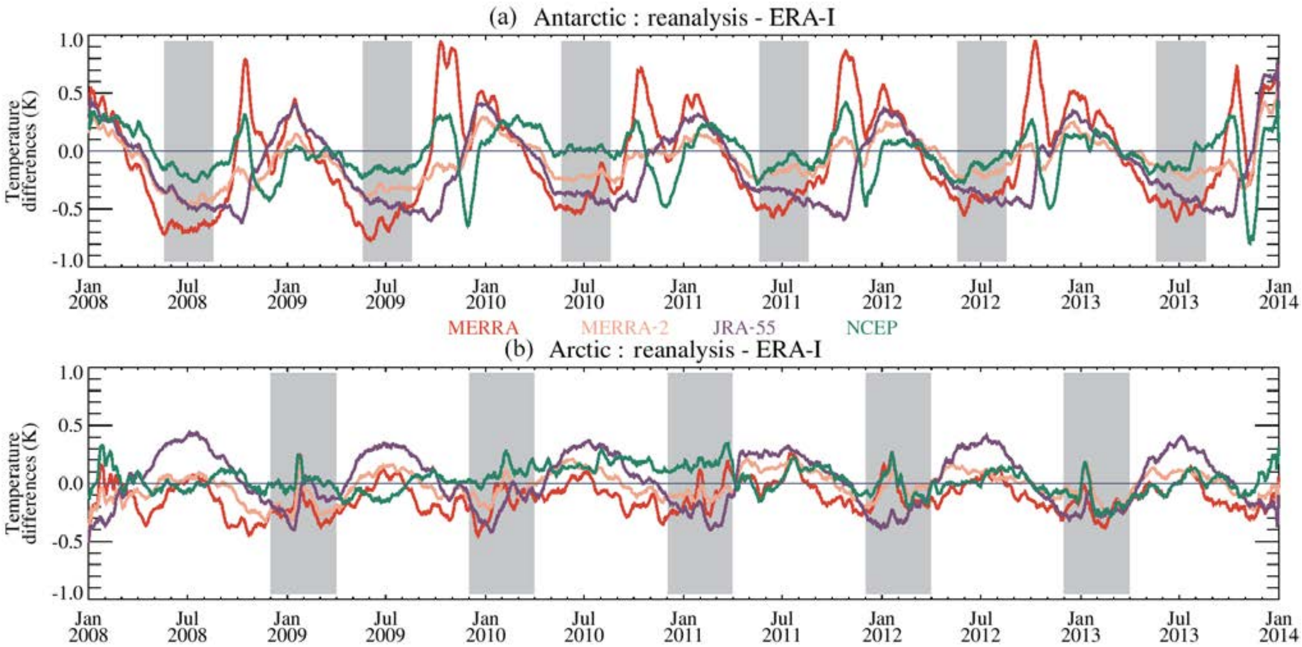


Figure 10.1: *(a) Time series (for 12 UT) from 2008 to 2013 of 10-day boxcar-smoothed temperature differences for MERRA, MERRA-2, JRA-55, and CFSR/CFSv2 relative to ERA-Interim at 46 hPa, averaged over the 60° Antarctic polar cap. The four reanalyses being differenced against ERA-Interim are shown in the colors indicated in the legend between the two panels. Grey regions indicate the periods defined for the analysis of PSC-related metrics (see Section 10.7). (b) Same, but for the Arctic. From Lambert and Santee (2018).*

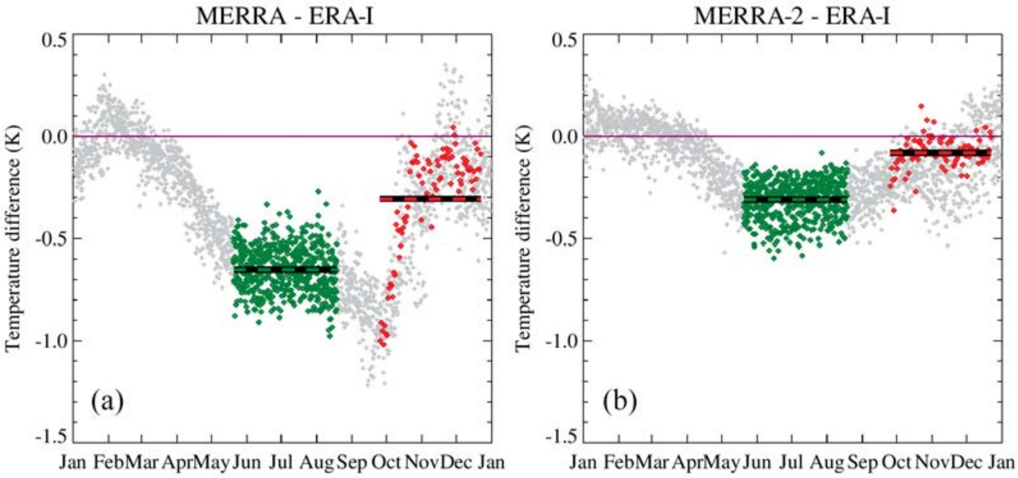


Figure 10.2: *Daily mean temperature differences (grey) of (a) MERRA and (b) MERRA-2, relative to ERA-Interim at 62 hPa (representative of the Concordiasi balloon float heights (see Section 10.6) in the 60° Antarctic polar cap for 2008-2013. The green-black dashed line indicates the mean of the 2008-2013 differences (green symbols) during the PSC analysis window. The red-black dashed line indicates the mean of the differences (red symbols) over the time span of 90% of the Concordiasi balloon measurements in 2010 (Hoffmann et al., 2017a). From Lambert and Santee (2018).*

ERA-Interim, JRA-55, and MERRA-2 (evaluations were also done for MERRA but were not included in the paper). The full suite of diagnostics examined by *Lawrence et al.* (2018) includes minimum temperatures poleward of ±40° latitude, the area of temperatures (poleward of ±30° latitude) below PSC existence thresholds, the winter mean volume of lower stratospheric air with temperatures below PSC existence thresholds, maximum gradients in scaled PV as a function of equivalent latitude, the area of the vortex exposed to sunlight, and approximate dates of the breakup of the polar vortices. As noted by *Lawrence et al.* (2018), reanalysis comparisons are particularly critical to assess the uncertainties in these types of diagnostics because they are not quantities that can be compared directly with observations. The key results of *Lawrence et al.* (2018) are summarized below.

Figure 10.3 compares Southern Hemisphere (SH) extended winter season (MJJASO) minimum temperatures poleward of 40°S in the lower stratosphere from MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 with those from the reanalysis ensemble mean (REM). Since the REM values that go into each season's mean vary from day to day and are expected to change with any large change in any of the reanalyses, the differences from the REM quantify only how far each reanalysis is from that mean during each season, thus giving an idea of the range of values (which could be interpreted as an uncertainty in the diagnostic) but not of the absolute changes in those values. The standard deviations shown on the right of **Figure 10.3** help further quantify the spread among the reanalyses. The reanalyses converge towards much better agreement in the later years at all levels. There are step-like changes in agreement among the reanalyses around 1998 (especially ERA-Interim and MERRA-2), when the reanalyses (albeit not all at exactly the same time) changed from assimilating Tiros Operational Vertical Sounder (TOVS) to advanced TOVS (ATOVS) radiances; the latter provide higher-resolution constraints on stratospheric temperatures. On average, the individual reanalyses agree with the REM to within about 0.5 K in the most recent decade, though differences in the early years commonly exceed 3 K. The standard deviations of the differences increase with altitude (indicating larger maximum differences between the reanalyses) and show a modest decrease over the time period, with an abrupt decrease seen around 1998 in most reanalyses. CFSR/CFSv2 generally shows larger variance than the other reanalyses in the period since about 2002, and less of a change around 1998. These features are consistent with those from comparisons of other polar processing diagnostics. In the Northern Hemisphere (NH), the reanalyses agree much better throughout the 36 years (to within about 1.5 K before 1999), though convergence toward better agreement is also seen in most of the reanalyses (excepting CFSR/CFSv2) (*Lawrence et al.*, 2018); this is not unexpected since the input data density (from, *e.g.*, sondes and ground-based measurements) in the years before the TOVS/ATOVS transition was much greater in the NH, and thus the reanalyses' temperatures before that transition were much better constrained than those in the SH. *Lawrence et al.* (2018) also evaluated the area of temperatures below PSC thresholds, with results consistent with those shown here for minimum temperatures.

Figure 10.4 shows differences from the REM of daily maximum PV gradients (a measure of vortex strength) with respect to equivalent latitude averaged over the DJFM season in the NH, as well as the standard deviations of the daily differences. The climatological maximum values of this diagnostic increase with height from around $1-6\times10^{-6}$ s^{-1} deg^{-1} at about 430 K to over 20×10^{-6} s^{-1} deg^{-1} above about 600 K (*Lawrence et al.*, 2018). There is no obvious systematic decrease in the differences from the REM, and a very slight apparent decrease in the standard deviations for most reanalyses. Not shown is a small

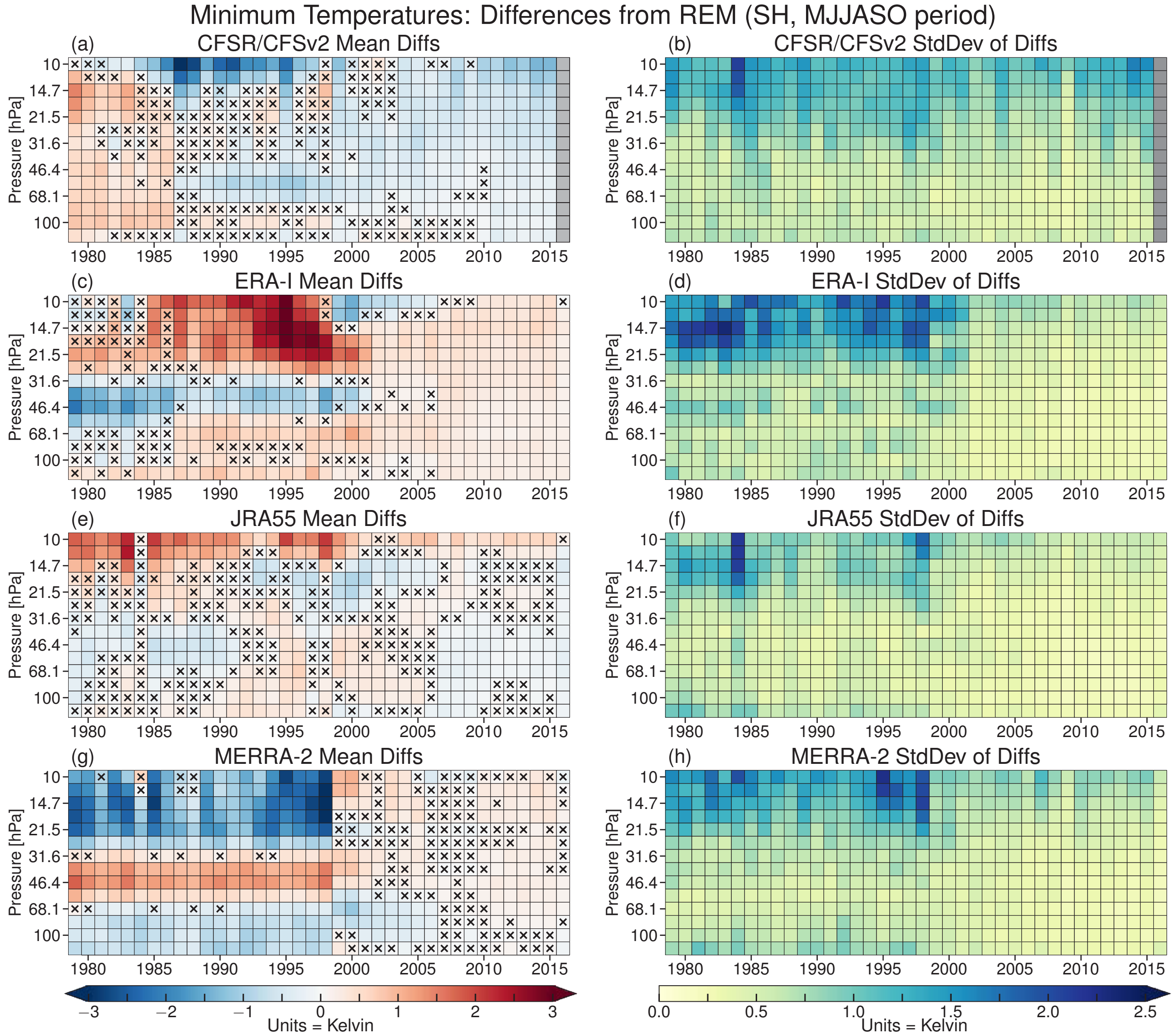


Figure 10.3: *SH extended winter season (MJJASO) (a, c, e, g) averages and (b, d, f, h) standard deviations of minimum daily temperature differences for each reanalysis from the reanalysis ensemble mean (REM, see text) as a function of year and pressure for the 1979 through 2017 winters, concatenated into pixel plots as described by Lawrence et al. (2018). Columns of grey pixels indicate years with no data. Pixels with x symbols inside indicate years and levels where the differences from the REM are insignificant according to our bootstrapping analysis (see Lawrence et al., 2018). In the average difference panels, negative values (reanalysis less than REM) are shown in blue and positive values (reanalysis greater than REM) are shown in red; in the standard deviation panels, yellows/deep blues represent low/high standard deviations of the reanalysis differences, respectively. From Lawrence et al. (2018).*

systematic decrease in the difference between MERRA and MERRA-2 for a narrow range of levels from 580 K to 700 K. JRA-55 shows generally stronger maximum PV gradients than the REM up through about 750 K, whereas CFSR/CFSv2 shows weaker gradients through most of that range. ERA-Interim and MERRA-2 are generally closer to the REM, with regions of stronger and weaker gradients alternating with height. As can be seen from the standard deviations, the variability in the daily differences increases strongly above about 520 K. The standard deviations also highlight a period of large variance in the differences between about 1995 and 2002 in the highest levels shown (above about 580 K) in ERA-Interim, JRA-55, and MERRA-2; this is a period when many data inputs were changing (including the TOVS to ATOVS transition) and during which the timing of those changes and the way they were handled (*e.g.*, whether a simple switch or using both older and newer inputs for a time) vary among the reanalyses. The maximum PV gradient diagnostic for the SH shows an improvement in agreement with the REM after about 1999, albeit not as strong as that for temperature diagnostics (*Lawrence et al.*, 2018).

Figure 10.5 shows the winter-mean (DJFM) volume of air with temperature below the NAT PSC threshold for the NH for each of the reanalyses, expressed as a fraction of the volume of air in the vortex. (The altitude for the depth in the volume calculation is obtained using the theta to altitude conversion approximation of *Knox* (1998).) The

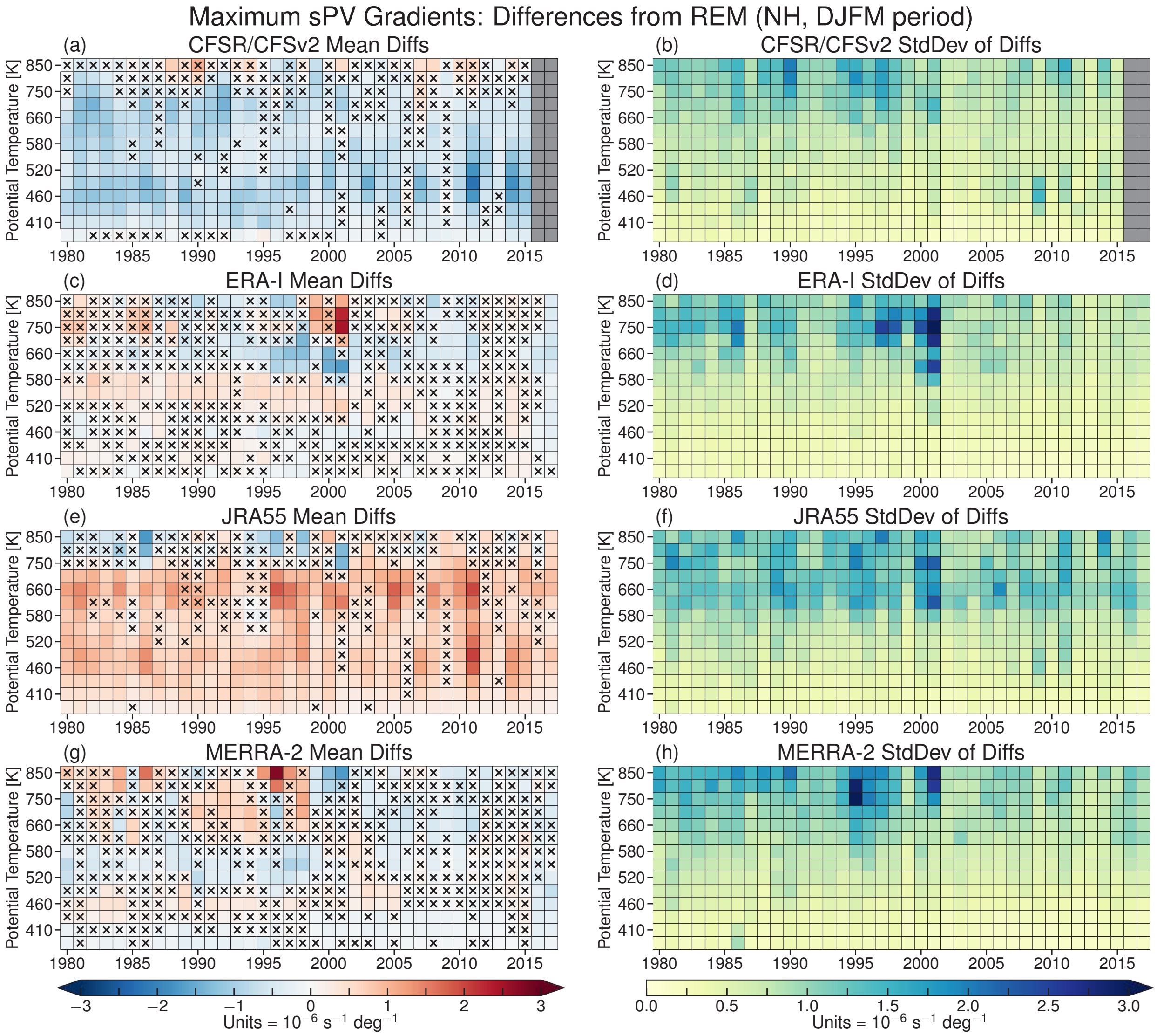


Figure 10.4: *As in* ***Figure 10.3****, but for NH maximum PV gradients for DJFM. Units are in 10^{-6} s^{-1} deg^{-1}. From Lawrence et al. (2018).*

extent of the bars shows the sensitivity of the diagnostic to using ± 1 K offsets in the threshold temperatures. Those threshold temperatures are calculated according to *Hanson and Mauersberger* (1988) for standard pressure surfaces (12 levels per decade in pressure, as have been used for several NASA satellite instrument data sets) using climatological HNO_3 and H_2O profiles (from all January Cryogenic Limb Array Etalon Spectrometer (CLAES) data and all January UARS MLS data, respectively); these values are most accurate for conditions that are neither denitrified nor dehydrated (see *Lawrence et al.*, 2018, and references therein, for further details). The threshold temperatures are then assigned to the "corresponding" standard theta level; the bar ranges for the ± 1 K offsets thus help to estimate uncertainties from the approximations for the HNO_3 and H_2O profiles and for the conversion to theta levels. The large interannual variability in the NH PSC potential is reflected accurately in all of the reanalyses throughout the period shown. Differences between the reanalyses are small throughout the period and do not show an obvious convergence to closer agreement. JRA-55 often shows a slightly larger volume than the other reanalyses. The sensitivity to the ± 1 K offsets is, as expected, largest when the volume itself is large. The corresponding diagnostics for the SH are shown in *Lawrence et al.* (2018); they indicate generally smaller differences, consistent with less interannual variability and much larger volumes with PSC potential in the Antarctic.

Lawrence et al. (2018) calculate vortex area and sunlit vortex area based on a vortex boundary defined by a climatological winter profile of the PV at the location of maximum PV gradients; that paper provides a detailed discussion of why this choice of vortex-edge definition is appropriate. Vortex area and sunlit vortex area show small but persistent biases in vortex size, with JRA-55 usually having a smaller vortex than ERA-Interim and CFSR/CFSv2 and the MERRA-2 vortex size difference from the REM varying with altitude. Consistent with a generally smaller vortex, JRA-55 tends to have earlier vortex decay dates than the other reanalyses evaluated; MERRA-2 has the latest vortex decay dates below about 550 K, and CFSR/CFSv2 the latest above that.

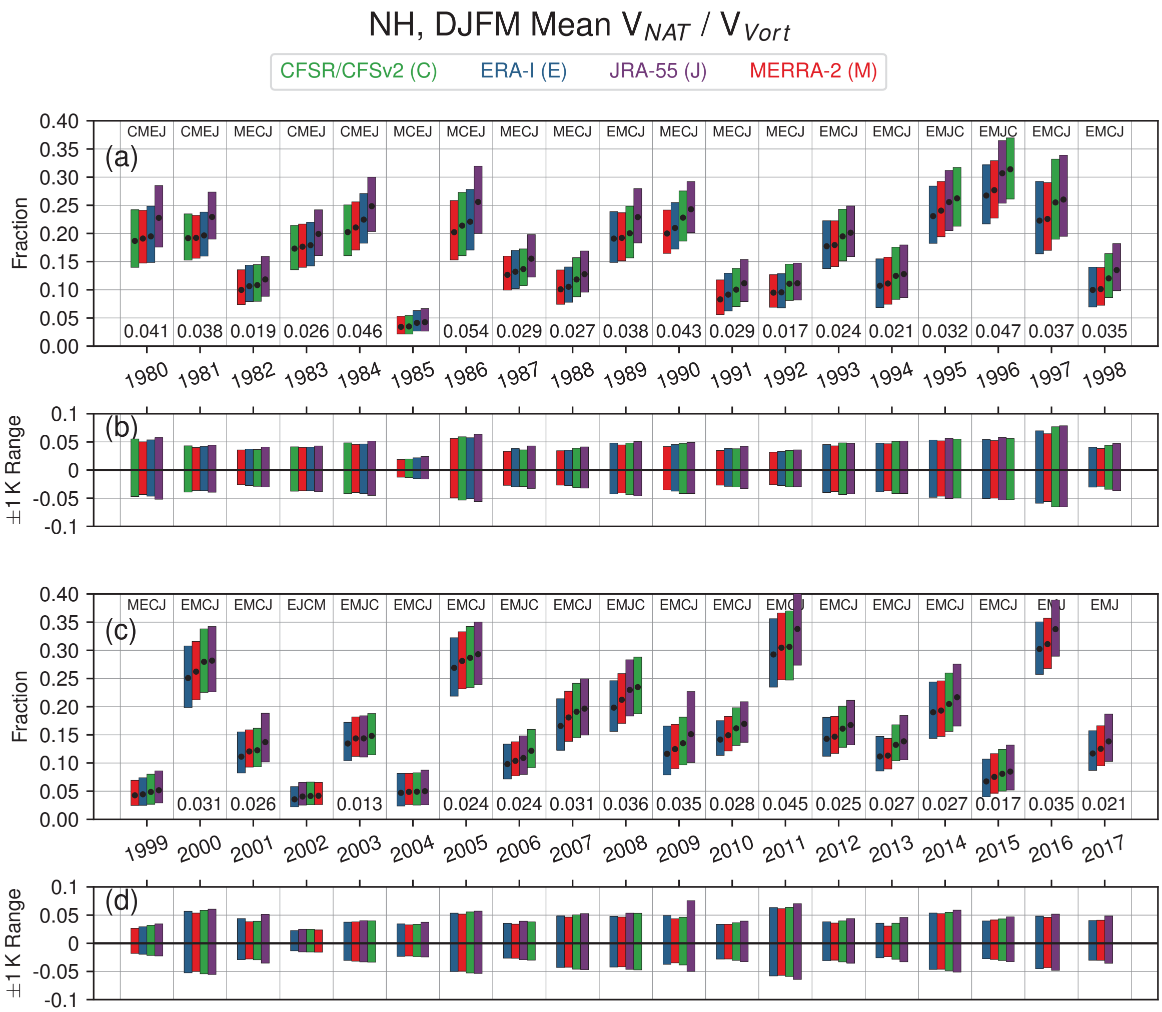


Figure 10.5: *Winter-mean (DJFM) fraction of vortex volume between the 390 K and 580 K isentropic surfaces with temperatures below T_{NAT} in the NH (a and c), and range of values obtained for the ± 1 K NAT threshold sensitivity tests (b and d). The bars are ordered from lowest to highest central values. The numbers at the bottom of (a) and (c) show the range of central values (that is, rightmost minus leftmost central value). Green, blue, purple, and red indicate CFSR/CFSv2, ERA-Interim, JRA-55, and MERRA-2, respectively. From Lawrence et al. (2018).*

Differences in vortex area and decay dates are small and do not obviously improve over the period studied.

10.5 Polar diabatic heating rates

Diabatic heating rates in the polar vortex regions are important to polar processing studies: Diabatic descent is the main driver of the downward transport of trace gases, including inorganic chlorine (Cl_y), and the replenishment of ozone in the lower stratospheric vortex, and it is thus critical to account for it in studies aimed at using observations to quantify chemical ozone loss (see, *e.g.*, *Manney et al.*, 2003a; *WMO*, 2007, and references therein, for reviews of methods for observational ozone loss studies). Although descent in the polar vortices can be estimated using quasi-conserved tracers such as N_2O or CO (*Ryan et al.*, 2018, and references therein), such estimates can be biased by mixing processes (*Livesey et al.*, 2015, and references therein), and in models and reanalyses the vertical motions of stratospheric air parcels are often constrained thermodynamically based on net diabatic heating, as initially laid out by *Murgatroyd and Singleton* (1961) and *Dunkerton* (1978).

Diabatic heating in the polar lower stratosphere is generally controlled by radiative effects – the sum of longwave cooling from thermal emission, and shortwave heating from solar absorption (when sunlight is present). For longwave cooling, the most important factors are temperature and the concentrations of major greenhouse gases carbon dioxide, water vapor, and ozone. Shortwave heating is determined primarily by the solar zenith angle and ozone amount. The physics of radiative transfer and the basic spectroscopic parameters needed to calculate net diabatic heating have been known for some time, and the foundation for understanding radiative heating and cooling in the stratosphere is well established (*e.g.*, *Mertens et al.*, 1999; *Olaguer et al.*, 1992; *Kiehl and Solomon*, 1986; *Ramanathan*, 1976). However, there are a number of important issues that remain for improving the accuracy of radiative heating and cooling calculations in the stratosphere, such as corrections to broadband schemes used in models, variations in the representation of water vapor longwave radiative effects, and uncertainties in solar near-infrared spectral irradiances (*e.g.*, *Menang*, 2018; *Maycock and Shine*, 2012; *Forster et al.*, 2001).

We use the zonal mean model-generated diabatic heating rates from the MERRA-2, ERA-Interim, JRA-55, and CFSR reanalyses described by *Martineau et al.* (2018; Section 3.6.1 of that paper, also see *Wright*, 2017), which are provided on a common grid to facilitate comparisons. These are on a standard pressure grid and given as the potential temperature tendency due to all physics; further details, including the processes included and differences in products provided, are given by *Martineau et al.* (2018) [1]. For the figures shown here we construct daily averages from the six-hourly fields provided and average those over the polar cap (60°-90°) in each hemisphere. The climatological comparisons are done for 1980 through 2010 so that CFSR can be included (diabatic heating rates for CFSRv2 were not archived).

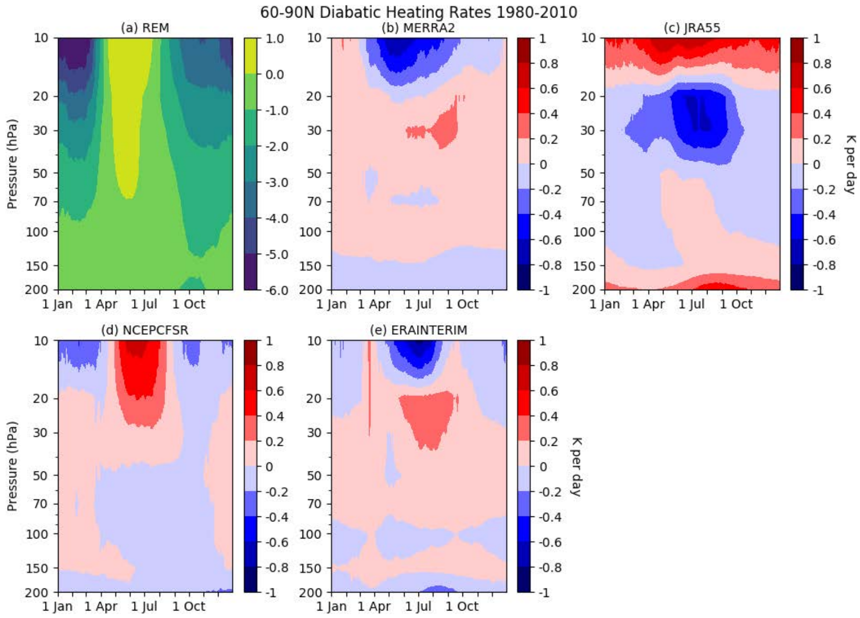


Figure 10.6: *Climatological 1980-2010 north polar cap (60°-90°N) diabatic heating rates in the lower stratosphere from four reanalyses: (a) REM for MERRA-2, ERA-Interim, JRA-55, and CFSR; (b through e) differences (reanalysis – REM, so positive values indicate that reanalysis values are higher than those for the REM) of each reanalysis from the REM climatology.*

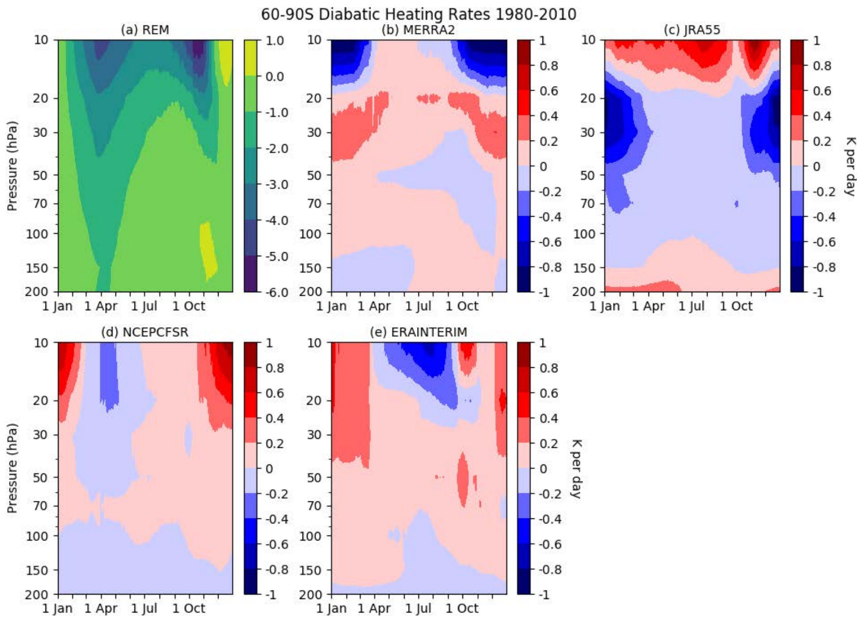


Figure 10.7: *As in **Figure 10.6** but for the south polar cap.*

Figure 10.6a shows NH polar cap averaged heating rates in the lower stratosphere for the REM (as defined in *Section 10.4*) constructed from MERRA-2, ERA-Interim, JRA-55, and CFSR. As expected, the strongest diabatic cooling (descent) is seen in late December through March, and stronger descent is seen at higher altitudes. Increased negative values in late January and February are coincident with the most common timing of strong sudden stratospheric warmings (SSWs), during which the temperatures far above radiative equilibrium give rise to stronger diabatic cooling (see, *e.g.*, *Manney et al.*, 2008, 2009b, and references therein). Differences of each reanalysis from the REM are shown in **Figure 10.6b** through **10.6e**. The largest differences between the reanalyses and the REM are at the lowest pressures, with differences up to about ±1 K/day near and above (at lower pressures than) 30 hPa. During the cold season (November through April) that we are most interested in here, the magnitude of the differences from the REM is typically no more than about 0.4 K/day except above (at pressures below) 20 hPa; cold-season differences are thus typically within about 10 %. While there is much interannual variability (in line with the large interannual variability in temperatures), examination of individual years indicates generally consistent biases between the reanalyses, with JRA-55 generally showing lower values (stronger diabatic cooling) than the other reanalyses in

[1] also see the footnote on diabatic heating rates in reanalyses in *Section 12.1.3*.

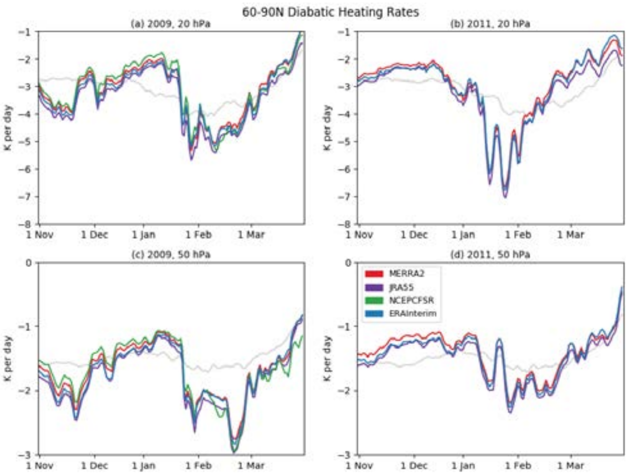


Figure 10.8: *North polar cap (60 - 90 °N) average diabatic heating rates in the lower stratosphere from four reanalyses for 2009 (left panels) and 2011 (right panels) in the NH at 20 hPa (top panels) and 50 hPa (bottom panels) during the cold seasons. Coloured lines show the reanalyses as given in the legend. The grey line shows the REM 1980 - 2010 climatology for the same time of year as a reference. Values from CFSR are not available in 2011.*

a layer between approximately 50 hPa and 15 hPa in the warm season. The seasonality of the differences from the REM at the lowest pressures shown is strongly influenced by JRA-55, which has a weaker seasonal cycle in the diabatic heating rates than the other reanalyses.

Figure 10.7 shows a similar REM climatology and reanalysis differences for the SH polar cap. The strongest radiative cooling (largest negative values) is seen in late spring during the final warming (October and November), with strong cooling also seen in fall. As in the NH, the largest differences among the reanalyses occur during the warm season and in late spring, and at levels above (pressures below) about 20 hPa except for JRA-55 warm-season differences near 50 hPa to 15 hPa. Outside of those times and regions, the magnitude of differences is typically less than 0.4 K/day, and less than 0.2 K/day below (at pressures higher than) about 30 hPa, thus within about 15 %. As was the case in the NH, JRA-55 has a weaker seasonal cycle in the diabatic heating in the SH than the other reanalyses.

In light of the large changes in temperature agreement before and after the transition from assimilating TOVS to ATOVS data (*e.g.*, *Lawrence et al.*, 2018; *Long et al.*, 2017), we examined separately climatologies between 1980 and 1998, and between 2000 and 2010, for each hemisphere. Overall, the agreement among reanalysis diabatic heating rates does not differ substantially during the two periods, suggesting that the differences in the reanalyses' radiative transfer models and their inputs (*e.g.*, differences in handling ozone and water vapor) are likely a larger factor than changes in the satellite radiances used to define temperatures for the radiation calculations. Maximum values of differences at pressures below 30 hPa are slightly smaller for most of the reanalyses in the ATOVS period, but the differences are small and qualitative differences appear negligible in the climatology.

To examine the day-to-day differences among the reanalyses, we show time series at 20 hPa and 50 hPa for individual winters, choosing a relatively quiescent and a very disturbed winter in each hemisphere. **Figure 10.8** shows the NH cold seasons of 2008/2009 (a year with a strong, prolonged vortex-split SSW, *e.g.*, *Kuttippurath and Nikulin*, 2012; *Labitzke and Kunze*, 2009; *Manney et al.*, 2009a) and 2010/2011 (at the time, the year with the largest Arctic ozone loss on record, *e.g.*, *Balis et al.*, 2011; *Manney et al.*, 2011b; *Sinnhuber et al.*, 2011). At the onset time of the SSW in late January 2009, there is an abrupt increase in diabatic cooling (stronger negative values) at both levels shown, as temperatures rose far above radiative equilibrium. A previous period of unusually strong diabatic cooling occurred at both levels in November 2009, and a period of unusually weak diabatic cooling occurred immediately before the SSW onset. In contrast, in 2011, a sharp increase in cooling in late January at 20 hPa was associated with a minor SSW that did not strongly affect the lower stratospheric vortex; except during this period at the higher levels, diabatic cooling in 2011 was close to, but often somewhat weaker than, climatological values, consistent with lower-than-average temperatures. While the reanalyses show small biases (about 0.5 K at 20 hPa and 0.3 K at 50 hPa), they all follow the day-to-day variations quite well, such that any of these would give a representative picture of the daily evolution of diabatic heating. CFSR often shows slightly less cooling than the other reanalyses and occasionally shows qualitatively different behavior for a few days to a week or so (*e.g.*, at 50 hPa in mid-February and late March 2009).

Figure 10.9 shows 2002 (the winter with the only major SSW on record in the SH, *e.g.*, *Shepherd et al.*, 2005, and references therein) and 2006 (a year with one of the deepest ozone holes on record, *e.g.*, *WMO*, 2018). In contrast to the NH, the diabatic heating rates in the SH are usually near the climatological values. In 2006, diabatic cooling at 20 hPa was slightly

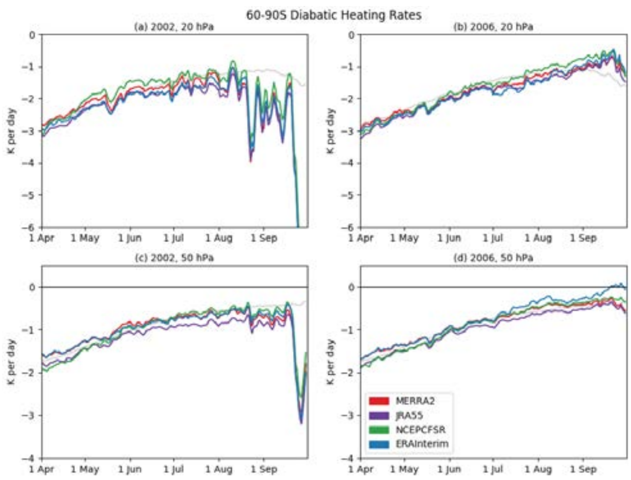


Figure 10.9: *As in **Figure 10.8** but for the south polar cap average in 2002 (left panels) and 2006 (right panels).*

weaker than the climatology at the end of the season, consistent with lower-than-usual temperatures (closer to radiative equilibrium) at that time. The contrast with 2002 is dramatic – the SSW occurred in late September, when there was an abrupt increase in diabatic cooling at all levels examined. Unusually large wave activity was seen for about a month prior to the SSW (*e.g.*, *Krüger et al.*, 2005; *Newman and Nash*, 2005), leading to modest temperature increases, consistent with the strengthening of the diabatic cooling over that in the climatology during the month prior to the SSW. During both winters, similar to the NH results, all of the reanalyses capture the day-to-day variability well, with some small overall biases apparent. The spread among the reanalyses is slightly larger than that in the NH, especially at 50 hPa, with a range among the reanalyses near 0.5 K at both levels shown. Again, CFSR often shows slightly less cooling than the other reanalyses, but ERA-Interim shows the weakest diabatic cooling among the reanalyses from July through September 2006. JRA-55 shows stronger diabatic cooling than the other reanalyses at 50 hPa during most of both winters shown.

This evaluation of diabatic heating rates in the lower stratosphere indicates generally good agreement among the reanalyses and that each of the reanalyses compared can reasonably be used in studies of polar processing, with some limitations; in particular, the weaker seasonal cycle in JRA-55 than in any of the other reanalyses suggests some caution in using heating rates from that reanalysis in the middle stratosphere and higher.

10.6 Concordiasi superpressure balloon comparisons

In this section we review the work presented by *Hoffmann et al.* (2017a) in the S-RIP special issue on the superpressure balloon measurements made during the Antarctic Concordiasi campaign (*Rabier et al.*, 2010) in September 2010 to January 2011 to evaluate meteorological analyses and reanalyses. The study covers the ERA-Interim, MERRA, MERRA-2, and NCEP-NCAR R1 reanalyses, as well as the higher-resolution ECMWF operational analysis (OA, see *Section 10.1* for details). The comparison was performed with respect to long-duration observations from 19 superpressure balloon flights, with most of the balloon measurements conducted at altitudes of 17 - 18.5 km and latitudes of 60° - 85° S. The five meteorological data sets considered by *Hoffmann et al.* (2017a) differ significantly in spatial and temporal resolution, with the truncation of the models limiting the capability of representing small-scale fluctuations.

10.6.1 Temperatures and winds

Hoffmann et al. (2017a) (see their Table 3) found positive temperature biases in the range of 0.4 K to 2.1 K, zonal wind biases in the range of – 0.3 m s^{-1} to 0.5 m s^{-1}, and meridional wind biases below 0.1 m s^{-1} for all data sets compared to the balloon measurements. The largest biases and standard deviations were typically found for NCEP-NCAR R1, which may be attributed to the fact that this data set is independent, whereas the Concordiasi balloon observations were assimilated into the other reanalyses. However, significant differences between ECMWF (OA and Interim), MERRA, and MERRA-2 are manifest, despite the balloon data being assimilated into each of these products. Hoffmann et al. suggest that this is a result of inter-model dependencies, such as the relative weighting of observations, types of forecast models, and the particular assimilation procedures of each model.

The superpressure balloon observations are a valuable source of data with which to study small-scale fluctuations in the atmosphere that are mostly attributed to gravity waves, as demonstrated in several studies (*e.g.*, *Jewtoukoff et al.*, 2015; *Vincent and Hertzog*, 2014; *Plougonven et al.*, 2013; *Hertzog et al.*, 2008, 2012). Statistics of high-pass filtered balloon data were used by *Hoffmann et al.* (2017a) (see their Table 4) to assess the representation of small-scale fluctuations in the meteorological analyses. ECMWF OA reproduces about 60 % of the standard deviations of the temperature and wind fluctuations of the balloons, and the temperature results are consistent with the recent studies of *Jewtoukoff et al.* (2015) and *Hoffmann et al.* (2017b). Therefore, the ECMWF operational model resolves the atmospheric gravity wave spectrum with higher fidelity than the reanalyses since ERA-Interim, MERRA, and MERRA-2 reproduce only about 30 % of the standard deviations of the temperature and wind fluctuations seen in the high-pass filtered balloon data, and NCEP-NCAR R1 reproduces only about 15 % for temperature and 30 % for the winds. These results are correlated with the spatial resolutions of the analyses, which determine the ability of the forecast models to simulate realistic gravity wave patterns.

Large-scale biases and standard deviations of temperatures and horizontal winds at different latitudes were averaged over the entire time period of the campaign by *Hoffmann et al.* (2017a) and are reproduced in **Figure 10.10**. An increasing temperature bias from middle to high latitudes is seen in all reanalyses. NCEP-NCAR R1 shows the largest temperature warm bias (3.1 K) at high latitudes of 80 - 85° S, but a bias is also present in reanalyses that assimilated the Concordiasi balloon observations, although to a lesser extent: MERRA (1.4 K), MERRA-2 (1.3 K), ERA-Interim (1.1 K), and ECMWF OA (0.5 K). Southern Hemisphere winter pole temperature biases were also reported for earlier winters in other studies (*Boccara et al.*, 2008; *Parrondo et al.*, 2007; *Gobiet et al.*, 2005), and the Concordiasi study indicates that such biases were still present in 2010/2011. *Gobiet et al.* (2005) speculate that the assimilation of microwave radiances from satellite measurements into ECMWF analyses may be a reason for the temperature bias. The NCEP-NCAR R1 warm bias found by *Hoffmann et al.* (2017a) is comparable to those found in earlier studies. Other analyses have smaller biases, which *Hoffmann et al.* attribute to the assimilation of the Concordiasi data.

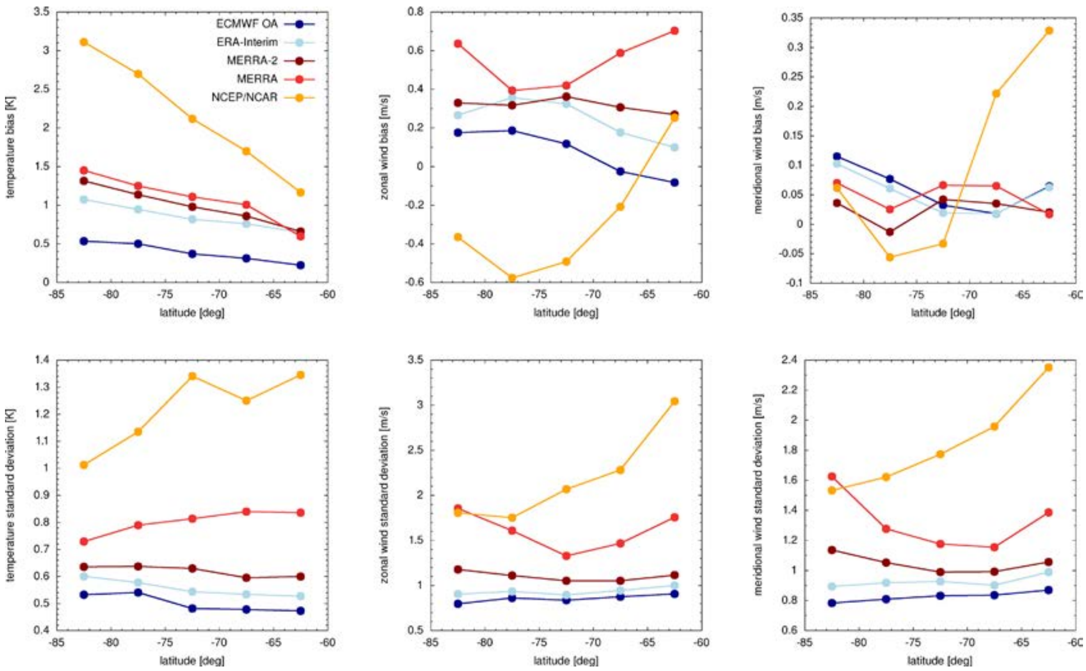


***Figure 10.10:** Bias and standard deviations of temperature and horizontal winds of meteorological analyses minus Concordiasi balloon data at different latitudes. From Hoffmann et al. (2017a).*

Moreover, the wind biases as well as temperature and wind standard deviations shown in **Figure 10.10** are also generally largest for NCEP-NCAR R1, whereas they are smaller and more similar to each other for both ECMWF data sets and MERRA-2.

10.6.2 Trajectories

Hoffmann et al. (2017a) used the balloon tracking observations to evaluate reanalysis-driven trajectory calculations with the Lagrangian transport model MPTRAC, where the vertical motions of simulated trajectories were nudged to pressure measurements of the balloons. Absolute horizontal transport deviations (AHTDs) and relative horizontal transport deviations (RHTDs) are commonly used to compare trajectory calculations with observations or to evaluate results for different model configurations (*Stohl*, 1998; *Stohl et al.*, 1995; *Rolph and Draxler,* 1990; *Kuo et al.*, 1985). AHTD at a particular time is the average Euclidean distance between the observed balloon and modeled air parcel positions projected to the Earth's surface. RHTDs are calculated by dividing the AHTD of individual air parcels by the length of the corresponding reference trajectory. Of course, the accuracy of calculated trajectories depends strongly on the fidelity of the particle dispersion model being used as well as the details of its configuration (*e.g.*, the time step, which is 30 seconds in these runs, consistent with the sampling rate of the balloon data). Our purpose here, however, is not to evaluate the trajectories themselves, but rather to compare the air motions calculated using the same model setup but different reanalyses to drive the model with long-duration balloon observations as a tool for assessing transport in the reanalyses.

Figure 10.11 shows transport deviations for the different meteorological data sets calculated from over a hundred samples of 15-day trajectories, and at the end of this time the AHTDs are in the range 1400 km to 2200 km. Estimates of the growth rates of the AHTDs are typically within 60 km day^{-1} to 170 km day^{-1}. The RHTDs are in the range 4 - 12 % after 2 days, but converge to a smaller range of 6 - 9 % after 15 days. Although the transport deviations grow rather steadily, the relative differences between the data sets tend to get smaller over time. The largest transport deviations and growth rates were found for NCEP-NCAR R1, which did not assimilate the Concordiasi balloon observations. The results agree well with those reported by *Boccara et al.* (2008) for the Vorcore campaign

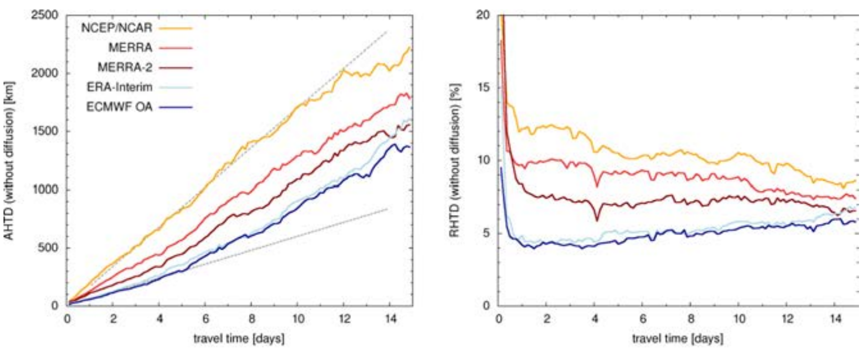


***Figure 10.11:** Absolute (left, AHTD) and relative (right, RHTD) horizontal transport deviations of simulated and observed balloon trajectories for different meteorological analyses. Dotted grey lines represent AHTD growth rates of 60 km day^{-1} and 170 km day^{-1}. From Hoffmann et al. (2017a).*

in 2005, where mean spherical distances of about 1650 km (with an interquartile range of 800 - 3600 km) for ECMWF analyses and 2350 km (1400 - 3800 km) for NCEP-NCAR R1 data were found for 15-day trajectories.

10.7 PSC thermodynamic-consistency diagnostics

In this section we review the work presented by *Lambert and Santee* (2018) in the S-RIP special issue on comparisons of reanalysis temperatures with COSMIC GNSS-RO temperatures and with independent absolute temperature references derived from theoretical considerations of PSC formation defined by the equilibrium thermodynamics of STS and water-ice clouds. The PSC thermodynamic-consistency diagnostics rely on the near-simultaneous and colocated measurements of nitric acid, water vapour and cloud phases provided from the long-term precise formation flying of the CALIOP and Aura MLS instruments within the afternoon "A-Train" satellite constellation. The initial A-Train configuration of the CALIPSO and Aura spacecraft from April 2006 to April 2008 resulted in an across-track orbit offset of ~200 km, with the MLS tangent point leading the CALIOP nadir view by about 7.5 minutes. From April 2008 until September 2018, Aura and CALIPSO were operated to maintain positioning within tightly constrained control boxes, such that the MLS tangent point and the CALIOP nadir view were colocated to better than about 10 - 20 km and about 30 seconds.

The analysis refines and extends the methodology used originally by *Lambert et al.* (2012) to investigate the temperature existence regimes of different types of PSCs. The CALIOP lidar PSC classification is used to identify the presence of STS and ice PSCs in the MLS geometric FOV at the along-track resolution (165 km × 2.16 km). MLS provides ambient gas-phase H_2O and HNO_3 volume mixing ratios, which are required to calculate the theoretical equilibrium temperature dependence of the STS (T_{eq}) and ice (T_{ice}) PSCs, based on the expressions of *Carslaw et al.* (1995) and *Murphy and Koop* (2005), respectively. Observed and calculated temperature distributions are compared for both the uptake of HNO_3 in STS and the ice frost point for each reanalysis data set and for MLS temperature. Viewing scenes having a distinct dominant PSC classification in a sample volume similar in size to the MLS gas-species resolution are selected, with the requirement that 75 % or more of the CALIOP pixels in the MLS geometric FOV have the same PSC classification. Scenes satisfying this requirement for CALIOP STS detections are denoted as LIQ; those for CALIOP ice detections are denoted as ICE. Reanalysis temperature biases are then quantified relative to the respective LIQ and ICE reference temperatures. The calculated standard deviations of the temperature differences are used to estimate the measurement precisions. *Lambert et al.* (2012) show that the resulting root mean square (RMS) temperature uncertainties for the LIQ and ICE references are smaller than those derived for the MLS retrieved temperatures and comparable to the measurement capabilities of the GNSS RO technique (bias < 0.2 K, precision > 0.7 K) in the lower stratosphere. The estimated measurement precisions for the STS equilibrium and ice frost points are 0.4 K and 0.3 K, respectively, in the 68 - 21 hPa pressure range, with the corresponding estimated measurement accuracies in the range of 0.7 - 1.6 K for STS and 0.4 - 0.7 K for ice.

The approach for PSC thermodynamic-consistency diagnostics is summarized as follows:

- Identify LIQ and ICE PSCs using the CALIOP lidar measurements.
- Accumulate the CALIOP PSC types (LIQ and ICE) at the MLS along-track resolution (165 km × 2.16 km), ensuring that the same PSC type is detected in at least 75 % of the MLS FOV.
- Calculate the theoretical temperature dependence of STS (T_{eq}) and ice (T_{ice}) PSCs under equilibrium conditions using the spatially and temporally colocated MLS gas-phase HNO_3 and H_2O measurements.
- Compare reanalyses and MLS data with (a) calculated and observed HNO_3 uptake in STS and (b) ice temperature distribution *vs*. the frost point temperature.
- Create LIQ and ICE temperature distributions for each reanalysis (all interpolated to the MLS measurement times and locations) as well as MLS.
- Calculate the median and mean temperature deviations from T_{eq} and T_{ice} and their standard deviations for LIQ and ICE classifications, respectively.

In the Antarctic, six PSC seasons from 20 May (d140) to 18 August (d230) were investigated from 2008 to 2013 for latitudes poleward of 60 °S in the lower stratosphere (100 - 10 hPa). In the Arctic, five PSC seasons from 2 December (d336) to 31 March (d090) were investigated from 2008/2009 to 2012/2013 for latitudes poleward of 60 °N.

10.7.1 Mountain waves

Small-scale temperature fluctuations are not captured accurately by the reanalyses because of their limited spatial resolution (*e.g.*, *Hoffmann et al.*, 2017a). An orographic gravity wave case study over the Palmer Peninsula was used by *Lambert and Santee* (2018) to show that, at 50 hPa, the temperature amplitudes resolved by the reanalyses can vary by up to a factor of two in extreme circumstances. The differences in the wave amplitudes are well correlated with the spatial resolutions of the reanalyses. It is thus important to identify regions characterized by small-scale temperature fluctuations and remove them from further consideration in this study. The high vertical resolution of the COSMIC temperatures allows examination of the spectrum of temperature variance over the height region of PSC activity. An estimate of the potential energy density of the wave disturbance, E_p,

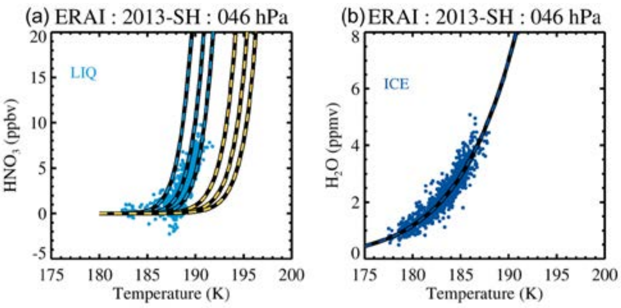


Figure 10.12: *Scatterplots of coincident MLS HNO_3 and H_2O vs ERA-Interim temperature for PSCs classified by CALIOP on the 46-hPa pressure level in the 2013 Antarctic winter. (a) MLS HNO_3 vs ERA-Interim temperature for liquid (LIQ) PSCs (light blue dots). Note that measurement noise can lead the MLS data processing algorithms to retrieve negative HNO_3 mixing ratios when abundances are low (e.g., under highly denitrified conditions). Though unphysical, such negative values must be retained to avoid introducing a positive bias into any averages calculated from the measurements. The theoretical equilibrium uptake of HNO_3 by STS is shown for representative ambient H_2O values by the light blue-black dashed lines. The yellow-black dashed lines show the corresponding NAT equilibrium curves. (b) MLS H_2O vs ERA-Interim temperature for ice (ICE) PSCs (dark blue dots). The dark blue-black dashed lines indicate the theoretical equilibrium for the frost point temperatures. Adapted from Lambert and Santee (2018).*

provides an effective means of applying low-pass filtering to reveal the large-scale temperature structure. A low-pass filter was applied by excluding profile matches with COSMIC temperature variances $> 0.5\,K^2$ ($E_p > 1.5\,J\,kg^{-1}$). COSMIC temperatures were restricted to below 200 K to focus on regions of potential PSC existence.

10.7.2 Thermodynamic equilibrium

PSC theoretical existence temperatures, calculated based on equilibrium thermodynamics, are dependent on the ambient partial pressures of H_2O in the case of the ice frost point, T_{ice} (*Murphy and Koop*, 2005), and also of HNO_3 for STS (*Carslaw et al.*, 1995).

Figure 10.12 shows scatterplots against ERA-Interim temperature of coincident MLS HNO_3 (left panel) and H_2O (right panel) for Antarctic PSCs classified by CALIOP at 46 hPa in 2013, along with the theoretical HNO_3 gas-phase uptake curves for STS and NAT. It is clear that LIQ PSCs are closely associated with the equilibrium STS curve. Similarly, the scatter of gas-phase H_2O is closely associated with the frost point temperature in the presence of ICE PSCs.

In **Figure 10.13** we show the variation in MLS gas-phase HNO_3 with ERA-Interim temperature for CALIOP PSC classifications at 31 hPa for one Southern Hemisphere winter. MLS HNO_3 data are separated into corresponding CALIOP PSC categories, allowing comparison of

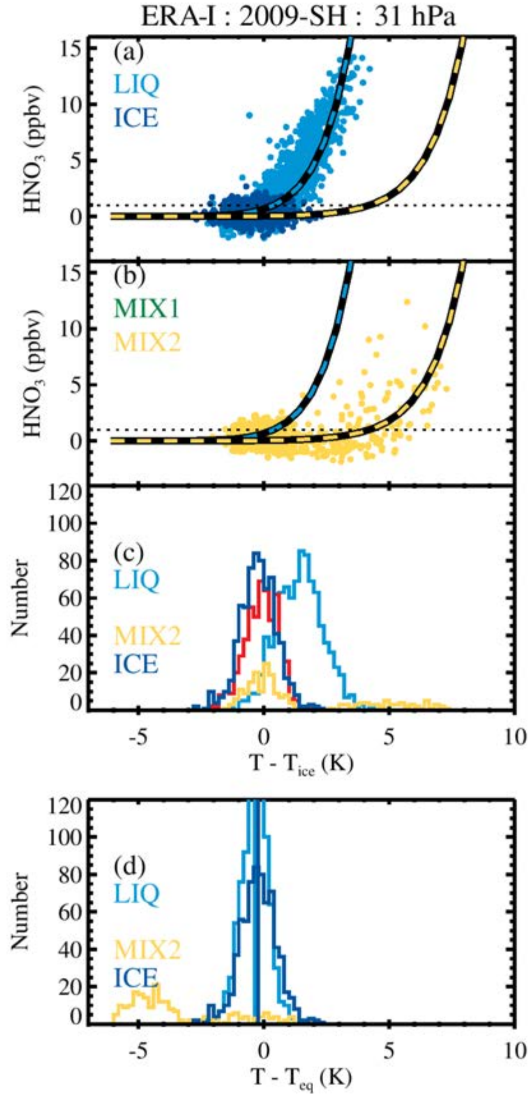


Figure 10.13: *Composite statistics for d140 - d230 (the bulk of the PSC existence period) of a representative year (2009) of the MLS gas-phase HNO_3 variation with ERA-Interim temperature corresponding to CALIOP PSC classifications at 31 hPa, with the added constraint that at least 75 % of the MLS FOV is filled with the same classification. (a) Scatter plot of HNO_3 vs temperature deviation from the frost point ($T - T_{ice}$) for PSCs classified as LIQ (light blue) and ICE (dark blue). (b) As for (a), but for MIX2 PSCs (yellow). Equilibrium STS (light blue-black dashed) and NAT (yellow-black dashed) curves show the theoretical uptake of HNO_3. Because of non-equilibrium effects, which cause larger temperature scatter, the CALIOP NAT classifications MIX1 and MIX2 are not used in this analysis, but this panel is shown to indicate the good discrimination between the solid and liquid uptake curve branches. Note that there are no MIX1 PSC detections for the case shown here. (c) Temperature histograms for HNO_3 mixing ratios > 1 ppbv for the LIQ PSC type; data in the ICE classification are not subject to this constraint. The red histogram indicates the distribution of LIQ PSCs that have HNO_3 below the 1 ppbv threshold. (d) Temperatures transformed according to the STS equilibrium curve for the LIQ classification and NAT equilibrium curve for the MIX2 classification; the ICE classification remains the same as in (c) for comparison. From Lambert and Santee (2018).*

observed and modeled uptake of HNO_3 in different types of PSCs. The scatter of MLS HNO_3 against the temperature deviation from the frost point (calculated using MLS H_2O) is shown in **Figure 10.13a** for LIQ and ICE PSCs. Uptake of gas-phase HNO_3 in the presence of liquid-phase LIQ PSCs follows the STS equilibrium curve. In contrast, HNO_3 abundances are very low (typically ~2 ppbv or less) in the presence of ICE PSCs. In **Figure 10.13b**, uptake of gas-phase HNO_3 in clouds identified by CALIOP as being solid NAT MIX2 shows significant non-equilibrium variation, lying between the STS and NAT equilibrium curves. For this reason, we do not use either of the CALIOP NAT classifications, MIX1 or MIX2, in this analysis. Histograms of the temperature distributions of **Figure 10.13a** and **b** are shown in **Figure 10.13c**. Light blue (LIQ) and yellow (MIX2) represent distributions for HNO_3 mixing ratios > 1 ppbv, whereas red (LIQ) represents distributions for HNO_3 mixing ratios < 1 ppbv. The tails of the temperature distributions for LIQ (light blue and red histograms) do not reach temperatures lower than those in the distribution for ICE (dark blue), and no peaks indicating the existence of PSCs at a frost-point depression near $T_{ice} - 3$ K are observed. In **Figure 10.13d**, the temperatures are transformed with respect to the corresponding equilibrium curves: STS for the LIQ classification and NAT for the MIX2 classification, with the ICE classification remaining the same as in **Figure 10.13c** for comparison. As a result, the LIQ histogram narrows and is shifted (*Lambert et al.*, 2012). This is just an illustrative case; since there are seven temperature data sources, collected over six years on six pressure levels, the total number of histograms is 252 for each hemisphere. Statistics of the temperature difference distributions were generated for each reanalysis data set for each year and pressure level.

10.7.3 Temperature difference profiles

Intercomparisons of the reanalysis temperature statistics are displayed in **Figure 10.14** as vertical profiles over 100 - 10 hPa averaged for the Antarctic winters 2008 - 2013. For ICE PSCs, median temperature bias values fall in the range −1.0 to +0.1 K with standard deviations of ~0.7 K, both of which are larger than those of their LIQ PSC counterparts. Median values for the ICE reference are more uniform throughout the profile than those for the LIQ reference, which become increasingly negative with decreasing pressure. Median values for the LIQ reference are consistently biased low relative to the corresponding values for ICE by ~0.5 K, although standard deviations for the LIQ reference are smaller than those for ICE. The largest bias is found for MLS. In addition to differences with respect to the LIQ and ICE reference points, we also show comparisons with the COSMIC temperatures (90° - 60° S, mean variance < 0.5 K^2, and COSMIC temperatures < 200 K). Biases for the COSMIC reference are generally smaller than those for either LIQ or ICE. MERRA does not assimilate COSMIC GNSS-RO data, and it tends to exhibit the largest bias with respect to the COSMIC reference. The comparisons with thermodynamic equilibrium temperatures are completely independent of any data assimilated by the reanalyses. Comparing the standard deviation of the temperature differences (SD), it is apparent that, except at 100 hPa, the (PSC temperature − Reanalysis) SDs (**Figure 10.14d, e**) are smaller than the (COSMIC − Reanalysis) SDs (**Figure**

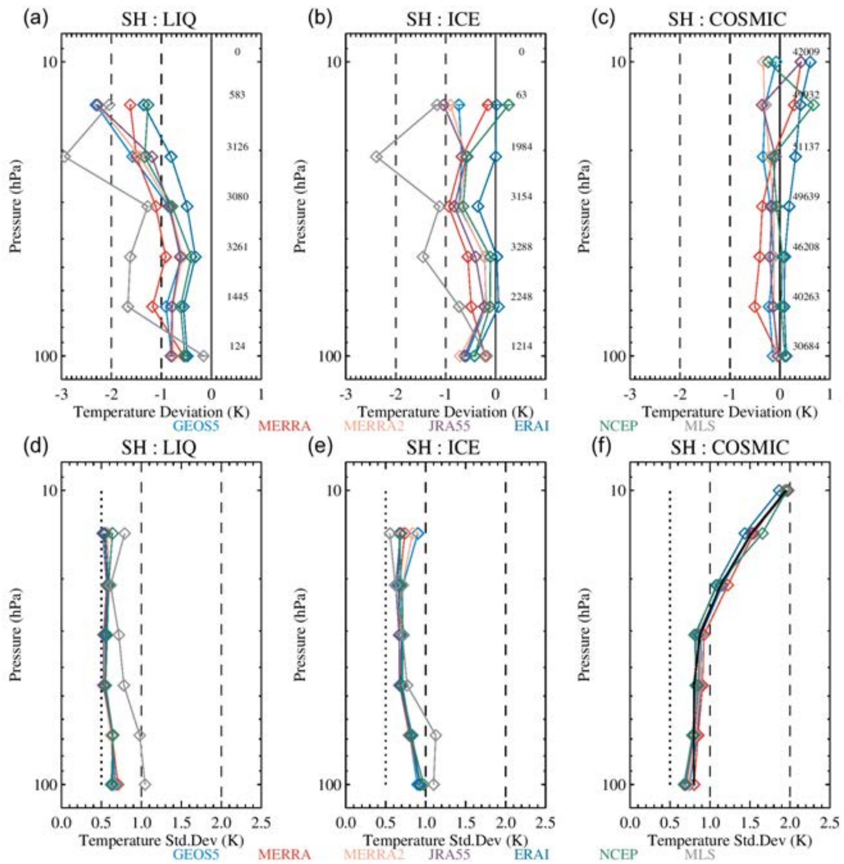


__Figure 10.14:__ Vertical profiles of median temperature deviations from T_{eq} for (a) LIQ PSCs and (b) ICE PSCs for the temperature distributions accumulated over Antarctic PSC seasons (20 May to 18 August, d140 - d230) from 2008 to 2013. Lines for the different reanalyses (MERRA, MERRA-2, JRA-55, ERA-Interim, and NCEP (CFSR/CFSv2)), GEOS-5.9.1, and MLS are colour coded (see legend); the numerical values on the right-hand side of the panel indicate the total number of observations in the distribution at the corresponding pressure level. Error bars for the median temperatures are displayed but are generally smaller than the symbol sizes. (c) as for (a,b) except for deviations with respect to COSMIC GNSS-RO. (d,e,f) The standard deviations of the corresponding temperature distributions shown in (a,b,c). Dotted lines indicate a standard deviation of 0.5 K. From Lambert and Santee (2018).

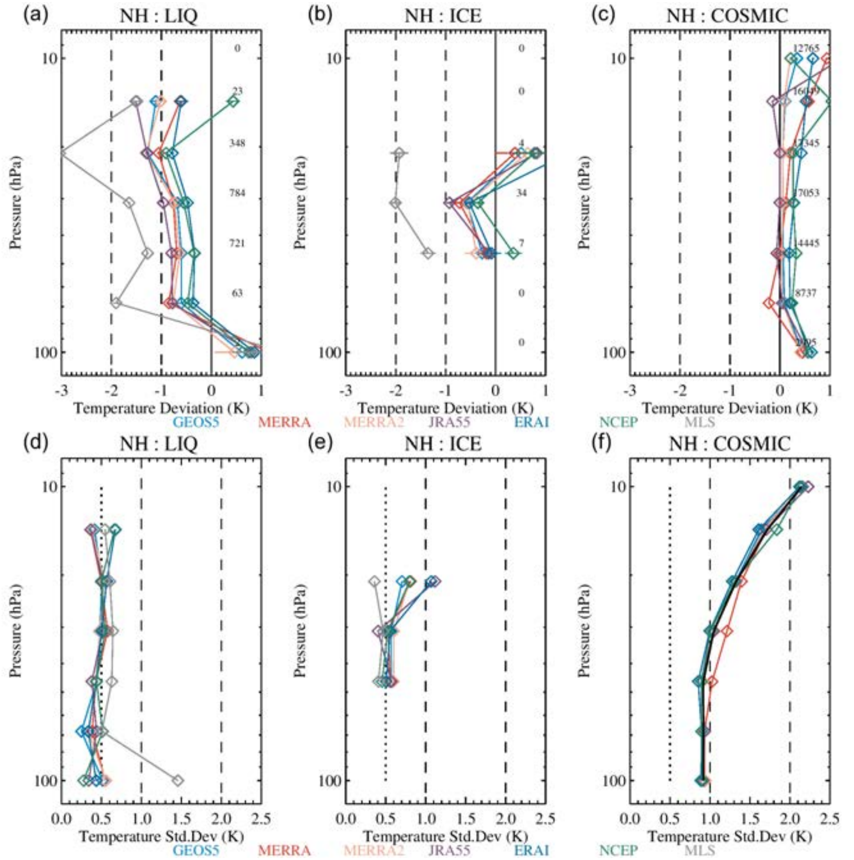


Figure 10.15: *As for* ***Figure 10.14****, but for Arctic PSC seasons (2 December to 31 March, d336 - d090) from 2008/2009 to 2012/2013. From Lambert and Santee (2018).*

10.14f). Therefore, in terms of random errors, the PSC-based temperature references perform better than the COSMIC temperature measurements. **Figure 10.15** shows the corresponding observations for the Arctic, where limited ice cloud production and hence few data points preclude a robust conclusion for ICE PSCs. Again, MLS shows the largest bias.

Although we have not directly matched A-Train locations with COSMIC occultations, we can estimate the differences between the COSMIC temperatures and the thermodynamic reference temperatures by elimination of the reanalysis temperature biases, $\overline{\Delta T_{re}}$. The temperature bias difference profiles for all reanalyses are tightly clustered (not shown), especially over the pressure range 68 - 21 hPa, justifying the assumption that $\overline{\Delta T_{re}}$ can be eliminated. The biases for ($\overline{\Delta T_{LIQ}} - \overline{\Delta T_{COSMIC}}$) have similar magnitude and profile shape over the pressure range 68 - 21 hPa in both hemispheres, with $\overline{\Delta T_{LIQ}}$ being smaller than $\overline{\Delta T_{COSMIC}}$ by about 0.5 K to 1.0 K. Likewise, the biases for $\overline{\Delta T_{ICE}}$ are about 0 K to 0.5 K smaller than $\overline{\Delta T_{COSMIC}}$ in the Antarctic, but there are too few data points to make a meaningful comparison in the Arctic.

10.7.4 Summary of the temperature differences

The accuracy and precision of several contemporary reanalysis data sets were evaluated through comparisons with (a) COSMIC GNSS-RO temperatures and (b) absolute temperature references obtained from the equilibrium properties of certain types of PSCs. A concise summary of the ranges of the mean temperature biases of the reanalyses relative to the LIQ (− 1.6 K to − 0.3 K) and ICE (− 0.9 K to + 0.1 K) equilibrium references as well as COSMIC (− 0.5 K to + 0.2 K) is given in **Figure 10.16**, which depicts bias ranges for the pressure domain 68 - 21 hPa. For all reference points, the coldest reanalysis biases tend to be found in the Antarctic and the warmest in the Arctic. The fact that GNSS-RO data are not assimilated in MERRA is evident in its higher biases with respect to COSMIC temperatures in the Antarctic (where there is a relative paucity of conventional measurements) compared to other reanalyses.

Reanalysis temperatures are found to be lower than the absolute reference points by 0.3 K to 1 K for LIQ and 0 K to 1 K for ICE at the altitudes of peak PSC occurrence (68 - 32 hPa). For LIQ, the negative biases are larger above 32 hPa than below that level. Median deviations for LIQ are consistently biased lower than those for ICE by ~ 0.5 K. At 46 hPa, median differences of the reanalyses with respect to the reference temperatures all depart from zero, and their scatter falls within the range of about 0.6 K for LIQ and 0.5 K for ICE. Although the biases are larger for LIQ, their standard deviations (~ 0.6 K) are smaller than those for ICE (~ 0.7 K). To put these comparisons with LIQ and ICE reference temperatures into context, temperature measurements from long-duration balloon flights have typical nighttime biases of 0.5 K, precisions of 0.4 K (*Pommereau et al.*, 2002), and measured standard deviations of 1.0 K to 1.3 K for temperature differences with respect to ECMWF operational analyses (*Knudsen et al.*, 2002).

Finally, the polar temperatures from recent full-input reanalyses are in much better agreement than were the reanalyses from previous decades. As a consequence, strong justification is warranted in modeling studies that seek to match the simulated chlorine activation and/or ozone loss by making arbitrary systematic adjustments to the reanalysis temperatures of 1 - 2 K or higher.

10.8 Chemical modeling diagnostics

In this section, we compare simulated fields of nitric acid, water vapour, several chlorine species, nitrous oxide, and

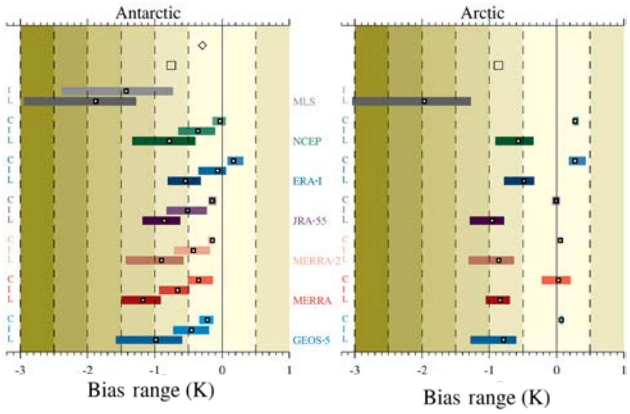


Figure 10.16: *Temperature bias ranges of the reanalyses (MERRA, MERRA-2, JRA-55, ERA-Interim, and NCEP (CFSR/CFSv2)), and MLS, relative to the LIQ (labelled "L" on the y-axis) and ICE ("I" on the y-axis) equilibrium references, and COSMIC ("C" on the y-axis), for Antarctic (left) and Arctic (right) winters 2008-2013, poleward of 60°, and for pressure levels from 68 hPa to 21 hPa. The bias ranges are obtained from the extrema of the yearly mean bias values over 68-21 hPa weighted by the yearly standard deviations. The horizontal coloured bars indicate the ranges of the minimum to maximum bias for MLS and each of the reanalyses as indicated in the legend. White squares with black border indicate the mean bias over 68-21 hPa. Open square (diamond) symbols indicate the mean values of* $\overline{\Delta T_{LIQ}} - \overline{\Delta T_{COSMIC}}$ *(*$\overline{\Delta T_{ICE}} - \overline{\Delta T_{COSMIC}}$*) over 68-21 hPa. There are insufficient statistics for a reliable comparison with the ICE reference in the Arctic. Note that MLS has not been compared directly to COSMIC. Background shading indicates 0.5 K increments in the bias range. From Lambert and Santee (2018).*

ozone with those observed by Aura MLS and Envisat MIPAS. As noted in *Section 10.1*, chemical model simulations provide a means of assessing the net effects of multiple (in some cases competing) spatially and temporally varying differences between reanalyses. Thus results from a chemical model driven by different reanalyses may provide further insights into which reanalyses are well suited for polar process studies beyond those obtained through simpler, more direct diagnostics.

It must be borne in mind that some of the differences seen between observed and simulated fields may arise through inherent shortcomings in the formulation or setup of the chemical model itself and not through the meteorological fields being used to drive it. Ideally the same set of simulations should be performed by a range of chemical models and the full suite of runs examined to elucidate how underlying differences in their chemistry and physics manifest when different models are forced by different reanalyses. It was not practical to carry out such an extensive investigation in the context of this project, however, and only a single model is applied here. Nevertheless, because our purpose is intercomparison of the reanalyses (not model evaluation or scientific study of specific polar processing events), using multiple reanalyses to drive the same (albeit imperfect) chemical model over the same time period can illuminate how various differences between reanalyses combine to affect simulated trace gas fields.

We have chosen to use the chemistry transport model developed for the Belgian Assimilation System for Chemical ObsErvations (BASCOE). The BASCOE CTM has been validated against several satellite data sets and shown to successfully reproduce observed stratospheric composition in the polar regions (*Huijnen et al.*, 2016; *Lindenmaier et al.*, 2011; *Daerden et al.*, 2007); its performance is thus adequate for this reanalysis intercomparison study. Experiments were performed for five reanalyses: MERRA, MERRA-2, JRA-55, ERA-Interim, and CFSR/CFSv2.

10.8.1 Details of the BASCOE system and experimental setup

We use here the BASCOE CTM (version 6.2). The transport module and its setup are described in *Chapter 5* for the mean age of air simulations (for full details see *Chabrillat et al.*, 2018). It is based on the Flux-Form Semi-Lagrangian (FFSL) scheme (*Lin and Rood*, 1996). We use a low-resolution latitude-longitude grid (2° × 2.5°) that is common to all five experiments, but the vertical grids differ between experiments because each of them retains the native sigma-pressure vertical grid of its input reanalysis. All five reanalyses have been expressed as spherical harmonics and identically truncated to wavelength number 47 before derivation of the wind fields on the common horizontal grid. On the basis of mass conservation, vertical velocities (expressed as $\omega = dp/dt$) are derived afterwards within the FFSL module. Hence, in contrast to diabatic transport models that use the isentropic coordinate, such as SLIMCAT (*Chipperfield*, 2006) or CLaMS (*Grooß et al.*, 2011), this transport scheme does not compute heating rates, nor does it read them from the input reanalysis. Diabatic transport models have distinct advantages in modeling the stratosphere, because the use of isentropic levels reduces spurious vertical mixing by providing a true separation between horizontal and vertical motion (*Chipperfield et al.*, 1997). These issues were especially important with the previous generation of reanalyses because they often suffered from spurious surface pressure increments caused by data assimilation (*Pawson et al.*, 2007; *Meijer et al.*, 2004). In modern kinematic transport models, these issues are mitigated by pre-processing the input wind fields in order to correct for the small inconsistencies in the pressure tendency compared with the divergence fields (*Chabrillat et al.*, 2018; *Chipperfield*, 2006). While the simpler kinematic approach used by the BASCOE CTM may provide less realistic simulations of transport in and around the polar vortex (*Hoppe et al.*, 2016), it should ease the interpretation of results in the context of S-RIP because the temperature fields impact only the chemical and microphysical processes, while the surface pressure and wind fields determine entirely the transport processes. This approach also allows us to keep the sigma-pressure vertical grid of each input reanalysis, providing an intercomparison that takes their different vertical resolutions into account.

The photochemical scheme and parameterization of PSCs are described by *Huijnen et al.* (2016). The kinetic rates for heterogeneous chemistry are determined by the parameterization of *Fonteyn and Larsen* (1996), using classical expressions for the uptake coefficients on sulphate aerosols (*Hanson and Ravishankara*, 1994) and on PSCs (*Sander et al.*, 2000). The surface area density of stratospheric aerosols uses an aerosol number density climatology based on SAGE II observations (*Hitchman et al.*, 1994). Ice PSCs are presumed to exist at any grid point in the winter/spring polar regions where water vapour partial pressure exceeds the vapour pressure of water ice (*Murphy and Koop*, 2005). NAT PSCs are assumed when the HNO_3 partial pressure exceeds the vapour pressure of condensed HNO_3 at the surface of NAT PSC particles (*Hanson and Mauersberger*, 1988). The surface area density is set to $2\times10^{-6}\,cm^2cm^{-3}$ for ice PSCs and $2\times10^{-7}\,cm^2cm^{-3}$ for NAT PSCs. The sedimentation of PSC particles causes denitrification and dehydration. This process is approximated by an exponential decay of HNO_3 with a characteristic timescale of 20 days for grid points where NAT particles are supposed to exist, and an exponential decay of HNO_3 and H_2O with a characteristic timescale of 9 days for grid points where ice particles are supposed to exist (*Huijnen et al.*, 2016).

Experiments were performed for five reanalyses: MERRA, MERRA-2, JRA-55, ERA-Interim, and CFSR/CFSv2. The simulations span 13 months, from March 2009 to April 2010, to allow examination of chemical processing during one Antarctic and one Arctic winter. Each experiment starts on 1 March 2009 from the CTM simulation described in *Huijnen et al.* (2016), *i.e.*, a higher-resolution ($1.125° \times 1.125°$) experiment that started one year earlier from a combined analysis of Envisat MIPAS and Aura MLS chemical observations (*Errera et al.*, 2008, 2016, 2019) and was driven by ECMWF analyses similar to ERA-Interim. We first investigate the period from May 2009 through October 2009 over the Antarctic (*Section 10.8.3*), followed by the period from November 2009 through April 2010 over the Arctic (*Section 10.8.4*). Hence, the five experiments were spun up with their respective input reanalyses during the two months of March and April 2009.

10.8.2 Analysis approach

The BASCOE model output is interpolated to the geolocations and times of the satellite observations to facilitate comparisons, and both observations and model results are vertically interpolated onto a potential temperature grid with 10 K spacing. Daily averages of observed and modeled quantities are calculated over two equivalent latitude ranges representing the inner vortex core (75°-90°) and outer vortex collar region (60°-75°) in both hemispheres. Temperature, PV, equivalent latitude, and other parameters at the MLS geolocations are obtained from the MLS DMP files; as described in *Section 10.2.1*, DMP files containing associated meteorological information at the MLS measurement locations are available for all five reanalyses considered here. Thus results from the model simulations driven by each of the five reanalyses are compared to corresponding MLS data analyzed using the respective set of DMPs. Since no DMP files are available for MIPAS data, they are analyzed in terms of geographic latitude. In the following two subsections, we examine two case studies: the 2009 Antarctic winter (May through October) and the 2009/2010 Arctic winter (November through April).

10.8.3 Case study: 2009 Antarctic winter

Model/measurement comparisons of the evolution of various chemical constituents are shown as a function of potential temperature (300 K to 800 K) for the inner and outer vortex in **Figures 10.17** and **10.18**, respectively. Observations (from MLS and MIPAS) are presented in the left-hand column, with corresponding model results in the middle column. The simulation driven by MERRA-2 is selected as a representative example. The right-hand column shows the reanalysis ranges, that is, the differences between the maximum and the minimum from the full set of five simulations, to illustrate the spread between the runs forced by the different reanalyses.

Overall, the model performs well in reproducing the observed stratospheric conditions inside the vortex during Antarctic winter, as has been demonstrated previously (*Huijnen et al.*, 2016). There are, however, areas where agreement is less satisfactory (**Figure 10.17**). Again, it should be emphasized that some of the model/measurement discrepancies may be attributable to fundamental deficiencies in the formulation of the model itself, independent of the reanalysis being used to drive it. The observations indicate higher peak HNO_3 abundances than in the model (forced by MERRA-2), so although sequestration of HNO_3 in PSCs is fairly well modeled, the simulated values start off from a lower maximum. Renitrification, or the redistribution of HNO_3 to lower altitudes via evaporation of sedimenting PSC particles, is visible in the observations but not represented in the model, likely because the simulated HNO_3 is permanently removed rather than being transported downward by sedimentation. The decline in HCl, indicative of chlorine activation, begins about two weeks earlier, progresses more rapidly, and is more complete in the model, such that nearly all of the available HCl in the Antarctic inner vortex is consumed by late May. This overestimation of HCl depletion at the beginning of Antarctic winter is in contrast to results from other chemical models, which typically see a delay in chlorine activation relative to observations (*e.g.*, *Wohltmann et al.*, 2017; *Grooß et al.*, 2018, and references therein). Because the deep vortex core is essentially in darkness until mid-winter, active chlorine remains tied up in forms other than ClO (primarily the ClO dimer, ClOOCl, as well as Cl_2

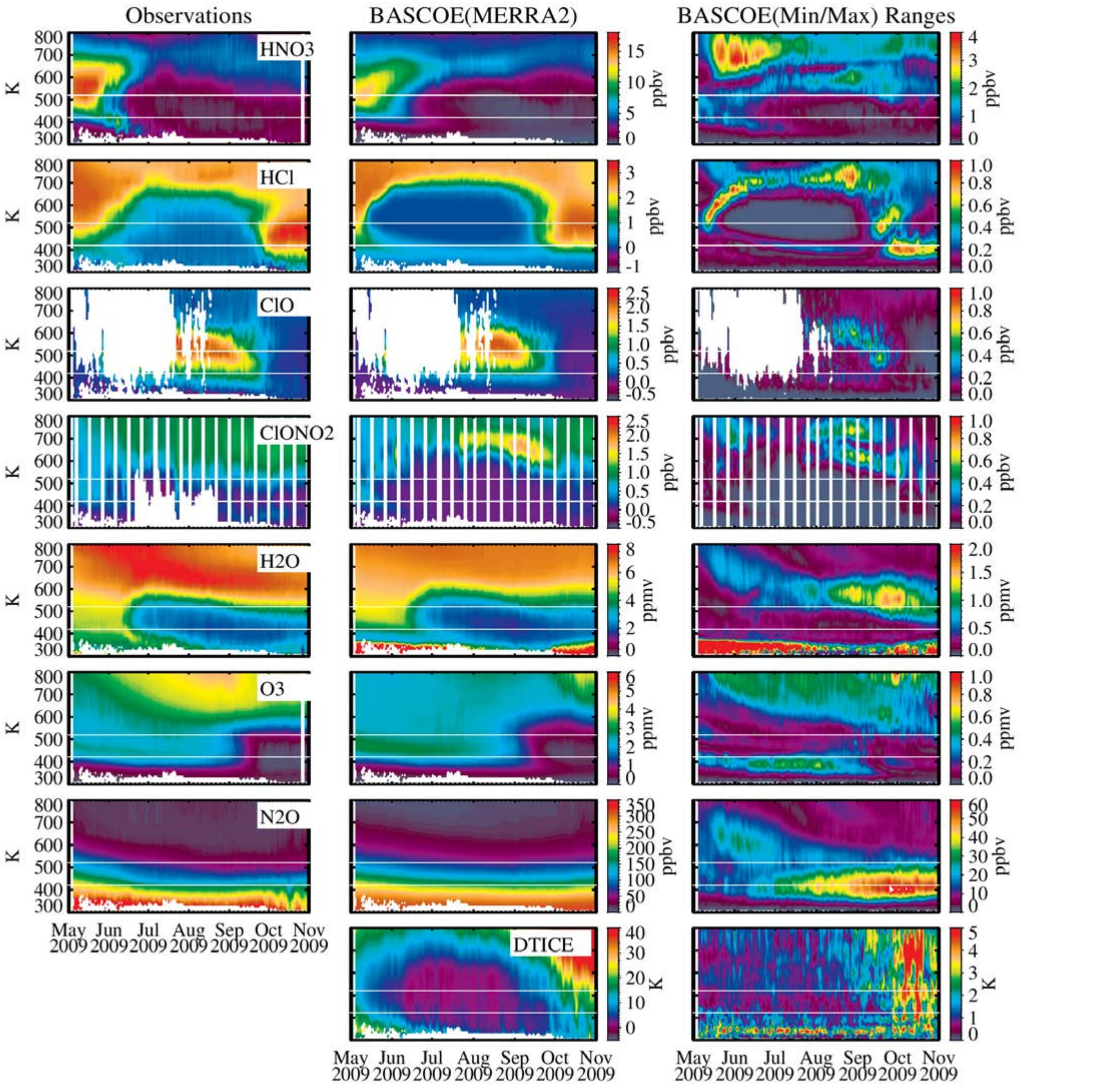


Figure 10.17: *Time/height cross sections of daily averages calculated over the region of the inner vortex (75 ° - 90 ° S) for the 2009 Antarctic winter (May through October). The first column shows measurements of HNO_3, HCl, ClO, $ClONO_2$, H_2O, O_3, and N_2O. The $ClONO_2$ data are from MIPAS, averaged over geographic latitude; all other species were measured by Aura MLS and averaged over equivalent latitude. The middle column shows corresponding results from the BASCOE simulation driven by MERRA-2. The third column shows the range (maximum – minimum) of model simulations driven by all five reanalyses. The bottom row of the middle column shows the time/height cross section of the MERRA-2 temperature deviation from the ice frost point calculated using MLS H_2O (DTICE), and the bottom row of the third column shows the range in reanalysis temperatures. Blank spaces arise for several reasons: (1) MIPAS is observing in another mode or data are otherwise missing, e.g., MIPAS measurements are affected by the presence of PSCs, (2) the potential temperature surface falls below the lowest recommended MLS retrieval pressure level, or (3) no measurements in sunlight are available in polar night. The latter situation pertains to the ClO panels; because active chlorine is converted to ClO in daylight, only data from the ascending portions of the orbit are used for ClO (see Section 10.2.1). Horizontal white lines mark the 420 K and 520 K potential temperature surfaces, which are examined in detail in subsequent figures.*

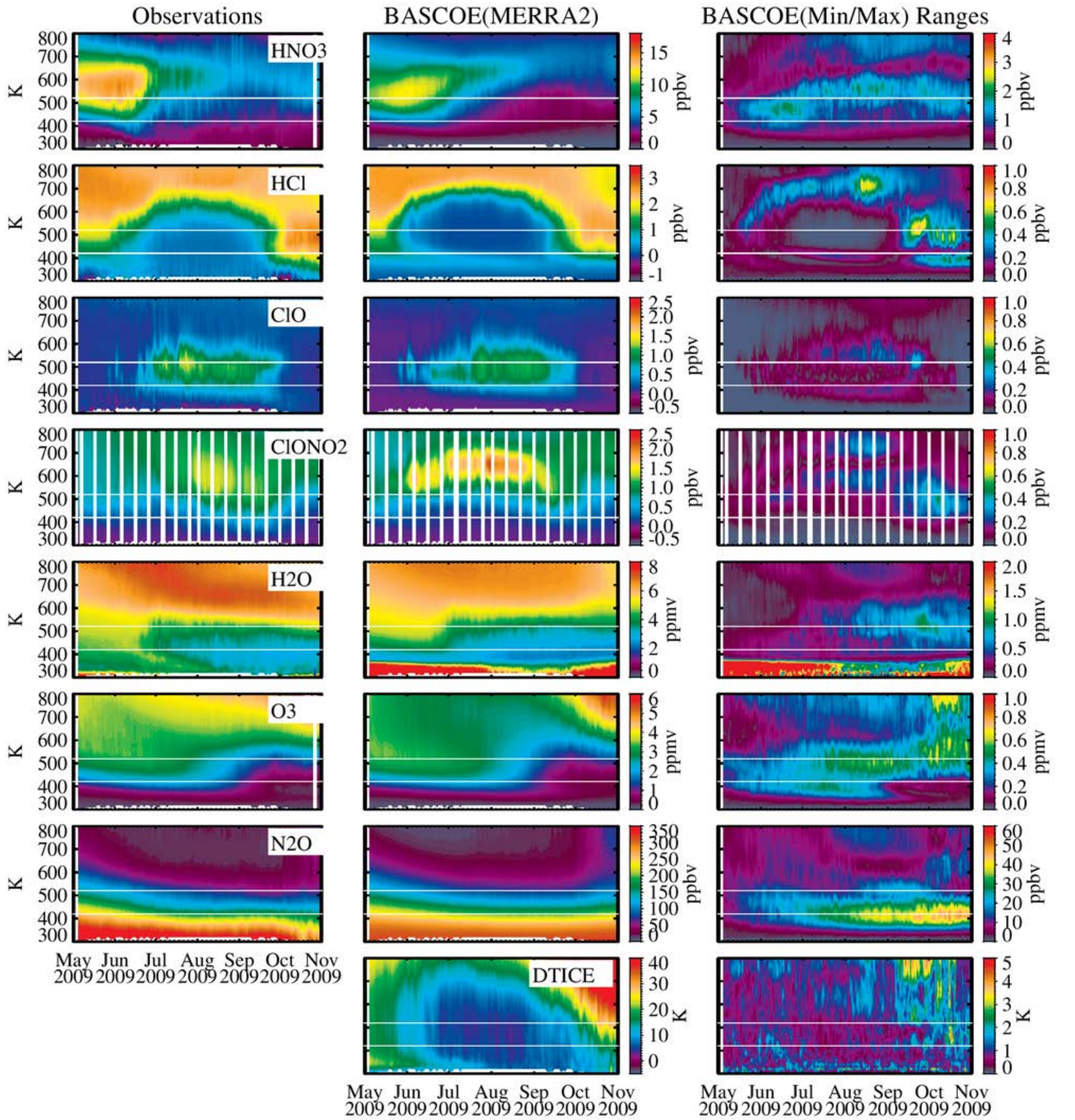


***Figure 10.18:** As in **Figure 10.17** but for the region of the outer vortex (60 ° - 75 ° S).*

and HOCl); the resulting patchiness complicates interpretation of the ClO panel. Nevertheless, **Figure 10.17** shows that the modeled and measured morphology and magnitude match fairly closely. As was the case for HCl, modeled $ClONO_2$ depletion is spatially and temporally more extensive than observed. The model/measurement discrepancies in PSC processes and chlorine activation may be attributable to the relatively simple PSC parameterizations implemented in the BASCOE model (*e.g., Errera et al.*, 2019).

Examination of N_2O as a tracer of stratospheric air motions, along with H_2O and O_3 in the non-dehydrated, non-ozone-depleted regions where those species act as conserved tracers (*e.g.*, above 600 K), suggests that either confined diabatic descent inside the winter polar vortex is weaker in the simulations than indicated by the observations, or mixing across the vortex edge is greater. The latter possibility would be consistent with the findings of *Hoppe et al.* (2014), who reported that model simulations based on the FFSL transport scheme (see *Section 10.8.1*) tend to underestimate the strength of the transport barrier at the edge of the polar vortex compared to observations. On the other hand, some models nudged to reanalyses have been shown to have difficulty in accurately reproducing the strength of the diabatic descent inside the winter polar vortices (*e.g.*, *Froidevaux et al.*, 2019; *Khosrawi et al.*, 2018). Disentangling the contributions of advection and mixing in the model representation of

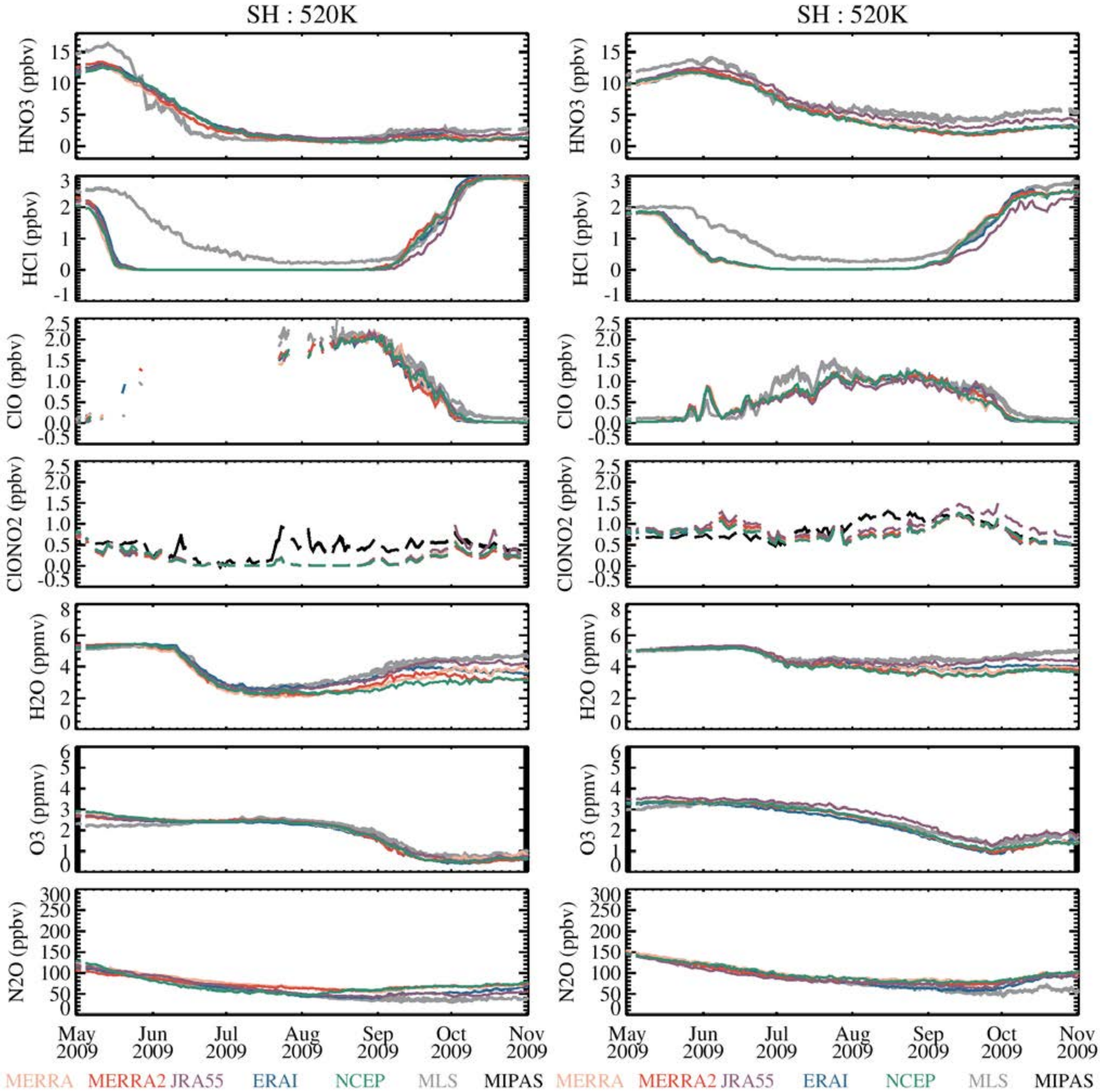


Figure 10.19: *Time series of daily averages calculated over the region of the inner vortex (75°-90°S, left) and outer vortex (60°-75°S, right) at 520K. MIPAS $ClONO_2$ data (black lines) are averaged in geographic latitude bands. For all other species, averages of MLS data are calculated over equivalent latitude based on all five reanalyses (grey lines, largely indistinguishable). Corresponding BASCOE results for each of the simulations driven by the reanalyses (MERRA, MERRA-2, JRA-55, ERA-Interim, and NCEP (CFSR/CFSv2)) are colour-coded as indicated in the legend.*

transport is challenging, because both processes suffer from large and competing uncertainties that depend not only on the input reanalysis but also on the offline (or nudged) transport model (*Minganti et al.*, 2020, and references therein).

Differences between the individual realizations of the model are captured in the right-hand column of **Figure 10.17**. For the most part, the model runs forced by the different reanalyses produce very similar results. However, substantial disparities are evident in a few places, in particular in those regions where the gradients are largest.

Figure 10.18 paints a generally similar picture for the outer vortex. One notable difference is that the collar region is more consistently in sunlight, affording greater coverage of daytime ClO abundances. Here it is again clear in the HCl and $ClONO_2$ that the modeled chlorine activation is stronger and more extensive than that observed; the ClO also indicates slightly earlier activation in the model, although the peak ClO abundances reached in midwinter are slightly smaller than those recorded by MLS. In terms of the reanalysis ranges, again the largest spread between simulations driven by different reanalyses is found where

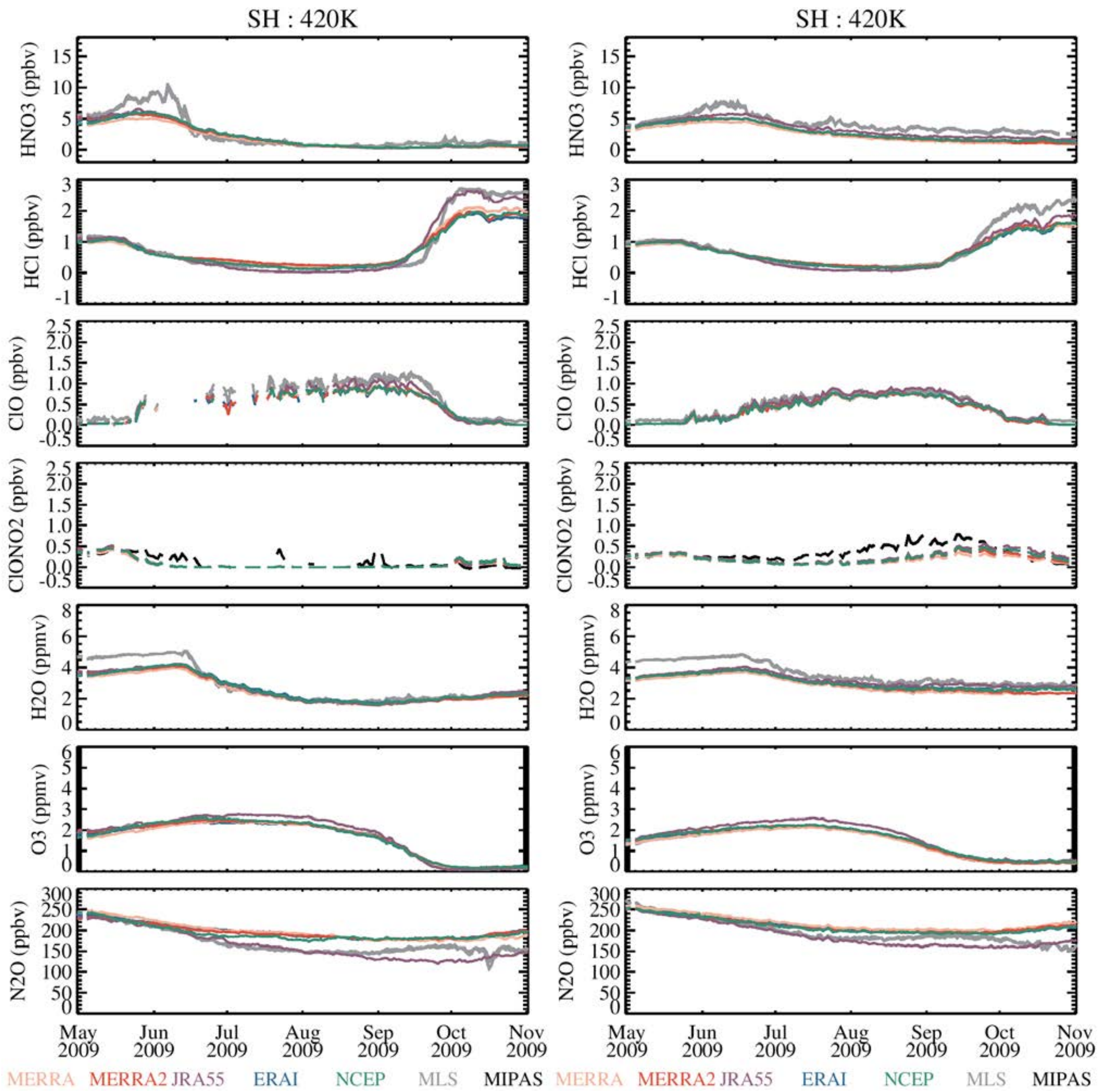


Figure 10.20: *As for **Figure 10.19** but at 420 K.*

the gradients are largest, but in general the discrepancies are smaller than those in the vortex core.

Figures 10.19 and **10.20** show slices through the data at the 520 K and 420 K potential temperature levels, respectively, marked by the horizontal white lines in the cross section plots (**Figures 10.17** and **10.18**). Averages for the inner vortex core are shown on the left and for the outer vortex collar region on the right. Conclusions similar to those discussed above can be drawn from **Figure 10.19**. Initially HNO_3 is ~20 % lower in the model than observed, but thereafter its sequestration in PSCs is reasonably well reproduced in all runs. As before, the most dramatic differences between the observed and simulated behavior are seen in HCl in early and mid-winter. However, the simulation lines are fairly tightly clustered, with the exception of JRA-55 at the end of the season, implying that the discrepancies arise from issues in the underlying model and not the reanalyses being used to force it. Although dehydration is simulated well by the model, at the end of winter H_2O recovers to larger values in the observations than in the model runs, again except for JRA-55, which agrees well with MLS. These results for water vapour suggest that the model does not overestimate the mixing across the polar vortex edge, because such erroneous mixing would lead to overestimations of H_2O inside the vortex (unless ongoing dehydration processes promptly removed the excess water vapor being mixed in throughout mid-winter), whereas the model delivers in all cases an underestimation.

After the month of July, no simulation reproduces the observed low values in N_2O in either the core or the

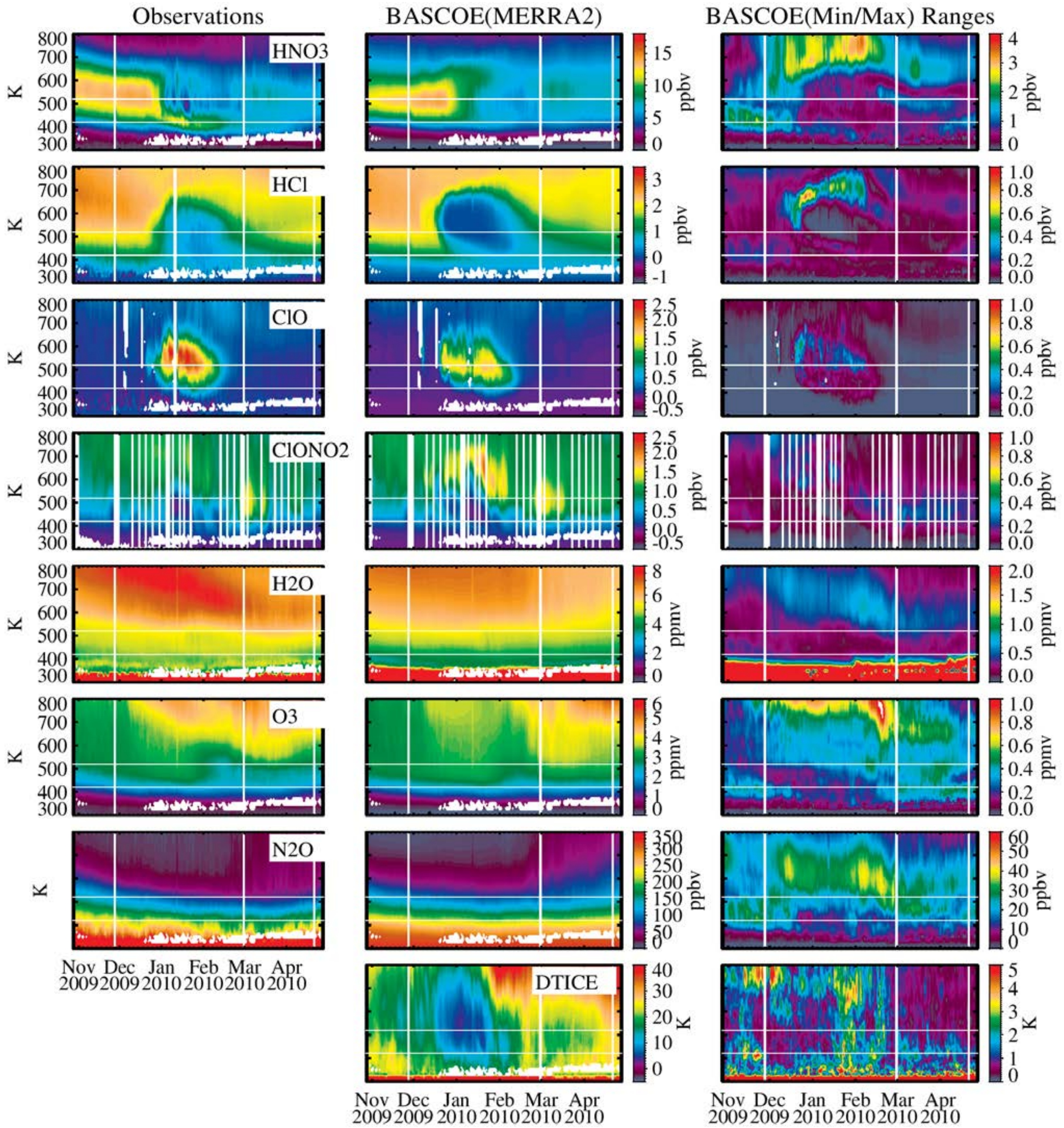


Figure 10.21: *As in* ***Figure 10.17*** *but for the 2009/2010 Arctic winter (November through April).*

collar of the vortex, although in the vortex core ERA-Interim and JRA-55 N_2O abundances are closer to those measured by MLS than the other reanalyses. Erroneous mixing across the vortex edge is probably not responsible for the overestimation in the simulations using MERRA, MERRA-2 and CFSR/CFSv2 because this issue is expected to have a similar impact in all simulations. Rather, the N_2O results at 520 K suggest that the model underestimates the strength of confined diabatic descent in the vortex interior, except for JRA-55 and ERA-Interim during the first half of the season. After the month of August, the rate of N_2O decrease is underestimated in all five simulations. As noted above, inaccurate depiction of downward transport is a known model problem. The underestimation of N_2O decrease may be interpreted as an underestimation of the diabatic descent in the vortex, but recent calculations of the Transformed Eulerian Mean budget of N_2O, using the same reanalyses and CTM, suggest that it may instead be due to underestimation of the impact of horizontal mixing with the BASCOE CTM (*Minganti et al.*, 2020). The representation of descent is also a serious issue at 420 K (**Figure 10.20**), where the contrast between JRA-55 and the other reanalyses

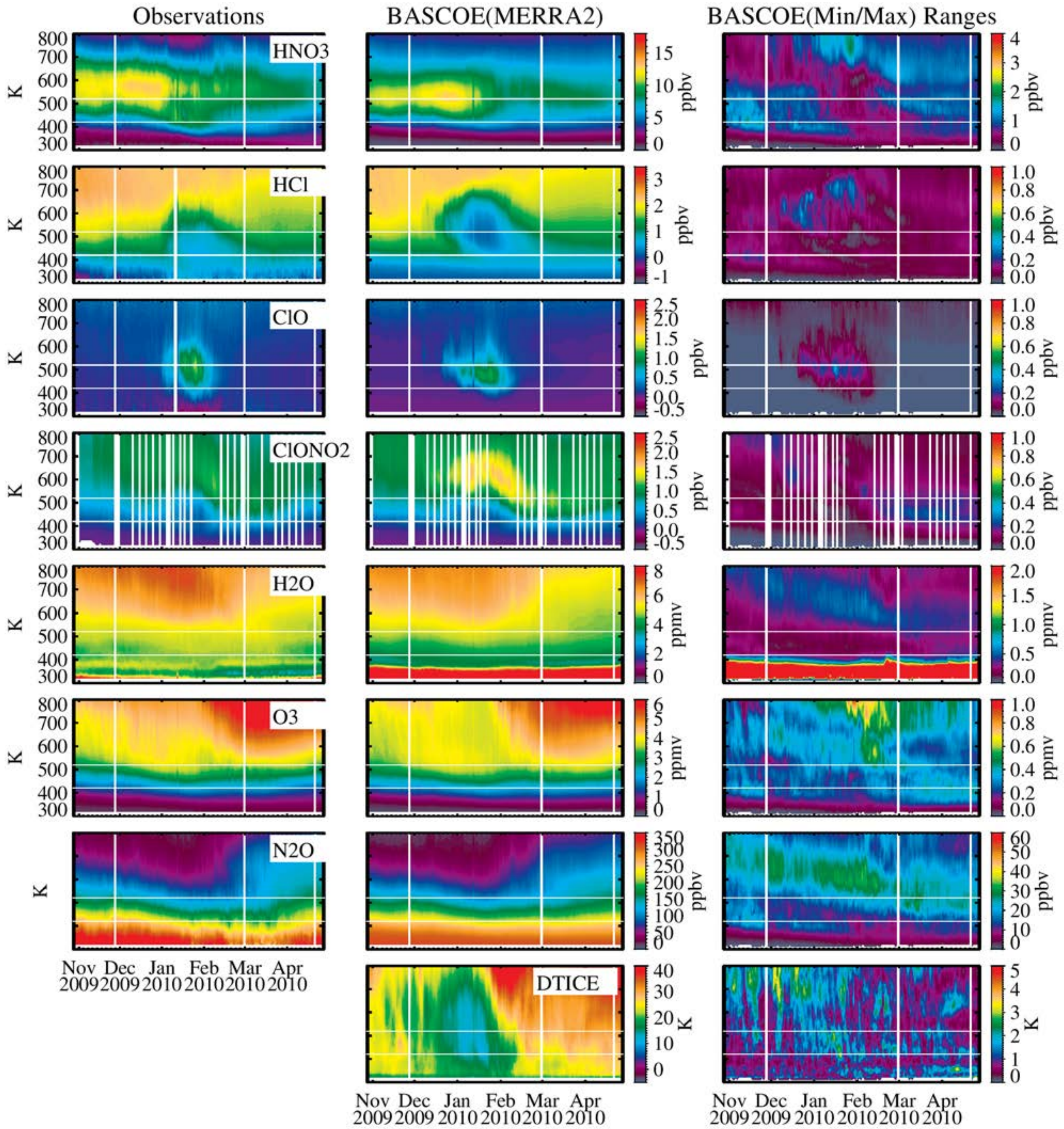


Figure 10.22: *As in* ***Figure 10.18*** *but for the Arctic.*

is even greater. At this level, however, JRA-55 seems to suggest slightly stronger descent than the MLS observations throughout the vortex, whereas the other reanalyses indicate weaker descent, as at 520 K. Other noteworthy results at 420 K include the stronger renitrification seen by MLS than any of the simulations. Faithful reproduction of observed denitrification and renitrification features has often been a challenge for models (*e.g.*, *Braun et al.*, 2019; *Khosrawi et al.*, 2018); as suggested earlier, the lack of renitrification in these BASCOE runs is likely attributable to the HNO_3 being permanently removed rather than transported downward by sedimentation. In addition, the HCl increase in October is much larger (and closer to that observed) with JRA-55 than with any other reanalysis.

The areas of largest disagreement between measured and modeled behavior are generally consistent across all five simulations, suggesting the presence of underlying problems with BASCOE not associated with the particular reanalysis being used to force the model. However, considerable spread between the simulations becomes evident in middle and late winter, particularly in the H_2O and N_2O fields. Apparently deficiencies in modeled diabatic descent within the vortex, together with other shortcomings in model physics

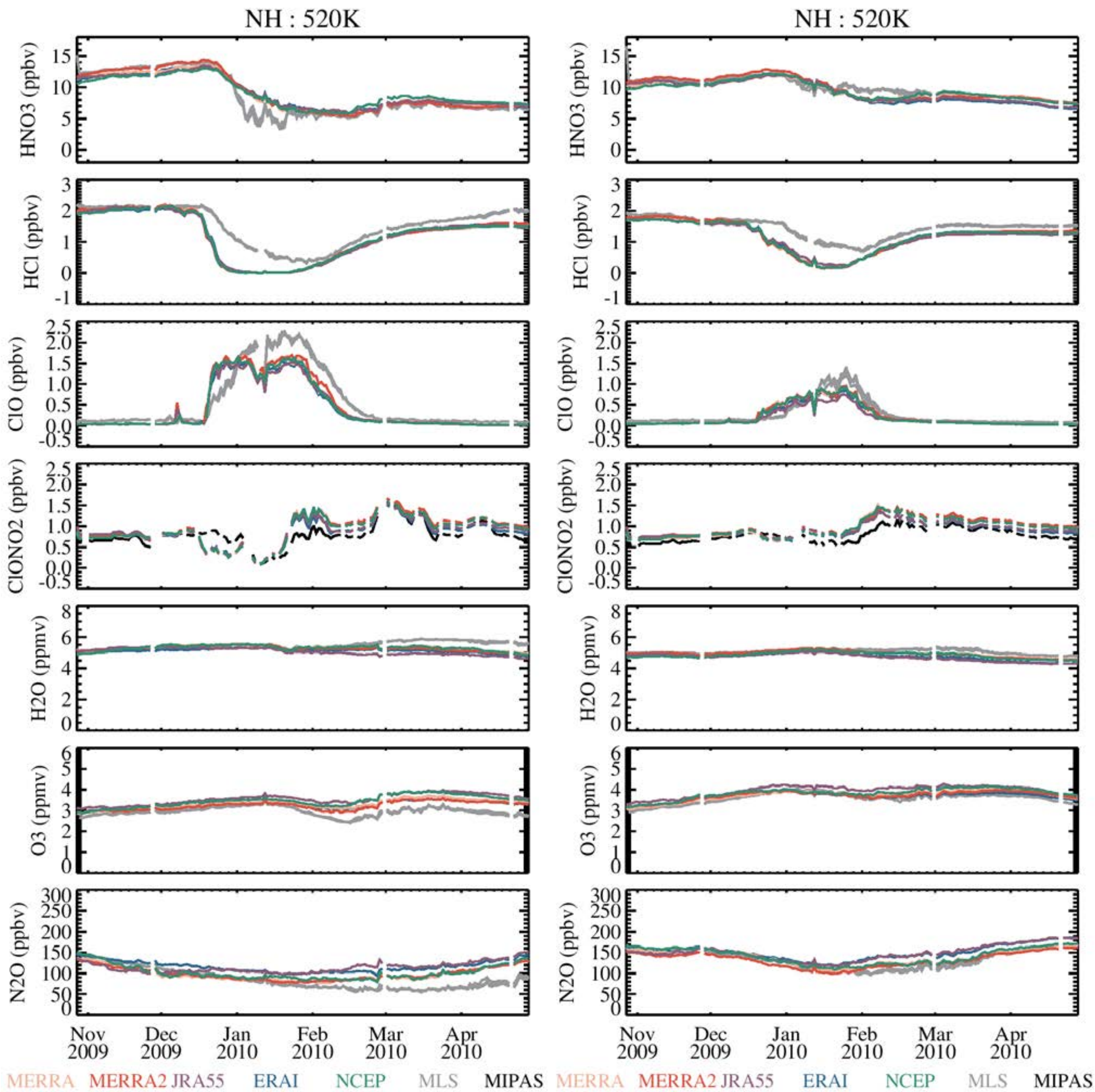


Figure 10.23: *As for* ***Figure 10.19*** *but for the Arctic.*

or chemistry schemes, manifest to varying degrees depending on the driving reanalysis, and these issues give rise to the divergence in simulated behavior seen in **Figures 10.19** and **10.20**.

10.8.4 Case study: 2009/2010 Arctic winter

Similar to the analysis for the Antarctic, cross section comparisons of the evolution over the 2009/2010 Arctic winter (November through April) are shown in **Figures 10.21** and **10.22** for the inner and outer vortex, respectively, and slices through the data at 520 K and 420 K are shown in **Figures 10.23** and **10.24**, respectively.

Again, the model performs well overall, although many of the same issues arise as for the Southern Hemisphere. At the beginning of the season, observed HNO_3 values are slightly larger than those modeled, although the agreement with MERRA-2 in particular is fairly good at 520 K. Here too MLS measurements indicate a clear signature of renitrification at 420 K that is lacking in any of the simulations (**Figure 10.24**). As before, HCl depletion at 520 K is substantially overestimated in all model realizations (**Figure 10.23**). In contrast to the situation in the Antarctic, however, peak ClO enhancement in the Arctic vortex core is considerably smaller and less persistent in the simulations than observed. Similar indications of weaker modeled diabatic descent inside the vortex as those discussed

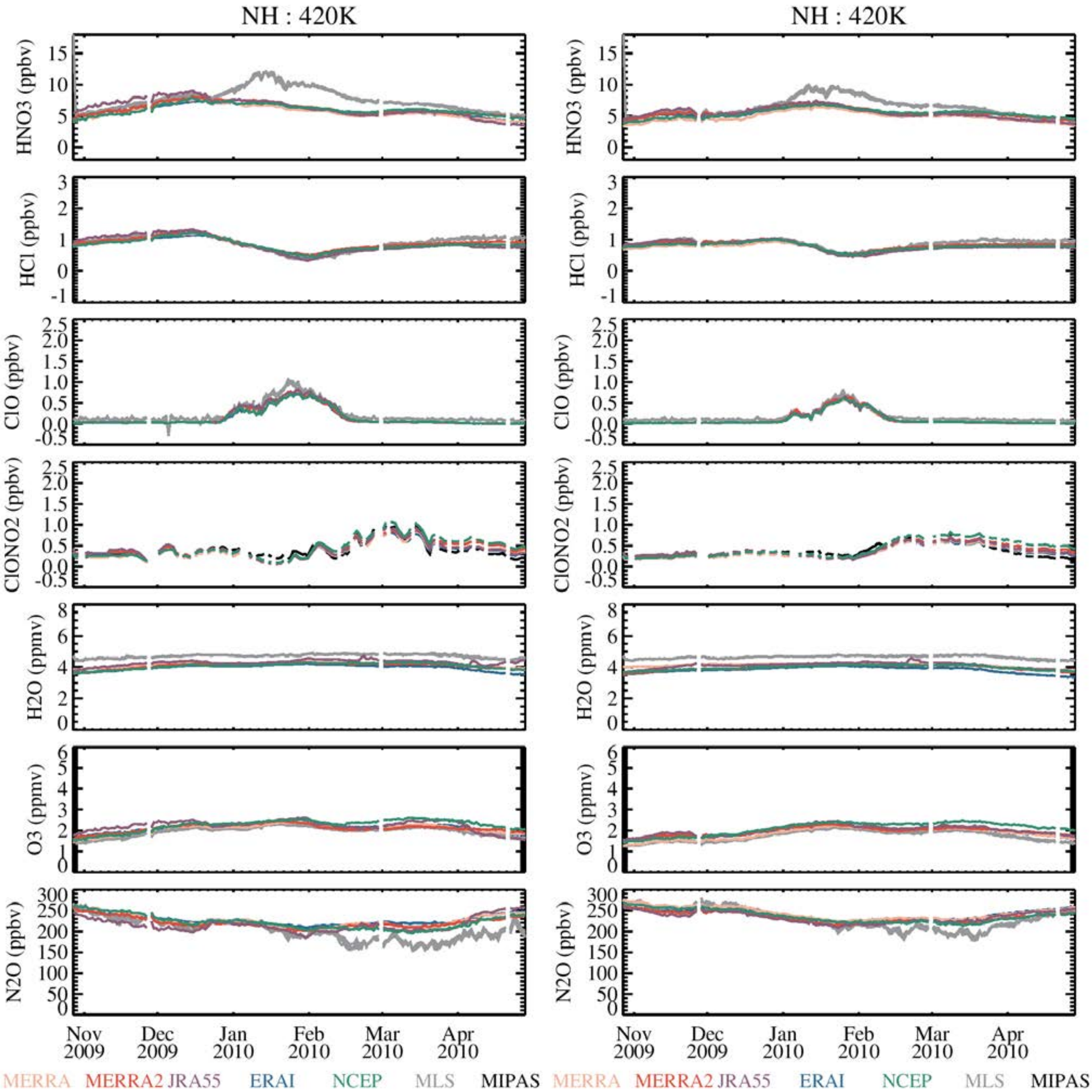


Figure 10.24: *As for* ***Figure 10.20*** *but for the Arctic.*

above are also seen at both levels in the Arctic. However, unlike in the Antarctic, where the simulations driven by JRA-55 and ERA-Interim more closely match observed behavior, in the Arctic they track the MLS measurements even less well than the other reanalyses. Overall, the maximum – minimum range of the various reanalyses is smaller for most species than that seen in the Antarctic.

10.8.5 Comparisons of chemical ozone loss

The ultimate goal of many stratospheric polar processing studies over the last several decades has been to quantify the degree of chemical ozone loss, which requires accounting for the effects of dynamics on the distribution of stratospheric ozone. Several different techniques have been developed to remove transport-induced changes in ozone and thus isolate the signature of chemical destruction (*e.g.*, *WMO*, 2007). Here we adopt the approach recently employed by *Strahan and Douglass* (2018). Ozone partial columns are calculated over the range of MLS retrieval pressure levels encompassing the majority of depletion, 261 - 12 hPa. Averages over 10-day periods in early winter (1 - 10 July) and late winter (11 - 20 September) are calculated in an attempt to reduce the effects of short-term dynamical fluctuations on ozone. The early-winter averaged partial column is then subtracted from the late-winter value to permit an estimate of chemical loss while mitigating dynamical variability to some extent. Although it is arguable whether transport variations are fully accounted for by this method, our purpose here is not to provide the most accurate quantification of chemical ozone loss, but rather to apply a convenient methodology for doing so consistently to modeled and measured fields to facilitate intercomparison of the reanalyses.

Results are shown in **Figure 10.25** for all five BASCOE simulations as well as MLS data (analyzed separately using meteorological information from the respective reanalyses), for averages over both the inner (dark shading) and outer (light shading) vortex regions. Estimates of ozone loss based on MLS are relatively insensitive to the choice of reanalysis used for interpolation of the measurements to isentropic surfaces and identification of vortex data points, although slightly larger differences between the MLS estimates for the inner and outer vortex are seen using JRA-55 and CFSR/CFSv2. In contrast, loss estimates based on the modeled O_3 fields do differ substantially, with MERRA indicating much smaller loss, closest to that calculated from MLS data. Estimates derived from the MERRA and MERRA-2 runs disagree by ~10 DU. ERA-Interim also provides relatively weak depletion, while JRA-55 indicates much stronger depletion than the other reanalyses. Thus, from a purely BASCOE-based perspective, using JRA-55 to drive the model instead of MERRA could yield ~25 DU (30 %) more chemical ozone loss in the Antarctic vortex core.

10.8.6 Discussion and implications

In summary, five recent full-input reanalyses were used to drive the BASCOE CTM, and the results were compared to satellite observations from MLS and MIPAS. The simulations spanned a full year, from May 2009 to May 2010, allowing chemical processing during winter to be examined in both hemispheres. Overall, the model reproduced the observed seasonal evolution of stratospheric constituents well, although some discrepancies between measured and modeled behavior were noted. In terms of reanalysis intercomparison, agreement between the individual model realizations was generally very good, and in most cases only small deviations between results from the different simulations were discernible. Thus the inter-simulation spread was generally smaller than the disparities between the model (regardless of how it was forced) and the observations.

Since the main areas of disagreement between measured and modeled behavior were typically replicated across all five simulations, as discussed above in connection with the case studies, they likely arise from underlying deficiencies in model physics or chemistry not associated with the particular reanalysis being used to drive the model. One notable exception was found in long-lived tracer abundances (N_2O in particular; also H_2O and O_3 to a lesser extent) in middle and late winter in both hemispheres. Although none of the model runs faithfully reproduced the observed tracer evolution in either the core or edge regions of the vortex, considerable spread between simulations developed at this time, especially in the core of the Antarctic vortex. Results from most simulations indicated weaker confined diabatic descent inside the vortex compared to observations, with the exception of the JRA-55 run in the Arctic at 420 K, which pointed to stronger descent throughout the vortex than implied by the MLS N_2O measurements. The divergence in simulated behavior suggests that issues with the modeled depiction of transport are worse for some reanalyses than others. It must be noted that these results were obtained with a kinematic transport model wherein diabatic descent is derived from the wind fields contained in the reanalyses. As noted in *Chapter 5* for age of air calculations, a different outcome may be expected with diabatic transport models that read the heating/cooling rates computed by the radiative transfer models of the reanalysis systems (see also *Martineau et al.*, 2018, and *Section 10.5*).

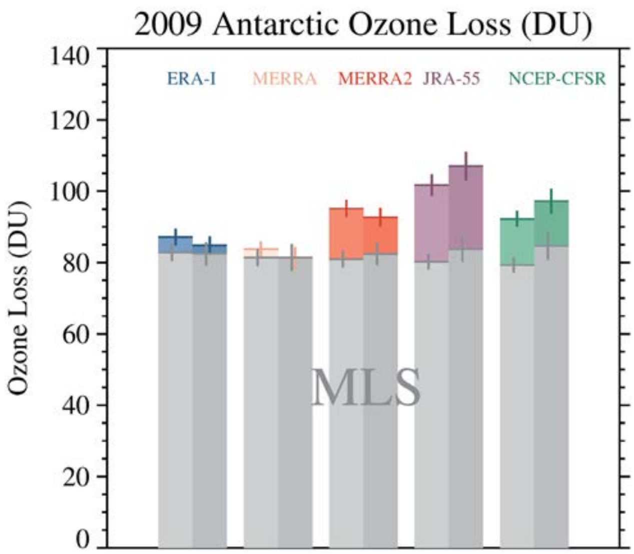


Figure 10.25: *Estimates of ozone loss for the 2009 Antarctic season (see text for details about the approach) for equivalent latitude ranges representing the outer vortex collar region (light shading: 60°-75°S) and the inner vortex core (dark shading: 75°-90°S), calculated from the BASCOE model output driven by each reanalysis (ERA-Interim, MERRA, MERRA-2, JRA-55, and NCEP (CFSR/CFSv2); coloured shading) and from MLS observations (grey shading). MLS data are analyzed separately for comparison to each simulation using meteorological information from the respective reanalysis; thus the five grey bars differ slightly. Note that in almost all cases, modeled ozone loss estimates exceed those derived from MLS measurements and are thus plotted as increments above the MLS values. Error bars represent the standard errors of the mean ozone losses calculated from the model data (coloured) and from the MLS observations (grey).*

10.9 Summary, key findings, and recommendations

This chapter employs an extensive set of diagnostics of relevance to polar chemical processing and dynamics to evaluate and intercompare recent full-input reanalyses, including MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2. The GMAO GEOS-591 operational analysis, a stable system providing consistent meteorological fields used by NASA satellite instrument teams, is also examined for PSC thermodynamic-consistency diagnostics.

To provide an overview and set the stage for more detailed intercomparisons, time series (2008 - 2013) of daily mean polar cap temperature differences at 46 hPa relative to ERA-Interim are presented first to examine seasonal (and interannual) variations. The rest of the chapter focuses on winter conditions. The comprehensive suite of polar temperature and vortex diagnostics considered in both hemispheres includes daily minimum lower stratospheric temperature poleward of 40 °, area and volume of stratospheric air with temperatures below PSC existence thresholds, maximum latitudinal gradients in PV (a measure of vortex strength), area of the vortex exposed to sunlight each day, vortex breakup dates, and polar cap average diabatic heating rates.

Comparisons with superpressure balloon measurements made during the Antarctic Concordiasi campaign in September 2010 to January 2011 are used to quantify biases in reanalysis temperatures and horizontal winds and the growth of errors along 15-day trajectories calculated using them, as well as to assess the capacity of the reanalyses to represent small-scale atmospheric fluctuations. The accuracy and precision of the reanalysis temperatures are evaluated through comparisons with both COSMIC GNSS-RO temperatures and independent absolute temperature references derived from theoretical considerations of PSC formation defined by the equilibrium thermodynamics of STS and water ice clouds. The thermodynamic-consistency diagnostics rely on near-simultaneous and colocated measurements of PSC classifications and gas-phase nitric acid and water vapour from CALIPSO CALIOP and Aura MLS satellite measurements, respectively.

Finally, because chemical model simulations synthesize the interplay among the spatially and temporally varying differences between reanalyses and exemplify how their net effects impact the bottom-line conclusions of typical real-life studies, simulated fields of nitric acid, water vapour, several chlorine species, nitrous oxide, and ozone are compared with those observed by Aura MLS and Envisat MIPAS. The key findings of this work, along with recommendations that follow from them, are summarized below.

Key findings

- In both hemispheres, differences in lower stratospheric daily polar cap averaged temperatures between ERA-Interim and other recent full-input reanalyses display annual cycles, with mainly positive deviations from ERA-Interim in summer and mainly negative deviations in winter. Thus intercomparisons of the same reanalyses undertaken for different time periods could indicate temperature deviations of roughly the same magnitude but opposite sign. Largest differences reach ~ 1 K in the Antarctic and ~ 0.5 K in the Arctic.

- Polar winter temperatures from recent full-input reanalyses are in much better agreement in the lower and middle stratosphere than were those from older reanalysis systems in common use in prior decades.

- In the Southern Hemisphere especially, a dramatic convergence toward better agreement between the reanalyses is seen after 1999 (when a major change in the input data occurred; see *Section 10.4* for more details). Average absolute differences in wintertime daily minimum temperatures poleward of 40°S from the reanalysis ensemble mean (REM) have been reduced from over 3 K prior to 1999 to generally less than 0.5 K in the most recent decade, and average differences from the REM in the area with temperatures below PSC thresholds have been reduced from over 1.5 % of a hemisphere to less than about 0.5 %. Other polar temperature and vortex diagnostics suggest a more complex picture, showing similar improvements for some reanalyses but persistent differences for others. Although convergence toward better agreement is also apparent in the Northern Hemisphere, the changes are less dramatic there because reanalysis temperatures are more consistent throughout the whole comparison period (1979 - 2015) than they are in the Southern Hemisphere.

- For many polar temperature and vortex diagnostics, the reanalyses generally agree better in the Antarctic than in the Arctic. The extremely cold conditions and relatively small interannual variability in the Antarctic mean that winter seasons tend to have similar polar processing potential and duration every year, and thus the sensitivity to differences in meteorological conditions among the reanalyses is low. In contrast, the generally warmer and more disturbed vortex and large interannual variability of Arctic winters lead to conditions that are frequently marginal, with temperatures hovering around PSC existence thresholds, and thus the sensitivity to reanalysis differences is high.

- Comparisons of polar-cap averaged diabatic heating rates in the lower stratosphere of both hemispheres show that MERRA-2, ERA-Interim, JRA-55, and CFSR give consistent results for the climatology and day-to-day evolution at pressures greater than about 20 hPa and should generally be suitable for polar processing studies. In most cases, the range of differences among the reanalyses is less than 0.5 K/day in the cold season, representing no more than about 10 - 15 % differences. These differences appear primarily as biases between the time series rather than as a failure of any reanalysis to capture day-to-day variations.

- Reanalyses differ in spatial and temporal resolution, and truncation of the models is an important factor in how well they represent small-scale fluctuations such as gravity waves. Comparisons of a subset of full-input reanalyses (ERA-Interim, MERRA, MERRA-2) with long-duration balloon observations in the Antarctic show that they reproduce the temperature and horizontal wind fluctuations of the balloons at about the 30 % level; thus a significant portion of the atmospheric gravity wave spectrum is not captured by the reanalyses. A case study of a mountain wave event shows that the temperature amplitudes resolved by the reanalyses can vary by up to a factor of two in extreme circumstances, with the differences being well correlated with the spatial resolutions of the reanalyses.

- Trajectory calculations from a Lagrangian transport model based on a subset of full-input reanalyses were evaluated using long-duration balloon observations in the Antarctic over the September 2010 to January 2011 period; relative horizontal transport deviations of 4 - 12 % and error growth rates of 60 - 170 km day^{-1} over 15-day trajectories were found for the different reanalyses.

- GNSS-RO data are not assimilated in MERRA, which consequently shows larger biases (of as much as ~ 0.5 K at 68 hPa) with respect to COSMIC temperatures compared to the other reanalyses, which do assimilate the GNSS-RO data. Differences between MERRA and COSMIC temperatures are particularly large in the Antarctic, where there is a relative paucity of conventional measurements.

- Absolute temperature references are derived from the thermodynamic equilibrium properties of certain types of PSCs and are completely independent of any data that is assimilated by the reanalyses. The reanalysis temperatures are found to be lower than these absolute reference points by up to 1 K.

- In poth polar regions, winter-long simulations from a chemistry transport model driven by different full-input reanalyses generally produce very similar results through most of the season for most species. However, substantial disparities between model runs are seen where composition gradients are largest. In particular, comparisons with satellite long-lived tracer measurements indicate that for most of the reanalyses the chemistry transport model underestimates the strength of confined diabatic descent inside the winter polar vortex. This underestimation of descent, together with other shortcomings in the model chemistry and physics schemes, manifests to varying degrees depending on the particular reanalysis used to force the model; as a consequence, considerable spread between the different simulations becomes evident by late winter.

- Results from a case study of a representative Antarctic winter (2009) show that estimates of chemical ozone loss based on satellite observations are relatively insensitive to the choice of reanalysis used for interpolation of the measurements to isentropic surfaces and identification of the vortex boundary. In contrast, chemical loss estimates based on simulated ozone fields from a chemistry transport model can differ substantially; in the case study, forcing the same model with different reanalyses yielded differences in the estimates of chemical ozone loss in the Antarctic vortex core as large as ~ 25 DU (20% - 30 %).

Recommendations

- Any of the recent full-input reanalyses (MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2) can be suitable for studies of lower stratospheric polar processing. However, substantial differences between the various reanalyses are found in some instances; therefore, the choice of which reanalysis to use in a given study may depend on the specific science questions being addressed.

- Temperature biases and other artifacts in older meteorological reanalyses often rendered them unsuitable for accurate modeling of the interannual variability in PSC formation and consequent denitrification, chlorine activation, and chemical ozone loss; in particular, ERA-40, NCEP-NCAR R1, and NCEP-DOE R2 are obsolete and should no longer be used for stratospheric studies of any kind.

- Because of the limitations of earlier reanalyses, it was not uncommon for modeling studies to try to match simulated chlorine activation and/or ozone loss by imposing arbitrary systematic adjustments of 1 - 2 K or more on the reanalysis temperatures. Increased confidence in the accuracy of current polar reanalysis temperatures provides tighter constraints on model parameterizations of microphysics/chemistry used to represent polar chemical processing. As a consequence, strong justification is warranted in modeling studies that seek to ascribe deficiencies in modeled chlorine activation and/or ozone losses to reanalysis temperature biases.

- Despite the overall good agreement between the polar temperatures from current full-input reanalyses, whenever feasible it is best to employ multiple reanalyses, even for studies involving recent winters for which it can reasonably be expected that differences between the reanalyses are small; using more than one reanalysis allows estimation of uncertainties and the potential impact of those uncertainties on the results, especially for quantities that cannot be directly compared with observations.

- If multiple reanalyses are used, they should be treated in the same manner to the extent possible to reduce the uncertainty in sources of differences; *e.g.*, data from one reanalysis on native model levels (*i.e.*, sigma coordinates) should not be used in conjunction with data from another reanalysis on pressure levels because of errors when interpolating from model levels to a coarser standard pressure grid.

- It would be helpful for the reanalysis centers to provide standard sets of products on standardized isobaric and isentropic levels. In particular, having PV available on model levels in future reanalyses would be extremely valuable. Vertical sampling of pressure and potential temperature levels comparable to that of model levels is also desirable.

- Reanalysis temperatures are generally unsuitable for assessment of trends in temperature-based diagnostics. One issue is that major changes in the input observational data in the assimilation systems are often made at approximately the same time in all of the reanalyses, hindering determination of the impact of such changes through reanalysis intercomparisons. Caution is especially advised for the estimation of trends in diagnostics that aggregate low temperatures over months and/or vertical levels in the Northern Hemisphere, such as the winter-mean fraction of the vortex volume with air cold enough for PSC existence; these types of diagnostics are particularly sensitive to the specific PSC threshold chosen, which is subject to non-negligible interannual variability.

Evaluation table

Figure 10.26 provides a summary evaluation of selected diagnostics examined in *Chapter 10*, as a quick reference to help users identify which reanalyses may be most suitable for a given issue related to stratospheric polar chemical processing. Only those diagnostics specifically examined either in this Chapter or in previously published papers are assigned a “score” in the table; otherwise they are marked “unevaluated” (tan shading). In particular, many of the polar processing diagnostics have not been formally assessed for the earlier ERA-40, NCEP-NCAR R1, and NCEP-DOE R2 reanalyses. However, given the considerable shortcomings in their representation of stratospheric temperatures and/or winds identified in prior studies, it is extremely unlikely that those reanalyses would perform well for the remaining diagnostics, and further evaluation of them was not undertaken here. Note that the results for some diagnostics used in this chapter are too complex to be amenable to such a rating system and have therefore been omitted. Most notably, the assessment of transport processes in the chemistry transport model yielded results that vary substantially by hemisphere, time of year, altitude, location in the polar vortex (core *vs*. edge region), and species, precluding simple categorization. Similarly, the PSC thermodynamic-consistency diagnostics have also been omitted from **Figure 10.26**.

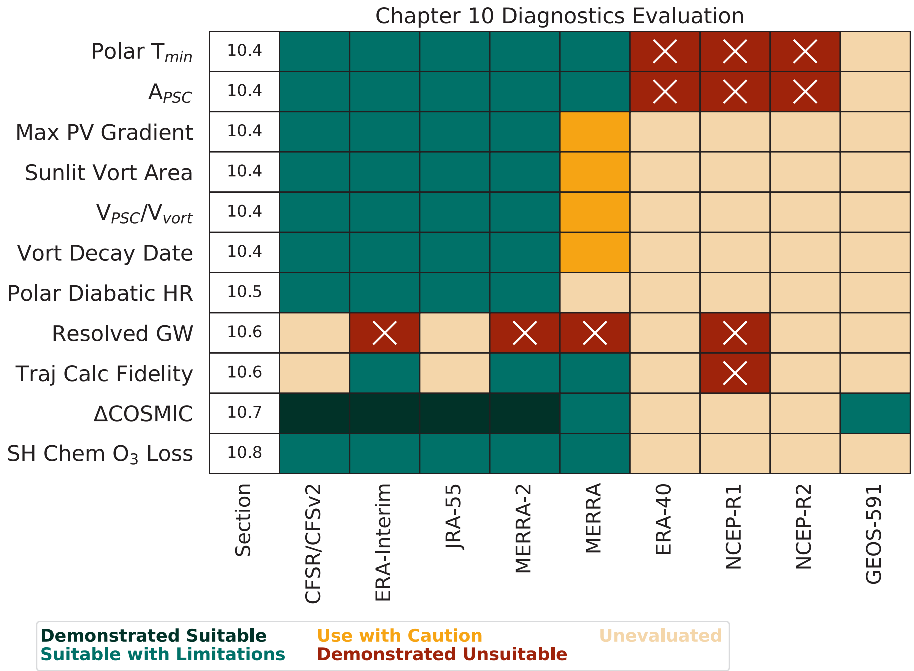


Figure 10.26: *Summary evaluation table for most of the Chapter 10 diagnostics. Along with a "score" capturing the overall performance of each reanalysis considered for each diagnostic (see table legend), the first column contains a pointer to the specific Chapter 10 section in which the diagnostic is discussed in detail.*

Data availability

Data availability for the reanalysis products is described in *Chapter 2* of this Report.

CALIOP data were obtained from the NASA Langley Research Center Atmospheric Science Data Center (*CALIPSO Science Team*, 2015a; *CALIPSO Science Team*, 2015b).

The Concordiasi Thermodynamical SENsor (TSEN) dataset is provided by the Laboratoire de Météorologie Dynamique (*Concordiasi Science Team*, 2010).

COSMIC data were obtained from the University Corporation for Atmospheric Research COSMIC Data Analysis and Archive Center (*COSMIC Science Team*, 2013).

GEOS-5.9.1 data were obtained from the Goddard Earth Sciences Data and Information Services Center (*Global Modeling and Assimilation Office*, 2013).

MLS data are archived at the NASA Goddard Earth Sciences Data Information and Services Center (*Froidevaux et al.*, 2015; *Lambert et al.*, 2015; *Manney et al.*, 2015; *Santee et al.*, 2015; *Schwartz et al.*, 2015a,b).

Acknowledgements

We thank the Microwave Limb Sounder team at JPL, especially Brian W. Knosp, Luis F. Millán, and Ryan A. Fuller, for computational, data processing, data management, and analysis support. We gratefully acknowledge members of the teams associated with the CALIOP, COSMIC, and MIPAS satellite missions and the Concordiasi campaign, as well as the reanalysis data centers. We thank Jonathon Wright for producing zonal mean diabatic heating rates files for the reanalyses. Helpful discussions with G. Stiller and T. von Clarmann were greatly appreciated, as were the comments and suggestions from authors of other S-RIP Report chapters. Work at the Jet Propulsion Laboratory, California Institute of Technology, was carried out under a contract with the National Aeronautics and Space Administration.

References

Anthes, R.A., 2011: Exploring Earth's atmosphere with radio occultation: contributions to weather, climate and space weather, *Atmos. Meas. Tech.*, **4**, 1077 - 1103, doi: 10.5194/amt-4-1077-2011.

Balis, D., *et al.*, 2011: Observed and modelled record ozone decline over the Arctic during winter/spring 2011. *Geophys. Res. Lett.*, **38**, L23801, doi: 10.1029/2011GL049259.

Boccara, G., A.Hertzog, C. Basdevant, and F. Vial, 2008: Accuracy of NCEP/NCAR reanalyses and ECMWF analyses in the lower stratosphere over Antarctica in 2005. *J. Geophys. Res. Atmos.*, **113**, doi: 10.1029/2008JD010116.

Bonavita, M., 2014: On some aspects of the impact of GPSRO observations in global numerical weather prediction. *Q. J. Roy. Meteor. Soc.*, **140**, 2546 - 2562, doi: 10.1002/qj.2320.

Braun, M. *et al., 2019*: Nitrification of the lowermost stratosphere during the exceptionally cold Arctic winter 2015 - 2016. *Atmos. Chem. Phys.*, **19**, 13 681 - 13 699, doi: 10.5194/acp-19-13681-2019.

CALIPSO Science Team, 2015a: CALIPSO Science Team: CALIPSO/CALIOP Level 1B, Lidar Profile Data, versions 3.01, 3.02 and 3.30. Hampton, VA, USA: *NASA Atmospheric Science Data Center* (ASDC), doi: 10.5067/CALIOP/CALIPSO/CAL_LID_L1-ValStage1-V301_L1B-003.01; 10.5067/CALIOP/CALIPSO/CAL_LID_L1-ValStage1-V3-01_L1B-003.02; 10.5067/CALIOP/CALIPSO/CAL_LID_L1-ValStage1-V3-01_L1B-003.30.

CALIPSO Science Team, 2015b: CALIPSO Science Team: CALIPSO/CALIOP Level 2, Polar Stratospheric Cloud Data, versions 1.00. Hampton, VA, USA: *NASA Atmospheric Science Data Center* (ASDC), doi: 10.5067/CALIOP/CALIPSO/CAL_LID_L2_PSCMASK-PROV-V1-00_L2-001.00.

Carslaw, K.S., B.P. Luo, and T. Peter, 1995: An analytic expression for the composition of aqueous HNO_3 - H_2SO_4 stratospheric aerosols including gas phase removal of HNO_3. *Geophys. Res. Lett.*, **22**, 1877 - 1880, doi: 10.1029/95GL01668.

Chabrillat, S., *et al.*, 2018: Comparison of mean age of air in five reanalyses using the BASCOE transport model. *Atmos. Chem. Phys.*, **18**, 14 715 - 14 735, doi: 10.5194/acp-18-14715-2018.

Chipperfield, M.P., 2006: New version of the TOMCAT/SLIMCAT off-line chemical transport model: Intercomparison of stratospheric tracer experiments. *Q. J. Roy. Meteor. Soc.*, **132**, 1179 - 1203, doi: 10.1256/qj.05.51.

Chipperfield, M.P., *et al.*, 1997: Model studies of chlorine deactivation and formation of $ClONO_2$ collar in the Arctic polar vortex. *J. Geophys. Res.*, **102**, 1467 - 1478, doi: 10.1029/96JD00442.

Concordiasi Science Team, 2010: Concordiasi Thermodynamical SENsor (TSEN) dataset. Available at: **http://www.umr-cnrm.fr/concordiasi-dataset/index.php**, last access on 14 July 2021.

COSMIC Science Team, 2013: COSMIC Data Analysis and Archive Center, Constellation Observing System for Meteorology, Ionosphere and Climate, University Corporation for Atmospheric Research, Atmospheric Profiles from COSMIC Occultation Data, *Research Data Archive at the National Center for Atmospheric Research*, Computational and Information Systems Laboratory, Boulder CO, Available at: **http://rda.ucar.edu/datasets/ds723.0/**, last access on 14 July 2021.

Daerden, F., *et al.*, 2007: A 3D-CTM with detailed online PSC-microphysics: analysis of the Antarctic winter 2003 by comparison with satellite observations. *Atmos. Chem. Phys.*, **7**, 1755 - 1772, doi: 10.5194/acp-7-1755-2007.

Davies, S. *et al.*, 2003: Modeling the effect of denitrification on Arctic ozone depletion during winter 1999/2000. *J. Geophys. Res.*, **108**, 8322, doi: 10.1029/2001JD000445.

Dunkerton, T., 1978: On the mean meridional mass motions of the stratosphere and mesosphere. *J. Atmos. Sci.*, **35**, 2325 - 2333, doi: 10.1175/1520-0469(1978)035<2325:OTMMMM>2.0.CO;2.

Errera, Q., *et al.*, 2008: 4D-Var assimilation of MIPAS chemical observations: ozone and nitrogen dioxide analyses. *Atmos. Chem. Phys.*, **8**, 6169 - 6187, doi: 10.5194/acp-8-6169-2008.

Errera, Q., *et al.*, 2016: Harmonisation and diagnostics of MIPAS ESA CH_4 and N_2O profiles using data assimilation. *Atmos. Meas. Tech.*, **9**, 5895 - 5909, doi: 10.5194/amt-9-5895-2016.

Errera, Q. *et al.*, 2019: Technical note: Reanalysis of Aura MLS chemical observations. *Atmos. Chem. Phys.*, **19**, 13 647 - 13 679, doi: 10.5194/acp-19-13647-2019.

Feng, W., *et al.*, 2005: Three-dimensional model study of the Antarctic ozone hole in 2002 and comparison with 2000. *J. Atmos. Sci.*, **62**, 822 - 837, doi: 10.1175/JAS3335.1.

Fischer, H., *et al.*, 2008: MIPAS: an instrument for atmospheric and climate research. *Atmos. Chem. Phys.*, **8**, 2151 - 2188, doi: 10.5194/acp-8-2151-2008.

Fonteyn, D. and N. Larsen, 1996: Detailed PSC formation in a two-dimensional chemical transport model of the stratosphere. *Ann. Geophys.*, **14**, 315 - 328, doi: 10.1007/s00585-996-0315-0.

Forster, P.M., M. Ponater, and W.Y. Zhong, 2001: Testing broadband radiation schemes for their ability to calculate the radiative forcing and temperature response to stratospheric water vapour and ozone changes. *Meteorol. Z.*, **10**, 387 - 393, doi: 10.1127/0941-2948/2001/0010-0387.

Froidevaux, L. *et al.*, 2008a: Validation of Aura Microwave Limb Sounder HCl measurements. *J. Geophys. Res.*, **113**, D15S25, doi: 10.1029/2007JD009025.

Froidevaux, L. *et al.*, 2008b: Validation of Aura Microwave Limb Sounder stratospheric ozone measurements. *J. Geophys. Res.*, **113**, D15S20, doi: 10.1029/2007JD008771.

Froidevaux, L., N. Livesey, and W. Read, 2015: MLS/Aura Level 2 Hydrogen Chloride (HCl) Mixing Ratio V004. Greenbelt, MD, USA, *Goddard Earth Sciences Data and Information Services Center* (GES DISC). doi 10.5067/Aura/MLS/DATA2010.

Froidevaux, L., *et al.*, 2019: Evaluation of CESM1 (WACCM) free-running and specified dynamics atmospheric composition simulations using global multispecies satellite data records. *Atmos. Chem. Phys.*, **19**, 4783 - 4821, doi: 10.5194/acp-19-4783-2019.

Fujiwara, M., *et al.*, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp17-1417-2017.

Gelaro, R., *et al.*, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate*, **30**, 5419 - 5454, doi: 10.1175/JCLI-D-16-0758.1.

Global Modeling and Assimilation Office (GMAO), 2013: GEOS5.9.1 v1.2 NRT Assimilation Products FP-IT (Forward Processing for Instrument Teams), available by subscription at: **https://gmao.gsfc.nasa.gov/products/index.php**.

Gobiet, A., *et al*., 2005: Climatological validation of stratospheric temperatures in ECMWF operational analyses with CHAMP radio occultation data. *Geophys. Res. Lett.*, **32**, doi: 10.1029/2005GL022617.

Gobiet, A., *et al.*, 2007: Retrieval of temperature profiles from CHAMP for climate monitoring: Intercomparison with Envisat MIPAS and GOMOS and different atmospheric analyses. *Atmos. Chem. Phys.*, **7**, 3519 - 3536, doi: 10.5194/acp-7-3519-2007.

Grooß, J.-U., *et al.*, 2011: Stratospheric ozone chemistry in the Antarctic: what determines the lowest ozone values reached and their recovery? *Atmos. Chem. Phys.*, **11**, 12 217 - 12 226, doi: 10.5194/acp-11-12217-2011.

Grooß, J.-U., *et al.*, 2018: On the discrepancy of HCl processing in the core of the wintertime polar vortices. *Atmos. Chem. Phys.*, **18**, 8647 - 8666, doi: 10.5194/acp-18-8647-2018.

Hanson, D. and K. Mauersberger, 1988: Laboratory studies of the nitric acid trihydrate: Implications for the south polar stratosphere. *Geophys. Res. Lett.*, **15**, 855 - 858, doi: 10.1029/GL015i008p00855.

Hanson, D.R. and A.R. Ravishankara, 1994: Reactive Uptake of $ClONO_2$ onto Sulfuric Acid Due to Reaction with HCl and H_2O. *J. Phys. Chem.*, **98**, 5728 - 5735, doi: 10.1021/j100073a026.

Hertzog, A., C. Basdevant, F. Vial, and C.R. Mechoso, 2004: The accuracy of stratospheric analyses in the northern hemisphere inferred from long-duration balloon flights. *Q. J. Roy. Meteor. Soc.*, **130**, 607 - 626, doi: 10.1256/qj.03.76.

Hertzog, A., *et al.*, 2008: Estimation of Gravity Wave Momentum Flux and Phase Speeds from Quasi-Lagrangian Stratospheric Balloon Flights. Part II: Results from the Vorcore Campaign in Antarctica. *J. Atmos. Sci.*, **65**, 3056 - 3070, doi: 10.1175/2008JAS2710.1.

Hertzog, A., M.J. Alexander, and R. Plougonven, 2012: On the Intermittency of Gravity Wave Momentum Flux in the Stratosphere. *J. Atmos. Sci.*, **69**, 3433 - 3448, doi: 10.1175/JAS-D-12-09.1.

Hitchman, M.H., M. McKay, and C.R. Trepte, 1994: A climatology of stratospheric aerosol. *J. Geophys. Res.*, **99**, 20 689 - 20 700, doi: 10.1029/94JD01525.

Ho, S.-P., *et al.*, 2012: Reproducibility of GPS radio occultation data for climate monitoring: Profile-to-profile inter-comparison of CHAMP climate records 2002 to 2008 from six data centers. *J. Geophys. Res.*, **117**, doi: 10.1029/2012JD017665.

Hoffmann, L., *et al.*, 2016: Lagrangian transport simulations of volcanic sulfur dioxide emissions: Impact of meteorological data products. *J. Geophys. Res.*, **121**, 4651 - 4673, doi: 10.1002/2015JD023749.

Hoffmann, L., *et al.*, 2017a: Intercomparison of meteorological analyses and trajectories in the Antarctic lower stratosphere using Concordiasi superpressure balloon observations. *Atmos. Chem. Phys.*, **17**, 8045 - 8061, doi: 10.5194/acp-17-8045-2017.

Hoffmann, L., *et al.*, 2017b: A decadal satellite record of gravity wave activity in the lower stratosphere to study polar stratospheric cloud formation. *Atmos. Chem. Phys.*, **17**, 2901 - 2920, doi: 10.5194/acp-17-2901-2017.

Höpfner, M. *et al.*, 2007: Validation of MIPAS $ClONO_2$ measurements. *Atmos. Chem. Phys.*, **7**, 257 - 281, doi: 10.5194/acp-7-257-2007.

Hoppe, C.M., *et al.*, 2014: The implementation of the CLaMS Lagrangian transport core into the chemistry climate model EMAC 2.40.1: application on age of air and transport of long-lived trace species. *Geosci. Model Dev.*, **7**, 2639 - 2651, doi: 10.5194/gmd-7-2639-2014.

Hoppe, C.M., F. Ploeger, P. Konopka, and R. Müller, 2016: Kinematic and diabatic vertical velocity climatologies from a chemistry climate model. *Atmos. Chem. Phys.*, **16**, 6223 - 6239, doi: 10.5194/acp-16-6223-2016.

Huijnen, V., *et al.*, 2016: C-IFS-CB05-BASCOE: stratospheric chemistry in the Integrated Forecasting System of ECMWF. *Geosci. Model Dev.*, **9**, 3071 - 3091, doi: 10.5194/gmd-9-3071-2016.

Hurst, D.F., *et al.*, 2014: Validation of Aura Microwave Limb Sounder stratospheric water vapor measurements by the NOAA frost point hygrometer. *J. Geophys. Res.*, **119**, 1612 - 1625, doi: 10.1002/2013JD020757.

Hurst, D.F., *et al.*, 2016: Recent divergences in stratospheric water vapor measurements by frost point hygrometers and the Aura Microwave Limb Sounder. *Atmos. Meas. Tech.*, **9**, 4447 - 4457, doi: 10.5194/amt-9-4447-2016.

Jewtoukoff, V., *et al.*, 2015: Comparison of Gravity Waves in the Southern Hemisphere Derived from Balloon Observations and the ECMWF Analyses. *J. Atmos. Sci.*, **72**, 3449 - 3468, doi: 10.1175/JAS-D-14-0324.1.

Khosrawi, F., *et al.*, 2018: Comparison of ECHAM5/MESSy Atmospheric Chemistry (EMAC) simulations of the Arctic winter 2009/2010 and 2010/2011 with Envisat/MIPAS and Aura/MLS observations. *Atmos. Chem. Phys.*, **18**, 8873 - 8892, doi: 10.5194/acp-18-8873-2018.

Kiehl, J.T. and S. Solomon, 1986: On the radiative balance of the stratosphere. *J. Atmos. Sci.*, **43**, 1525 - 1534, doi: 10.1175/1520-0469(1986)043<1525:OTRBOT>2.0.CO;2.

Knox, J.A., 1998: On converting potential temperature to altitude in the middle atmosphere. *Eos Trans. AGU*, **79**, 376, doi: 10.1029/98EO00290.

Knudsen, B.M., 1996: Accuracy of arctic stratospheric temperature analyses and the implications for the prediction of polar stratospheric clouds. *Geophys. Res. Lett.*, **23**, 3747 - 3750, doi: 10.1029/96GL03541.

Knudsen, B.M., *et al.*, 2001: Comparison of stratospheric air parcel trajectories based on different meteorological analyses. *J. Geophys. Res.*, **106**, 3415 - 3424, doi: 10.1029/2000JD900608.

Knudsen, B.M., *et al.*, 2002: Accuracy of analyzed stratospheric temperatures in the winter Arctic vortex from infrared Montgolfier long-duration balloon flights 2. Results. *J. Geophys. Res.*, **107**, doi: 10.1029/2001JD001329.

Krüger, K., B. Naujokat, and K. Labitzke, 2005: The unusual midwinter warming in the southern hemisphere stratosphere 2002: A comparison to northern hemisphere phenomena. *J. Atmos. Sci.*, **62**, 603 - 613, doi: 10.1175/JAS-3316.1.

Kuo, Y.-H., M. Skumanich, P.L. Haagenson, and J.S. Chang, 1985: The Accuracy of Trajectory Models as Revealed by the Observing System Simulation Experiments. *Mon. Weather Rev.*, **113**, 1852 - 1867, doi: 10.1175/1520-0493(1985)113<1852:TAOTMA>2.0.CO;2.

Kuttippurath, J. and G. Nikulin, 2012: A comparative study of the major sudden stratospheric warmings in the Arctic winters 2003/2004 - 2009/2010. *Atmos. Chem. Phys.*, **12**, 8115 - 8129, doi: 10.5194/acp-12-8115-2012.

Labitzke, K. and M. Kunze, 2005: Stratospheric temperatures over the Arctic: Comparison of three data sets. *Meteorol. Z.*, **14**, 65 - 74, doi: 10.1127/0941-2948/2005/0014-0065.

Labitzke, K. and M. Kunze, 2009: On the remarkable Arctic winter in 2008/2009. *J. Geophys. Res.*, **114**, doi: 10.1029/2009JD012273.

Lambert, A. and M.L. Santee, 2018: Accuracy and precision of polar lower stratospheric temperatures from reanalyses evaluated from A-Train CALIOP and MLS, COSMIC GPS RO, and the equilibrium thermodynamics of supercooled ternary solutions and ice clouds. *Atmos. Chem. Phys.*, **18**, 1945 - 1975, doi: 10.5194/acp-18-1945-2018.

Lambert, A., M.L. Santee, D.L. Wu, and J.H. Chae, 2012: A-train CALIOP and MLS observations of early winter Antarctic polar stratospheric clouds and nitric acid in 2008. *Atmos. Chem. Phys.*, **12**, 2899 - 2931, doi: 10.5194/acp-12-2899-2012.

Lambert, A., W. Read, and N. Livesey, 2015: MLS/Aura Level 2 Water Vapor (H_2O) Mixing Ratio V004, Greenbelt, MD, USA, *Goddard Earth Sciences Data and Information Services Center* (GES DISC). doi: 10.5067/AURA/MLS/DATA2009.

Lambert, A. *et al.*, 2007: Validation of the Aura Microwave Limb Sounder middle atmosphere water vapor and nitrous oxide measurements. *J. Geophys. Res.*, **112**, D24S36, doi: 10.1029/2007JD008724.

Lawrence, Z.D., *et al.*, 2015: Comparisons of polar processing diagnostics from 34 years of the ERA-Interim and MERRA reanalyses. *Atmos. Chem. Phys.*, **15**, 3873 - 3892, doi: 10.5194/acp-15-3873-2015.

Lawrence, Z.D., G.L. Manney, and K. Wargan, 2018: Reanalysis intercomparisons of stratospheric polar processing diagnostics. *Atmos. Chem. Phys.*, **18**, 13 547 - 13 579, doi: 10.5194/acp-18-13547-2018.

Lin, S.-J. and R.B. Rood, 1996: Multidimensional Flux-Form Semi-Lagrangian Transport Schemes. *Mon. Weather Rev.*, **124**, 2046 - 2070, doi: 10.1175/1520-0493(1996)124<2046:mffslt>2.0.co;2.

Lindenmaier, R., *et al.*, 2011: A study of the Arctic NO_y budget above Eureka, Canada. *J. Geophys. Res.*, **116**, doi: 10.1029/2011JD016207, 2011.

Livesey, N.J., M.L. Santee, and G.L. Manney, 2015: A Match-based approach to the estimation of polar stratospheric ozone loss using Aura Microwave Limb Sounder observations. *Atmos. Chem. Phys.*, **15**, 9945 - 9963, doi: 10.5194/acp-15-9945-2015.

Livesey, N.J. *et al.*, 2020: Version 4.2x Level 2 and 3 data quality and description document. *Tech. Rep. JPL* **D-33509** Rev. E, Jet Propulsion Laboratory, available at **http://mls.jpl.nasa.gov**.

Livesey, N.J., *et al.*, 2021: Investigation and amelioration of long-term instrumental drifts in water vapor and nitrous oxide measurements from the Aura Microwave Limb Sounder (MLS) and their implications for studies of variability and trends, *Atmos. Chem. Phys.*, **21**, 15409 - 15430, doi:10.5194/acp-21-15409-2021.

Long, C.S., *et al.*, 2017: Climatology and Interannual Variability of Dynamic Variables in Multiple Reanalyses Evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14,593 - 14,629, doi: 10.5194/acp-17-14593-2017.

Manney, G., *et al.*, 2015: MLS/Aura Level 2 Nitric Acid (HNO_3) Mixing Ratio V004, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), doi: 10.5067/AURA/MLS/DATA2012.

Manney, G.L., *et al.*, 1996: Comparison of U.K. Meteorological Office and U.S. National Meteorological Center stratospheric analyses during northern and southern winter. *J. Geophys. Res.*, **101**, 10,311 - 10,334, doi: 10.1029/95JD03350.

Manney, G.L., *et al.*, 2003a: Variability of ozone loss during Arctic winter (1991 to 2000) estimated from UARS Microwave Limb Sounder measurements. *J. Geophys. Res.*, **108**, 4149, doi: 10.1029/2002JD002634.

Manney, G.L., *et al.*, 2003b: Lower stratospheric temperature differences between meteorological analyses in two cold Arctic winters and their impact on polar processing studies. *J. Geophys. Res.*, **108**, 8328, doi: 10.1029/2001JD001149.

Manney, G.L. *et al.* 2005: Diagnostic comparison of meteorological analyses during the 2002 Antarctic winter. *Mon. Weather Rev.*, **133**, 1261 - 1278, doi: 10.1175/MWR2926.1.

Manney, G.L. *et al.*, 2007: Solar occultation satellite data and derived meteorological products: Sampling issues and comparisons with Aura Microwave Limb Sounder. *J. Geophys. Res.*, **112**, D24S50, doi: 10.1029/2007JD008709.

Manney, G.L. *et al.*, 2008: The evolution of the stratopause during the 2006 major warming: Satellite data and assimilated meteorological analyses. *J. Geophys. Res.*, **113**, D11115, doi: 10.1029/2007JD009097.

Manney, G.L., *et al.*, 2009a: Aura Microwave Limb Sounder observations of dynamics and transport during the record-breaking 2009 Arctic stratospheric major warming. *Geophys. Res. Lett.*, **36**, L12815, doi: 10.1029/2009GL038586.

Manney, G.L. *et al.*, 2009b: Satellite observations and modelling of transport in the upper troposphere through the lower mesosphere during the 2006 major stratospheric sudden warming. *Atmos. Chem. Phys.*, **9**, 4775 - 4795, doi: 10.5194/acp-9-4775-2009.

Manney, G.L., *et al.*, 2011a: Jet characterization in the upper troposphere/lower stratosphere (UTLS): applications to climatology and transport studies. *Atmos. Chem. Phys.*, **11**, 6115 - 6137, doi: 10.5194/acp-11-6115-2011.

Manney, G.L., *et al.*, 2011b: Unprecedented Arctic Ozone Loss in 2011. *Nature*, **478**, 469 - 475, doi: 10.1038/nature10556.

Martineau, P., J.S. Wright, N. Zhu, and M. Fujiwara, M., 2018: Zonal-mean data set of global atmospheric reanalyses on pressure levels. *Earth Syst. Sci. Data*, **10**, 1925 - 1941, doi: 10.5194/essd-10-1925-2018.

Maycock, A.C. and K.P. Shine, 2012: Stratospheric water vapor and climate: Sensitivity to the representation in radiation codes. *J. Geophys. Res.*, **117**, D13102, doi: 10.1029/2012JD017484.

Meijer, E.W., B. Bregman, A. Segers, and P.F.J. van Velthoven, 2004: The influence of data assimilation on the age of air calculated with a global chemistry-transport model using ECMWF wind fields. *Geophys. Res. Lett.*, **31**, doi: 10.1029/2004GL021158.

Menang, K.P., 2018: Assessment of the impact of solar spectral irradiance on near-infrared clear-sky atmospheric absorption and heating rates. *J. Geophys. Res.*, **123**, 6460 - 6468, doi: 10.1029/2018JD02834.

Mertens, C.J., M.G. Mlynczak, R.R. Garcia, and R.W. Portmann, 1999: A detailed evaluation of the stratospheric heat budget: 1. Radiation transfer. *J. Geophys. Res. Atmos.*, **104**, 6021 - 6038, doi: 10.1029/1998JD200100.

Minganti, D., *et al.*, 2020: Climatological impact of the Brewer–Dobson Circulation on the N_2O budget in WACCM, a chemical reanalysis and a CTM driven by four dynamical reanalyses. *Atmos. Chem. Phys.,* **20**, 12609 - 12631, doi: 10.5194/acp-20-12609-2020.

Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8**, 1339 - 1356, doi: 10.5194/gmd-8-1339-2015.

Murgatroyd, R.J. and F. Singleton, 1961: Possible meridional circulations in the stratosphere and mesosphere. *Q. J. Roy. Meteor. Soc.*, **87**, 125 - 135, doi: 10.1002/qj.49708737202.

Murphy, D.M. and T. Koop, 2005: Review of the vapour pressures of ice and supercooled water for atmospheric applications. *Q. J. Roy. Meteor. Soc.*, **131**, 1539 - 1565, doi: 10.1256/qj.04.94.

Newman, P.A. and E.R. Nash, 2005: The unusual southern hemisphere stratosphere winter of 2002. *J. Atmos. Sci.*, **62**, 614 - 628, doi: 10.1175/JAS-3323.1.

Olaguer, E.P., H. Yang, and K.K. Tung, 1992: A reexamination of the radiative balance of the stratosphere. *J. Atmos. Sci.*, **49**, 1242 - 1261, doi: 10.1175/1520-0469(1992)049<1242:AROTRB>2.0.CO;2.

Parrondo, M.C., *et al.*, 2007: Mid-winter lower stratosphere temperatures in the Antarctic vortex: comparison between observations and ECMWF and NCEP operational models. *Atmos. Chem. Phys.*, **7**, 435 - 441, doi: 10.5194/acp-7-435-2007.

Pawson, S., *et al.*, 2007: Stratospheric transport using 6-h-averaged winds from a data assimilation system. *J. Geophys. Res.*, **112**, doi: 10.1029/2006JD007673, 2007.

Pitts, M.C., L.R. Poole, and L.W. Thomason, 2009: CALIPSO polar stratospheric cloud observations: second-generation detection algorithm and composition discrimination. *Atmos. Chem. Phys.*, **9**, 7577 - 7589, doi: 10.5194/acp-9-7577-2009.

Pitts, M.C., L.R. Poole, A. Lambert, and L.W. Thomason, 2013: An assessment of CALIOP polar stratospheric cloud composition classification. *Atmos. Chem. Phys.*, **13**, 2975 - 2988, doi: 10.5194/acp-13-2975-2013.

Plougonven, R., A. Hertzog, A. and L. Guez, 2013: Gravity waves over Antarctica and the Southern Ocean: consistent momentum fluxes in mesoscale simulations and stratospheric balloon observations. *Q. J. Roy. Meteor. Soc.*, **139**, 101 - 118, doi: 10.1002/qj.1965.

Podglajen, A., A. Hertzog, R. Plougonven, and N. Žagar, 2014: Assessment of the accuracy of (re)analyses in the equatorial lower stratosphere. *J. Geophys. Res.*, **119**, 11,166 - 11,188, doi: 10.1002/2014JD021849.

Poli, P. and J. Joiner, 2004: Effects of horizontal gradients on GPS radio occultation observation operators. I: Ray tracing. *Q. J. Roy. Meteor. Soc.*, **130**, 2787 - 2805, doi: 10.1256/qj.03.228.

Poli, P., S.B. Healy, and D.P. Dee, 2010: Assimilation of Global Positioning System radio occultation data in the ECMWF ERA–Interim reanalysis. *Q. J. Roy. Meteor. Soc.*, **136**, 1972 - 1990, doi: 10.1002/qj.722.

Pommereau, J.-P., *et al.*, 2002: Accuracy of analyzed stratospheric temperatures in the winter Arctic vortex from infrared Montgolfier long-duration balloon flights 1. Measurements. *J. Geophys. Res.*, **107**, doi: 10.1029/2001JD001379.

Rabier, F., *et al.*, 2010: The Concordiasi Project in Antarctica. *Bull. Am. Meteor. Soc.*, **91**, 69 - 86, doi: 10.1175/2009BAMS2764.1.

Ramanathan, V., 1976: Radiative transfer within the Earth's troposphere and stratosphere: a simplified radiative-convective model. *J. Atmos. Sci.*, **33**, 1330 - 1346, doi: 10.1175/1520-0469(1976)033<1330:RTWTET>2.0.CO;2.

Rieder, H.E. and L.M. Polvani, 2013: Are recent Arctic ozone losses caused by increasing greenhouse gases? *Geophys. Res. Lett.*, **40**, 4437 - 4441, doi: 10.1002/grl.50835.

Rienecker, M.M. *et al.*, 2011: MERRA - NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624 - 3648, doi: 10.1175/JCLI-D-11-00015.1.

Rolph, G.D. and R.R. Draxler, 1990: Sensitivity of Three-Dimensional Trajectories to the Spatial and Temporal Densities of the Wind Field. *J. App. Meteorol.*, **29**, 1043 - 1054, doi: 10.1175/1520-0450(1990)029<1043:SOTDTT>2.0.CO;2.

Rößler, T., *et al.*, 2018: Trajectory errors of different numerical integration schemes diagnosed with the MPTRAC advection module driven by ECMWF operational analyses. *Geosci. Model Dev.*, **11**, 575 - 592, doi: 10.5194/gmd-11-575-2018.

Ryan, N.J., *et al.*, 2018: Assessing the ability to derive rates of polar middle-atmospheric descent using trace gas measurements from remote sensors. *Atmos. Chem. Phys.*, **18**, 1457 - 1474, doi: 10.5194/acp-18-1457-2018.

Sander, S.P., *et al.*, 2000: Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling. Supplement to evaluation 12: Update of key reactions. *Evaluation Number 13*, Publication **00–3**, JPL, available at: **http://jpldataeval.jpl.nasa.gov/**.

Santee, M.L. *et al.*, 2007: Validation of the Aura Microwave Limb Sounder HNO_3 measurements. *J. Geophys. Res.*, **112**, D24S40, doi: 10.1029/2007JD008721.

Santee, M.L. *et al.*, 2008: Validation of the Aura Microwave Limb Sounder ClO measurements. *J. Geophys. Res.*, **113**, D15S22, doi: 10.1029/2007JD008762.

Santee, M., N. Livesey, and W. Read, 2015: MLS/Aura Level 2 Chlorine Monoxide (ClO) Mixing Ratio V004. Greenbelt, MD, USA, *Goddard Earth Sciences Data and Information Services Center* (GES DISC). doi:10.5067/Aura/MLS/DATA2004.

Schwartz, M., N. Livesey, and W. Read, 2015a: MLS/Aura Level 2 Temperature V004. Greenbelt, MD, USA, *Goddard Earth Sciences Data and Information Services Center* (GES DISC), doi: 10.5067/AURA/MLS/DATA2021.

Schwartz, M., L. Froidevaux, N. Livesey, and W. Read, 2015b: MLS/Aura Level 2 Ozone (O3) Mixing Ratio V004. Greenbelt, MD, USA, *Goddard Earth Sciences Data and Information Services Center* (GES DISC). doi: 10.5067/Aura/MLS/DATA2017.

Schwartz, M.J., et al., 2008: Validation of the Aura Microwave Limb Sounder Temperature and Geopotential Height Measurements. *J. Geophys. Res.*, **113**, D15S02, doi: 10.1029/2007JD008805.

Shepherd, T.G., R.A. Plumb, and S.C. Wofsy, 2005: PREFACE (Special Issue on 2002 Antarctic Major Sudden Stratospheric Warming). *J. Atmos. Sci.*, **62**, 565 - 566, doi: 10.1175/JAS-9999.1.

Sinnhuber, B.-M., *et al.*, 2011: Arctic winter 2010/2011 at the brink of an ozone hole. *Geophys. Res. Lett.*, **38**, L24814, doi: 10.1029/2011GL049784.

Solomon, S., D. Kinnison, J. Bandoro,and R. Garcia, 2015: Simulation of polar ozone depletion: An update. *J. Geophys. Res. Atmos.*, **120**, 7958 - 7974, doi: 10.1002/2015JD023365.

Stohl, A., 1998: Computation, accuracy and applications of trajectories–A review and bibliography. *Atmos. Env.*, **32**, 947 - 966, doi: 10.1016/S1352-2310(97)00457-3.

Stohl, A., G. Wotawa, P. Seibert, and H. Kromp-Kolb, 1995: Interpolation Errors in Wind Fields as a Function of Spatial and Temporal Resolution and Their Impact on Different Types of Kinematic Trajectories. *J. App. Meteorol.*, **34**, 2149 - 2165, doi: 10.1175/1520-0450(1995)034<2149:IEIWFA>2.0.CO;2.

Stohl, A., *et al.*, 2005: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2. *Atmos. Chem. Phys.*, **5**, 2461 - 2474, doi: 10.5194/acp-5-2461-2005.

Strahan, S.E. and A.R. Douglass, 2018: Decline in Antarctic Ozone Depletion and Lower Stratospheric Chlorine Determined From Aura Microwave Limb Sounder Observations. *Geophys. Res. Lett.*, **45**, 382 - 390, doi: 10.1002/2017GL074830.

Tilmes, S., et al., 2006: Chemical ozone loss in the Arctic and Antarctic stratosphere between 1992 and 2005. Geophys. Res. Lett., 33, L20812, doi: 10.1029/2006GL026925, 2006.

Vincent, R.A. and A. Hertzog, 2014: The response of superpressure balloons to gravity wave motions. *Atmos. Meas. Tech.*, **7**, 1043 - 1055, doi: 10.5194/amt-7-1043-2014.

von Clarmann, T., *et al.*, 2009: Retrieval of temperature, H_2O, O_3, HNO_3, CH_4, N_2O, $ClONO_2$ and ClO from MIPAS reduced resolution nominal mode limb emission measurements. *Atmos. Meas. Tech.*, **2**, 159 - 175, doi: 10.5194/amt-2-159-2009.

Waters, J.W. *et al.*, 2006: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite. *IEEE Trans. Geosci. Remote Sens.*, **44**, 1075 - 1092, doi: 10.1109/TGRS.2006.873771.

Wegner, T. *et al.*, 2012: Heterogeneous chlorine activation on stratospheric aerosols and clouds in the Arctic polar vortex. *Atmos. Chem. Phys.*, **12**, 11,095 - 11,106, doi: 10.5194/acp-12-11095-2012.

Winker, D.M., *et al.*, 2009: Overview of the CALIPSO Mission and CALIOP data processing algorithms. *J. Atmos. Oceanic Technol.*, **26**, 2310 - 2323, doi: 10.1175/2009JTECHA1281.1.

WMO, 2007: Scientific assessment of ozone depletion: 2006. *Global Ozone Res. and Monit. Proj.* — Report No. **50**, Geneva, Switzerland.

WMO, 2018: Scientific assessment of ozone depletion: 2018. *Global Ozone Res. and Monit. Proj.* — Report No. **58**, Geneva, Switzerland.

Wohltmann, I., R. Lehmann, and M. Rex, 2017: A quantitative analysis of the reactions involved in stratospheric ozone depletion in the polar vortex core. *Atmos. Chem. Phys.*, **17**, 10,535 - 10,563, doi: 10.5194/acp-17-10535-2017.

Wright, J., 2017: S-RIP: Zonal-mean heating rates of global atmospheric reanalyses on pressure levels. *Centre for Environmental Data Analysis*, doi: 10.5285/70146c789eda4296a3c3ab6706931d56, last access on 14 July 2021.

Major abbreviations and terms

4D-Var	Four-Dimensional Variational data assimilation scheme
AHTD	Absolute Horizontal Transport Deviation
ATOVS	Advanced TIROS Operational Vertical Sounder
BASCOE	Belgian Assimilation System for Chemical ObsErvations
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
CDAAC	COSMIC Data Analysis and Archive Center
CFSR/CFSv2	NCEP Climate Forecast System Reanalysis/Climate Forecast System, version 2
CLAES	Cryogenic Limb Array Etalon Spectrometer
CNES	Centre National d’Etudes Spatiales
COSMIC	Constellation Observing System for Meteorology, Ionosphere and Climate
CPC	Climate Prediction Center
CTM	Chemical Transport Model
DAO	Goddard Space Flight Center Data Assimilation Office
DOE	Department of Energy
ECMWF	European Centre for Medium-range Weather Forecasts
ECMWF OA	ECMWF Operational Analysis
ERA-40	ECMWF 40-year reanalysis

ERA-Interim	ECMWF interim reanalysis
ERA5	the fifth major global reanalysis produced by ECMWF
ESA	European Space Agency
FLEXPART	Flexible Particle model
FOV	Field Of View
GEOS	Goddard Earth Observing System
GMAO	NASA Global Modeling and Assimilation Office
GNSS	global navigation satellite system
GNSS-RO	global navigation satellite system - radio occultation
GPS	Global Positioning System
ICE	used to indicate water-ice particle type in PSC classification schemes
IFS	Integrated Forecast System
JRA-55	Japanese 55-year Reanalysis
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MJJASO	May-June-July-August-September-October
MLS	Microwave Limb Sounder
MPTRAC	Massive-Parallel Trajectory Calculations
NAT	Nitric Acid Trihydrate
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NCEP-DOE R2	Reanalysis 2 of the NCEP and DOE
NCEP-NCAR R1	Reanalysis 1 of the NCEP and NCAR
NH	Northern Hemisphere
NMC	US National Meteorological Center (now NCEP)
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
ppmv / ppbv	parts per million by volume / parts per billion by volume
PSC	Polar Stratospheric Cloud
PV	Potential Vorticity
REM	Reanalysis Ensemble Mean
RHTD	relative horizontal transport deviation
RMS	Root Mean Square
SD	Standard Deviation (used here for standard deviation of temperature differences)
SH	Southern Hemisphere
S-RIP	SPARC Reanalysis Intercomparison Project
SSW	Sudden Stratospheric Warming
STS	Supercooled Ternary Solutions
TOVS	TIROS Operational Vertical Sounder
TSEN	Thermodynamical SENsor
UCAR	Universities for Cooperative Atmospheric Research
UKMO	UK Meteorological Office
V_{PSC}	Volume of air with temperatures below the PSC threshold
WMO	World Meteorological Organization

Chapter 11: Upper Stratosphere and Lower Mesosphere

Chapter lead authors

V. Lynn Harvey	(1) Laboratory for Atmospheric and Space Physics (2) Atmospheric and Oceanic Sciences Department, University of Colorado	USA
John Knox	Department of Geography, University of Georgia	USA

Co-authors

Jeff France	White Ridge Solutions	USA
Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Lesley Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK
Toshihiko Hirooka	Department of Earth and Planetary Sciences, Kyushu University	Japan
Peter Hitchcock	Department of Earth and Atmospheric Sciences, Cornell University	USA
Matthew Hitchman	Department of Atmospheric and Oceanic Sciences, University of Wisconsin – Madison	USA
Yoshio Kawatani	Japan Agency for Marine-Earth Science and Technology	Japan
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
John McCormack	Heliophysics Division, Science Mission Directorate, NASA Headquarters *previously at Space Science Division, Naval Research Laboratory*	USA
Yvan Orsolini	Norwegian Institute for Air Research	Norway
Takatoshi Sakazaki	Graduate School of Science, Kyoto University	Japan
Yoshihiro Tomikawa	(1) National Institute of Polar Research (2) The Graduate University for Advanced Studies (SOKENDAI)	Japan

Abstract. This chapter focuses on the uppermost levels in the reanalyses, where assimilated data sources are most sparse. The first part of the chapter includes a brief discussion of the effects of the model top and physical parameterizations relevant to the the Upper Stratosphere and Lower Mesosphere (USLM) . Long-term signatures of discontinuities in data assimilation and variability among reanalyses are then presented. A climatology of the basic state variables of temperature, horizontal winds, and residual circulation velocities is given. The climatology includes estimates of variability among the reanalyses. Annual cycles highlight the dependence of reanalysis difference on time of year. We then document dominant modes of variability in the reanalyses in the tropical regions and at high latitudes, and longer-term variability including solar cycle, volcanic, ENSO, and QBO signals. The tropical Semi-Annual Oscillation (SAO), the middle atmosphere Hadley circulation, and the occurrence of inertial instability are compared among the reanalyses. High latitude processes considered include polar vortex variability and extreme disruptions therein observed during “elevated stratopause” events. Planetary wave amplitudes are quantified and compared to observations. The chapter ends with a comparison of solar atmospheric tides, 2-day wave amplitudes, and 5-day wave amplitudes in the USLM.

Content

11.1 Introduction 535
11.1.1 Reanalysis products used in this chapter 535
11.1.2 Satellite observational products 536
11.1.3 Upper boundary conditions in the reanalyses 537
11.1.4 Physical parameterizations in the reanalyses specific to USLM 537
11.1.5 Long-term effects of data assimilation discontinuities 537
11.1.6 Variability among reanalyses 539
11.2 Climatology of the USLM 541
11.2.1 Seasonal zonal means 541
11.2.2 Annual cycles 546
11.2.3 Long-term variability 550
11.3 Tropical dynamics 554
11.3.1 Semi-annual oscillation 554
11.3.2 Middle-atmosphere Hadley circulation 556
11.3.3 Inertial instability 557
11.4 Polar dynamics 559
11.4.1 Polar vortices 559
11.4.2 Planetary waves 561
11.4.3 Elevated stratopause events 561
11.5 Tides and normal modes 564
11.5.1 Tides 564
11.5.2 Quasi-2-day wave 566
11.5.3 Quasi-5-day wave 569
11.6 Summary, key findings, and recommendations 571
References 574
Major abbreviations and terms 581

11.1 Introduction

It is now widely accepted that there are vertical coupling processes that allow the stratosphere to impact surface weather patterns (*Baldwin and Dunkerton*, 2001). Conversely, the upper atmosphere (thermosphere and ionosphere) is strongly influenced by near-surface processes that generate waves that propagate through the stratosphere and mesosphere (*Pedatella et al.*, 2018; *Liu et al.*, 2011; *Goncharenko et al.*, 2010). In between, in the region we refer to as the upper stratosphere and lower mesosphere (USLM), there is a two-way interaction that both affects and is affected by meteorological processes. The modeling of the USLM is therefore critically dependent on meteorological data in the lower and middle atmosphere (*e.g.*, *Pedatella et al.*, 2014) as well as the parameterization of meteorological processes (*e.g.*, sub-grid-scale gravity waves). This chapter compares the reanalysis datasets in the USLM. It is important to document differences at these upper levels because this is the region of the atmosphere where assimilated observations are sparse and differences among the reanalyses are large (*e.g.*, *Chapter 3*; *Long et al.*, 2017). Without an abundance of data to tether forecast models to observations in the USLM, differences in forecast model details may play a prominent role in explaining differences in reanalyses.

The USLM region, apparently first defined as such in the literature by *Gerrard et al.* (2002), is home to numerous dynamical, chemical and other processes of importance for understanding the Earth's atmosphere. Dynamically, at the shortest time scales, mesoscale gravity waves break as they amplify in the rarefied atmosphere or reach critical layers, causing "drag" that closes jet streams and drives meridional circulations, with concomitant thermal and chemical effects (*e.g.*, *Fritts and Alexander*, 2003). Migrating diurnal, semi-diurnal and terdiurnal tides lead to large temperature and wind perturbations in this region (*e.g.*, *Lilienthal et al.*, 2018; *Hagan and Forbes*, 2003). Planetary waves affect this region as well via upward propagation. Perhaps the most spectacular phenomenon, the sudden stratospheric warming (SSW), is driven by planetary wave absorption and breaking and leads to major changes in temperature, wind, and chemistry especially in the high latitudes on the time scales of days to weeks (*Butler et al.*, 2017). At longer time scales, the Semi-Annual Oscillation (*e.g.*, *Smith et al.*, 2017) and the Quasi-Biennial Oscillation (*Baldwin et al.*, 2001), both driven by complex combinations of wave forcings at multiple spatial scales, represent major dynamical and thermal reversals that define many aspects of the tropical and extratropical middle and upper atmosphere. At even longer time scales, changes in the USLM due to solar variability can and do affect its thermal, dynamical and chemical characteristics (*e.g.*, *Beig et al.*, 2008). There is also a rich interplay among these different phenomena that, in many cases, is still to be understood. Since the drivers of these phenomena generally originate in the lower atmosphere and impact regions extending up to the top of the atmosphere, the USLM is in some sense the gateway between "weather" in the troposphere and "space weather" in the thermosphere/ionosphere (*Baker et al.*, 2019). Clearly, an understanding of the USLM benefits not only researchers of the USLM, but those interested in weather and climate from the top to the bottom of the Earth's atmosphere.

In this chapter, we first describe the reanalyses and satellite observations used to evaluate the reanalyses. This is followed by a brief discussion of the effects of the model top and physical parameterizations relevant to the USLM. Long-term signatures of discontinuities in data assimilation and variability among reanalyses are then presented. A climatology of the basic state variables of temperature, horizontal winds, and residual circulation velocities are given. The climatology includes estimates of variability among the reanalyses. Annual cycles highlight the dependence of reanalysis differences as a function of time of the year. We then document dominant modes of variability in the reanalyses in the tropical regions and at high latitudes, and longer-term variability including solar cycle, volcanic, El Niño Southern Oscillation (ENSO), and Quasi-Biennial Oscillation (QBO) signals. The tropical Semi-Annual Oscillation (SAO), the middle-atmosphere Hadley circulation, and the occurrence of inertial instability are compared among the reanalyses. Polar phenomenology evaluated here comprise the polar vortices and extreme disruptions therein observed during "elevated stratopause" events. Planetary wave amplitudes are quantified and compared to observations. The chapter ends with a comparison of solar atmospheric tides, 2-day wave amplitudes, and 5-day wave amplitudes in the USLM.

11.1.1 Reanalysis products used in this chapter

Table 11.1 lists the reanalysis datasets examined in this chapter. These include reanalyses considered in the overall S-RIP project (*Fujiwara et al.*, 2017; also see, *e.g.*, the list given here **https://s-rip.ees.hokudai.ac.jp/pubs/reanalysis.html**) that cover the USLM region with upper air observations assimilated, *i.e.*, the Modern Era Retrospective analysis for Research and Applications version 2 (MERRA-2), MERRA, the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim), the ERA-40, and the latest ERA-5, the Japanese 55-year Reanalysis (JRA-55), the JRA-55C (with only conventional data assimilated), the JRA-55AMIP (with no data assimilated and only constrained by sea surface temperatures), the JRA-25 (covering 25 years from 1979 - 2004) and the combined data records from the Climate Forecast System Reanalysis of NCEP (CFSR) and the Climate Forecast System, version 2 (CFSv2). Note that CFSR is excluded in many comparisons shown in this chapter. The reason for this is that the CFSR top is lower than in the other reanalyses (see *Chapter 2*) and that no pressure levels above 1 hPa are post-processed.

When CFSR is included in comparisons it is only for completeness. For MERRA-2, the M2I3NPASM collection (*GMAO*, 2015a) was used in *Sections 11.1.6*, *11.2* and *11.3*, the M2I6NVANA collection (*GMAO*, 2015b) was used in *Section 11.4*, and the M2I3NVASM collection (*GMAO*, 2015c) was used in *Sections 11.5.2* and *11.5.3*. Users of MERRA-2 reanalyses should be aware of differences between "ANA" and "ASM" products (*GMAO*, 2017) especially when calculating long-term trends in the Hadley Cell (*Garfinkel et al.*, 2015) and when modeling lower stratospheric transport (*Orbe et al.*, 2017). State-of-the-art data assimilation (DA) models such as the Whole Atmosphere Community Climate Model with Data Assimilation Research Testbed are not run operationally and thus are not considered "reanalyses" for the purposes of S-RIP. Comparison of reanalyses, observations, and high-top DA models is the subject of future work. A few essential details of the reanalysis systems relevant to the USLM are given in the references in **Table 11.1**, including the model top altitude and the gravity wave parameterizations that play a key role in the characteristics of this region. The reader is directed to *Chapter 2* and *Fujiwara et al.* (2017) for more details on the models, including a comprehensive list of assimilated observations, model parameterizations, and changes to the models over time. In the following sections, key points regarding the reanalyses and the observations to which they are compared are briefly summarized for ease of reference.

Table 11.1: *List of reanalysis datasets used in this chapter, overall references, model top altitude, and gravity wave specifications. In the 4th column, ORO refers to the parametrization for orographic gravity waves while NON refers to that of non-orographic gravity waves.*

Reanalysis Dataset	Reference	Model Top (hPa)	Gravity Wave Drag Parameterizations
MERRA	*Rienecker et al.* (2011)	0.01	ORO: *McFarlane* (1987) NON: *Garcia & Boville* (1994)
MERRA-2	*Bosilovich et al.* (2015); *Gelaro et al.* (2017); *Molod et al.* (2015)	0.01	ORO: *McFarlane* (1987) NON: *Garcia & Boville* (1994); *Molod et al.* (2015)
ERA-40	*Uppala et al.* (2005)	0.1	ORO: *Lott & Miller* (1997) NON: none
ERA-Interim	*Dee et al.* (2011)	0.1	ORO: *Lott & Miller* (1997) NON: none
ERA5	*Hersbach & Dee* (2016)	0.01	ORO: *Lott & Miller* 1997; NON: *Orr et al.* (2010)
JRA-55 **JRA-55C** **JRA-55AMIP**	*Kobayashi et al.* (2015)	0.1	ORO: *Iwasaki et al.* (1989a, b) NON: none
JRA-25	*Onogi et al.* (2007)	0.4	ORO: *Iwasaki et al.* (1989a, b) NON: none
NCEP-CFSR	*Saha et al.* (2010)	~0.266	ORO: *Kim & Arakawa* (1995); *Lott & Miller* (1997) NON: none
CFSv2	*Saha et al.* (2014)	~0.266	ORO: *Kim & Arakawa* (1995); *Lott & Miller* (1997) NON: *Chun & Baik* (1998)

11.1.2 Satellite observational products

Throughout this chapter and in science studies that employ reanalyses, it is critically important to compare the reanalyses to independent data sources whenever possible. This analysis step acts to quantify model biases (either known or unknown) and establishes a level of consistency between the reanalysis fields and observations. The following satellite datasets appear in this chapter for this purpose.

The Earth Observing System (EOS) Microwave Limb Sounder (MLS) satellite data record spans August 2004 to the present and provides ~3500 vertical profiles of temperature, geopotential height (from which horizontal winds can be derived) and trace gases each day that cover the globe (*Waters et al.*, 2006). The retrieval methods and error estimates for the most recent version-4 data products are given by *Livesey et al.* (2017). Some of the comparisons among reanalyses shown in this chapter also include MLS data to provide a (mostly) independent reference point. Note that MERRA-2 assimilates MLS temperature at pressures 5 hPa and less starting in October 2004. Likewise, ERA-Interim, MERRA-2, and ERA5 assimilate MLS ozone profiles starting in 2008, 2004, and 2004, respectively (see *Chapter 2* and **Tables 2.10**, **2.20**, and **2.22**). An advantage of comparing reanalyses to MLS observations is that the consistent near-global coverage does not result in data gaps at polar latitudes.

The Thermosphere, Ionosphere, Mesosphere, Energetics and Dynamics (TIMED) satellite launched in December 2001 provides daily near-global measurements of ozone, temperature, and geopotential height from the troposphere up to the lower thermosphere. The Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) instrument aboard TIMED is a 10-channel radiometer that measures infrared Earth limb emissions. On any given day, SABER observes from the ~82° latitude in one hemisphere to 52° latitude in the other. The TIMED satellite then "yaws" to allow SABER to view the other pole every ~60 days. SABER temperature profiles have 2 km vertical resolution and estimates of precision are within 4 K throughout the mesosphere (*Remsberg et al.*, 2003). Since SABER data are not currently assimilated into any of the reanalysis systems they provide an important independent reference point for comparison in the USLM.

Ultimately, the observational datasets employed in each of the following sub-sections depend on the choice of the author(s) writing each sub-section. Future users of reanalysis data products in the USLM are encouraged to compare to other data sources not necessarily listed above. These include other national and international satellite data, ground-based radars and lidars, and sub-orbital rocket data. While some of these data sources are inherently geographically sparse, all provide an invaluable source of independent data with which to establish the fidelity of USLM reanalysis output products and place scientific results on more solid footing.

11.1.3 Upper boundary conditions in the reanalyses

The results in this chapter are, to a larger degree than in the other S-RIP chapters, impacted by the effects of proximity to the model top. The approximate vertical grid spacing of the reanalyses becomes coarser nearer the model top, from about 1.5 km for MERRA and MERRA-2 in the upper stratosphere to 8 km for the CFSR and CFSv2 in the lower mesosphere. See **Figure 3** in *Chapter 2* for more detailed information on vertical grid spacing. Sponge layers are employed in the models used in the reanalyses to avoid problems created by a "rigid lid" top, which would otherwise spuriously reflect wave energy. Different reanalyses implement sponge layers differently, thus the reader is referred to S-RIP *Chapter 2* for more information and references regarding sponge layers and model top treatments. Briefly, the MERRA/MERRA-2 sponge layer is applied by increasing the divergence damping coefficient in the topmost nine layers, while also reducing the order of advection in the top model level to a first-order scheme (*Bill Putman, personal communication*, 2018). The ERA reanalyses' sponge layer covers the atmosphere above 10 hPa in which an additional term is applied to horizontal diffusion specifically to absorb vertically propagating gravity waves. This means the K coefficient is multiplied by a factor that depends on wavenumber and model level, consistent with "enhanced hyperdiffusion in the sponge layer". Rayleigh drag is also applied above 10 hPa in ERA-40 and above the stratopause in ERA-Interim. In JRA-25 and JRA-55, gradually larger horizontal diffusion coefficients are applied for pressures of less than 100 hPa in the data assimilation system, and Rayleigh friction is also applied to the temperature deviations from the global average at pressures less than 50 hPa (see *Chapter 2*, **Table 2.3**). CFSR, like JRA, also uses gradually increasing horizontal diffusion coefficients with increasing height; like ERA, CFSR also uses Rayleigh drag beginning in the upper stratosphere, at pressures less than 2 hPa. It is important to keep in mind that sponge layers, while minimizing the obvious contamination effects of unrealistic gravity wave reflection, can also introduce spurious effects of their own in winds and temperature (*Shepherd et al.*, 1996).

11.1.4 Physical parameterizations in the reanalyses specific to USLM

The model parameterizations that have the largest impact in the USLM involve horizontal diffusion and the methods that account for small-scale gravity waves, both orographic and non-orographic. Horizontal diffusion is represented by implicit linear 4th order (for ERA-40, ERA-Interim, JRA-25, and JRA-55) or 8th order (for CFSR/CFSv2) diffusion in spectral space. For MERRA and MERRA-2, horizontal diffusion is accounted for using explicit 2nd order horizontal divergence damping. See *Chapter 2* **Table 2.8** for differences in the treatment of horizontal diffusion among the reanalyses. Each reanalysis center treats orographic gravity waves differently (see **Table 11.1** for references that describe each scheme; see *Chapter 2*, **Table 2.7** for additional details). Of the reanalysis datasets used in this chapter, only MERRA, MERRA-2, and CFSv2 apply non-orographic gravity wave drag schemes (see *Chapter 2*, **Table 2.7** and *Fujiwara et al.* (2017), **Table 3** and discussions therein for more details). The MERRA-2 non-orographic gravity wave parameterization has been modified from that used in MERRA by increasing the background source at certain latitudes and by increasing the intermittency (*Molod et al.*, 2015). As noted in the previous section, Rayleigh drag is used in several of the reanalyses, both in the USLM and below it, both to simulate non-orographic gravity wave drag (in the ERA-Interim) and also more generally as a damping/sponge effect. The use of Rayleigh drag to simulate non-orographic gravity wave drag is problematic, since these waves can have substantial phase speeds.

11.1.5 Long-term effects of data assimilation discontinuities

Reanalysis efforts aim to minimize the effects of discontinuities to a given assimilation system. However, there remain discontinuities in the data records due to differences in the data that are assimilated from year to year (*e.g., Chapter 3; Long et al.*, 2017; *Simmons et al.*, 2014). To discuss discontinuities, we need to describe (1) major observations assimilated, and (2) execution streams, because they are the main sources of discontinuities in reanalysis time series. What follows is a brief synopsis of *Chapter 3.3*, where additional details can be found.

Because radiosondes generally do not reach the upper stratosphere, the major observations that are assimilated in this region are satellite-based. The ERA-40 reanalysis used SSU data, which introduced discontinuities because of the multiple NOAA polar orbiters that provided SSU data. The ERA-Interim has biases as a result of the polar orbiter issue in 1985 and, in 1998, the transition from TOVS to ATOVS. The reader is referred to *McLandress et al.* (2014) for a characterization of these discontinuities and a numerical method to remove them. Similarly, the JRA-25 also has a discontinuity in 1998 because of the transition from TOVS to ATOVS.

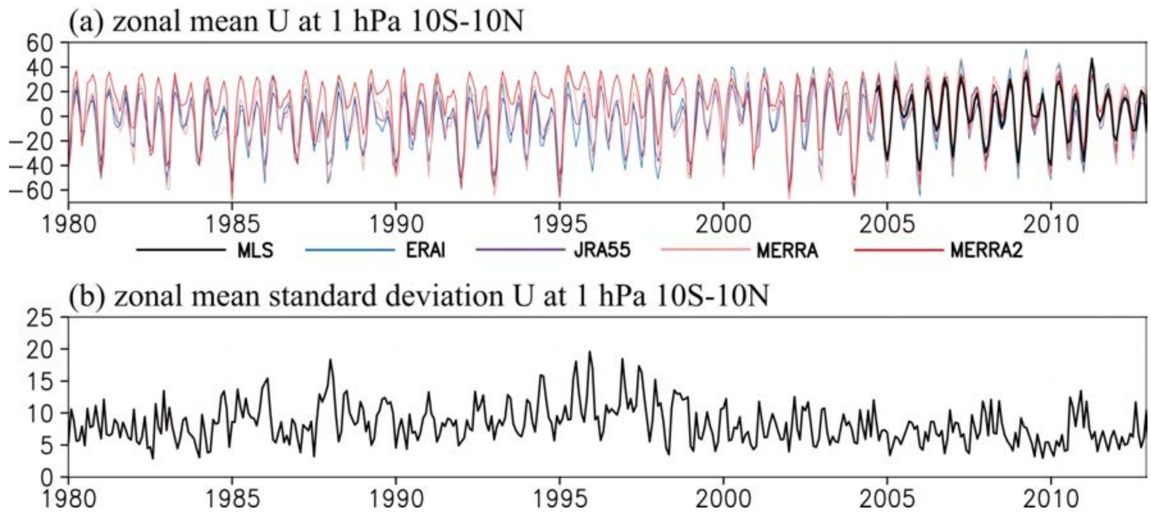


Figure 11.1: *35-year time series at 1hPa from 10°S-10°N of (a) zonal mean zonal wind in 4 reanalyses and MLS and (b) zonal mean zonal wind standard deviation among the reanalyses. Winds are in ms^{-1}. Figure 11.1a is modified from Kawatani et al. (2020).*

The CFSR has multiple discontinuities in the middle and upper stratosphere in 1986, 1989, 1994, 1999, 2005, 2009 due to multiple execution streams and a biased SSU bias correction method. MERRA-2 has discontinuities in 1995 (SSU), 1998 (transition to AMSU) and 2004 (transition to MLS Aura). In general, the discontinuities are less of a factor from 1998 onward, in the post-SSU era.

Now we examine a few examples of how the discontinuities affect phenomena in the USLM. **Figure 11.1** reveals some of the impacts of these discontinuities in winds near the stratopause (at 1 hPa, near 50 km). The black line in panel (**a**) shows monthly and zonally averaged zonal winds estimated by MLS geopotential heights (*Smith et al.,* 2017). This method obtains tropical zonal winds by cubic spline interpolation of the balanced winds across the Equator. The zonal winds estimated using MLS and SABER are in good agreement at most altitudes (see **Figure 1** of *Smith et al.*, 2017). **Figure 11.1** shows that reanalysis zonal winds are in agreement with those estimated from MLS data. A notable difference among the reanalyses is that the westerlies in MERRA-2 are larger than the other reanalyses during the 1980s and most of the 1990s. This MERRA-2 bias prior to 1998 results in larger standard deviations among the reanalyses, as shown in panel (**b**). The reader is referred to *Kawatani et al.* (2020) for representation of the equatorial zonal wind in the USLM in several reanalyses and satellite observations.

In order to quantify the spread among reanalyses, the 3D standard deviations among the reanalyses ($\sqrt{\sum(u_i - [u])^2/N}$) are calculated, where i labels the individual datasets and there are N datasets included. The square brackets denote the mean over all N reanalyses (*cf.*, *Kawatani et al.*, 2016). The standard deviation is calculated for each month using monthly mean zonal wind. **Figure 11.1** panel (**b**) shows that the zonal mean standard deviation among the reanalyses is larger prior to 1998 compared to the period after 1998, consistent with expectations.

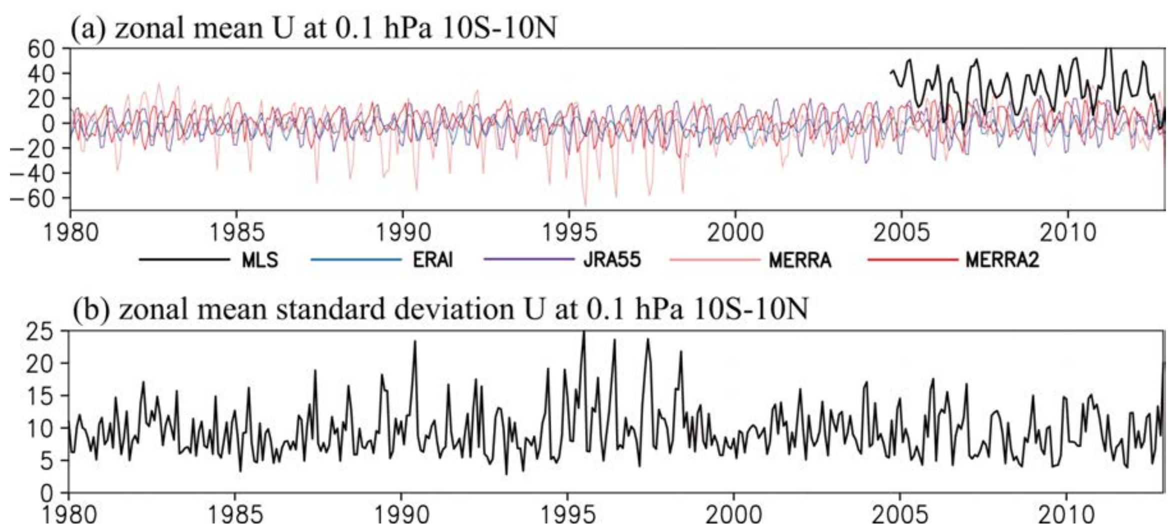


Figure 11.2: *Same as **Figure 11.1** except at 0.1 hPa.*

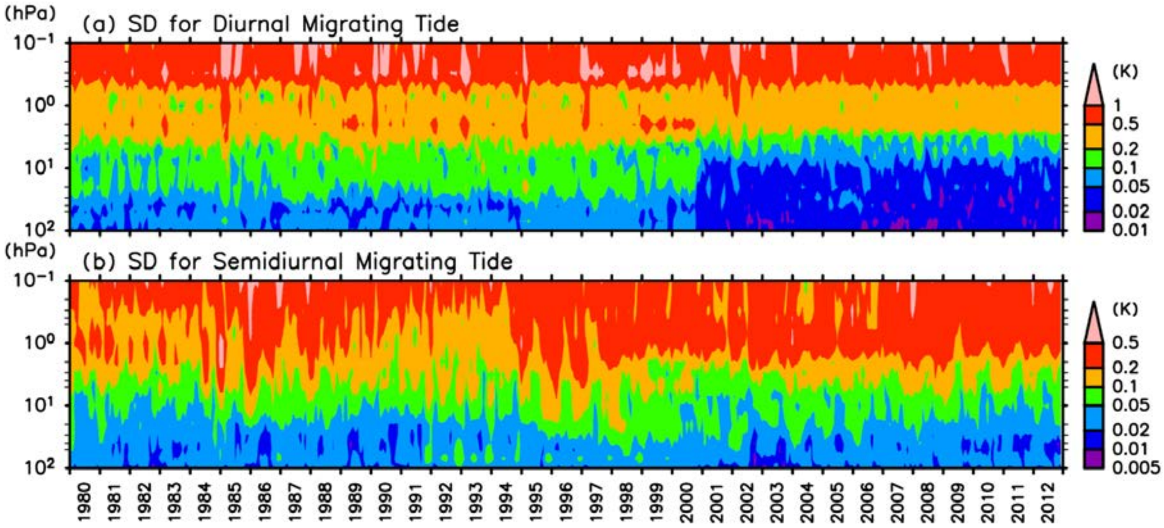


Figure 11.3: *32-year altitude-time section of the standard deviation averaged over 10 °S - 10 °N in the amplitude of the (a) diurnal and (b) semidiurnal temperature tide using MERRA, MERRA-2, ERA-Interim and JRA-55.*

Temporal discontinuities in the zonal winds are especially obvious at lower mesospheric altitudes. **Figure 11.2** of zonal winds at 0.1 hPa (near 60 km) shows strong easterlies in MERRA for the period before 1998 (panel **a**) that are not corroborated by the other reanalyses.This MERRA bias results in the standard deviations among the reanalyses being larger prior to 1998 (panel **b**). These results are consistent with *Das et al.* (2016) (see their **Figure 4a**, top panel) who showed 1) a MERRA easterly bias in the tropical zonal winds in the USLM (compared to rocket observations) during 1979 - 1991 and 2) that this easterly bias disappeared after 1998. Reanalyses zonal winds do not agree with the MLS winds due to different time-mean values (~ 28 m s^{-1} for the MLS and ~ 0 m s^{-1} for reanalyses, see also **Fig. 11.6**).

Atmospheric tides are also affected by temporal discontinuities in input data. In **Figure 11.3**, the impact of differences in the data assimilation stream on the diurnal (panel **a**) and semidiurnal (panel **b**) tides is depicted. See *Sakazaki et al.* (2018) for interannual variability in individual reanalysis datasets. In the stratosphere, the variance in the diurnal temperature tidal amplitudes among the reanalyses is significantly larger (note the logarithmic color scale) before 2000 compared to later years. In the lower mesosphere the variance among the reanalysis data sets is large (~ 1 K) and fairly steady throughout the entire record. The reanalyses are presumably strongly dependent on the tides simulated in the forecast model used in producing each reanalysis in this altitude region. An abrupt change due to the TOVS-to-ATOVS transition around 2000 is not apparent for the semidiurnal tide (panel b). However, a decadal variation is seen before ~2000 (*e.g.*, relatively large variance around 1985 and 1995 above 10 hPa level). This is caused by anomalously large interannual variations in ERA-Interim, which are likely related to the orbital drift of TOVS and the transition between different NOAA satellites carrying the TOVS (*Sakazaki et al.*, 2018)

11.1.6 Variability among reanalyses

As shown in the previous section, significant differences in the tropical zonal winds exist among the different reanalysis datasets. This section quantifies those differences for zonal mean temperature and zonal mean zonal winds over a range of latitudes and altitudes. These differences are likely conservative estimates and day-to-day differences among the reanalyses are likely larger. Data users are advised to keep in mind the magnitude of these differences when drawing scientific conclusions.

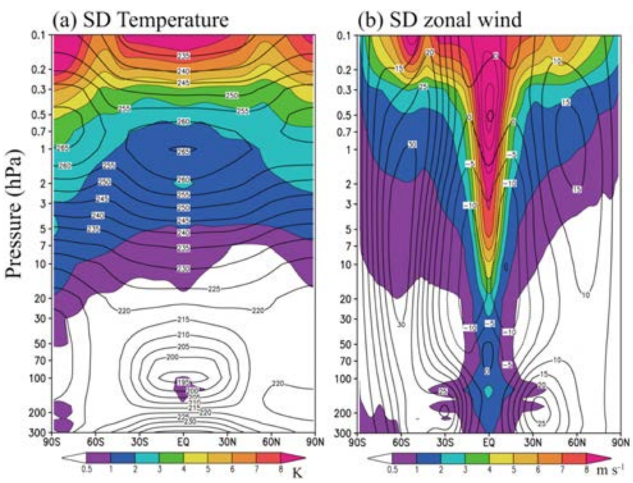


Figure 11.4: *Latitude-altitude distribution of zonal mean and time mean (1980 - 2012) standard deviation (color fill) of (a) temperature and (b) zonal winds among the four reanalyses (ERA-Interim, JRA-55, MERRA, MERRA-2). Annually averaged zonal-mean temperature contours every 5 K are added to (a) and zonal winds contours every 5 m s^{-1} are added to (b) where westerlies are solid and easterlies are dotted. Modified from Kawatani et al. (2020).*

In **Figure 11.4**, a comparison of cross-sections of zonal mean and time mean (1980 - 2012) standard deviations of temperature (panel **a**) indicates increasing differences among four reanalyses (ERA-Interim, JRA-55, MERRA, and MERRA-2) with height into the mesosphere at all latitudes. The latitudinal dependence of the differences is rather weak, though the differences are somewhat smaller in the equatorial upper stratosphere and slightly larger in the Southern Hemisphere (SH) stratosphere and mesosphere than in the Northern Hemisphere (NH). Above about 10 hPa there are limited data and different models create different solutions that lead to differences in variability. Another factor contributing to variability among the reanalyses are sponge layers in the mesosphere of the JRA-55 and ERA-Interim models, leading to the broad large standard deviation region above ~1 hPa. MERRA and MERRA2, which have a much higher top, do not have significant sponge-layer damping below 0.2 hPa (see *Chapter 2; Kawatani et al.*, 2020).

The zonal wind standard deviation (**b**) increases in the equatorial region and with increasing height. There is a notable "v-shape" of increased variability in the Tropics in **Figure 11.4b**. These results are consistent with Das et al. (2016) (see their **Figure 5a**, left panel) who showed root-mean-square differences between reanalysis and rocket observed zonal winds in the Tropics that increased linearly with altitude. One possible explanation is the relatively weak constraints of the thermal wind balance associated with satellite temperature observations over the Equator (*Kawatani et al.*, 2016). Without strong data constraints in the USLM (see previous section), differences among reanalyses such as MERRA and MERRA-2 are likely ascribable to model differences (*Krzysztof Wargan, personal communication*, 2018). The lack of a non-orographic gravity wave parameterization in JRA-55 may also contribute to its variability relative to other reanalyses (*Yayoi Harada, personal communication*, 2018).

Figure 11.5 depicts standard deviations of temperature in the lower mesosphere (**a**) and at the stratopause (**c**) and for zonal wind in the lower mesosphere (**b**) and at the stratopause (**d**) in plan view. Standard deviations are averages from 1980 to 2012. Overall, the temperature and zonal wind standard deviations are much larger in the mesosphere (top panels) than in the stratosphere (bottom panels). Panels (**a**) and (**c**) show that temperature differences among the reanalyses are relatively large in the polar region at both levels. There are no data assimilated at 0.1 hPa (except MLS in MERRA-2 after August 2004), so these differences result largely from model performance. Panel (**b**) shows that in the lower mesosphere, areas of large standard deviation in the zonal wind are globally spread with two maxima at the Equator and 50 °S. Smaller variability over the North Pacific at 0.1 hPa compared to other longitude sectors may be due to weaker zonal winds associated with the climatological Aleutian anticyclone (*Harvey and Hitchman*, 1996).

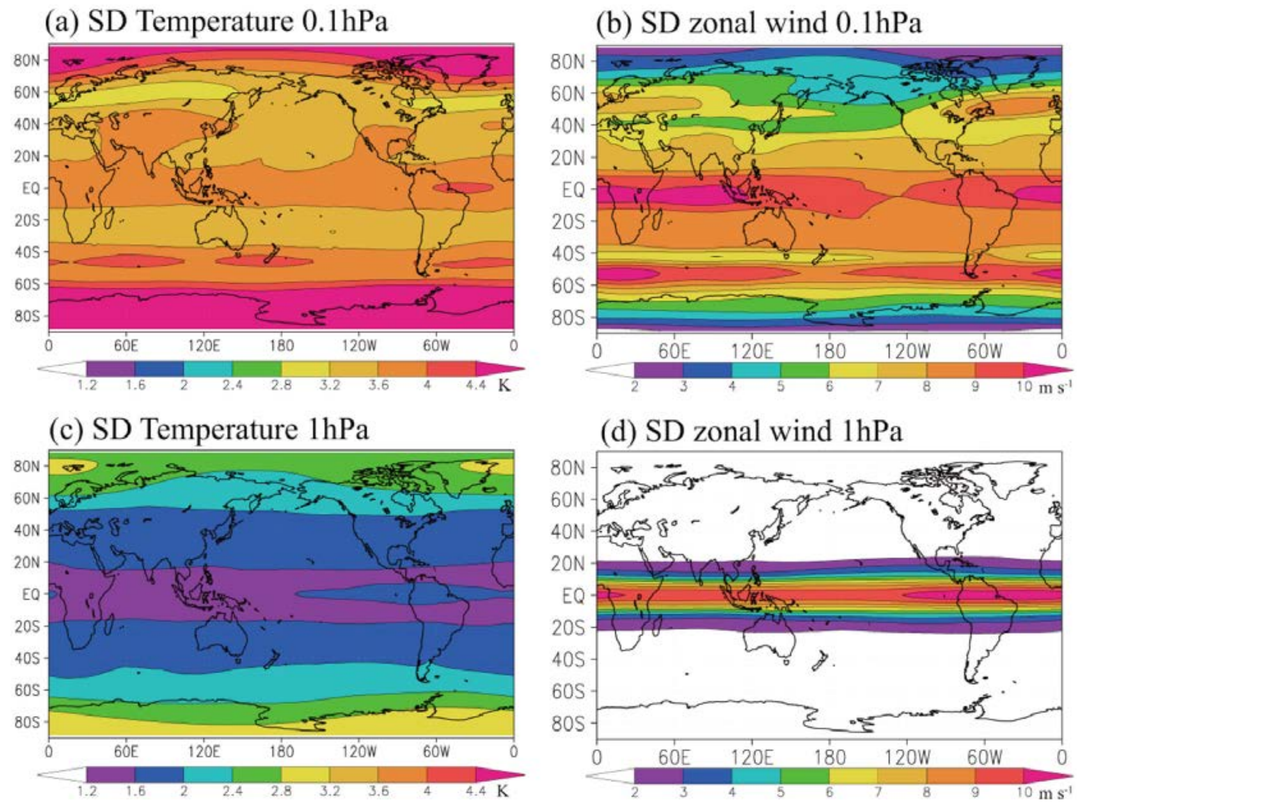


Figure 11.5: *Mercator maps at 0.1 hPa (top) and 1 hPa (bottom) of the standard deviation among reanalysis temperature (panels a and c) and zonal wind (panels b and d). Temperature is in degrees K and wind is in m s^{-1}. Figures 11.5b and 11.5d are modified from Kawatani et al. (2020).*

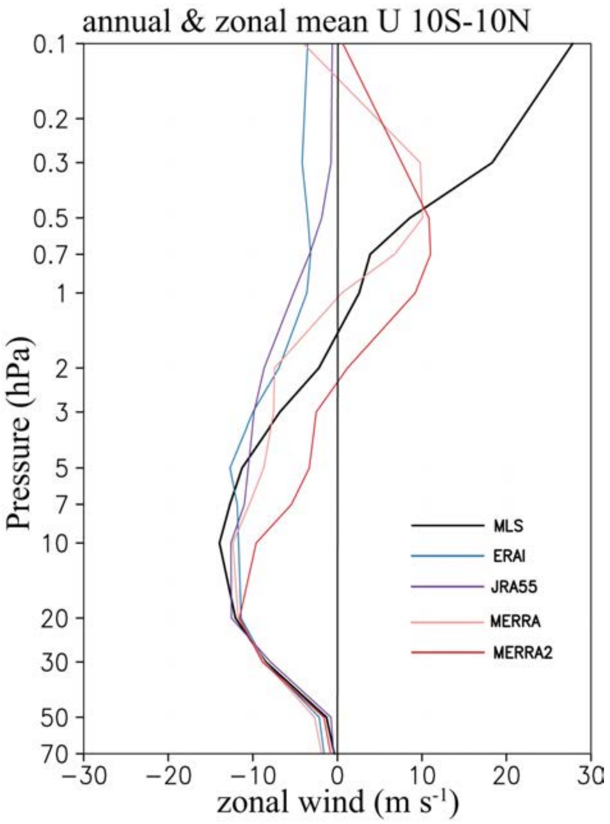


Figure 11.6: *Vertical profiles of climatological annual mean zonal mean zonal winds averaged from 10 °S - 10 °N in four reanalyses (MERRA, MERRA-2, ERA-Interim, and JRA-55). MLS gradient winds are shown in black. Modified from Kawatani et al. (2020).*

The largest spread in the SH occurs where a strong polar night jet exists. Large differences among the reanalyses may trace back to the large absolute value of variability in this region. At 1 hPa (panel **d**) areas of large standard deviation are well concentrated near the Equator with relatively small zonal variability. Note that the 0.1 hPa levels correspond to sponge layers in ERA-Interim and JRA-55. Due to this, the physical interpretation of differences in the mesosphere is less clear.

Figure 11.6 summarizes the climatological (1980 - 2012) annual average inter-reanalyses differences in the zonal mean zonal wind near the Equator, where wind differences maximize. The black line shows the MLS zonal wind averaged from September 2004 to August 2014. Westerly biases in MERRA-2 are large between 10 hPa and 1 hPa. Since MERRA-2 includes a non-stationary gravity wave parameterization it may result in a westerly bias (*Coy et al.*, 2016; *Molod et al.*, 2015). More investigation is needed to quantify this potential bias. Westerly MLS gradient winds are stronger than all of the reanalyses above 0.5 hPa. *Section 11.3* will demonstrate that this is due to the emergence of summer SAO easterlies in most reanalyses that offset equinox westerlies and time-average to be near zero. Weak time-mean zonal winds in ERA-I and JRA55 above 1 hPa are attributed to sponge layer effects. However, MERRA and MERRA2 time-mean winds also converge to zero but do not suffer as much from sponge layer effects at these altitudes. The reader is cautioned that satellite-derived gradient wind approximations are not particularly accurate in the equatorial mesosphere because of the importance of Reynolds stress terms, *e.g.*, from the diurnal tide, and thus the MLS wind values at 0.1 hPa may be overestimated (*McLandress et al.*, 2006). That said, MLS and SABER agree with each other (*Smith et al.*, 2017) and with rocketsonde observations (*Kishore Kumar et al.*, 2015). Overall, users of reanalysis zonal wind data in the tropical USLM are cautioned to keep these large differences in mind when drawing scientific conclusions.

11.2 Climatology of the USLM

In this section we describe the climatology of the USLM in different reanalyses, giving a closer look at this region than provided in *Chapter 3*. Reanalysis data used in this section include MERRA, MERRA-2, ERA-Interim, JRA-25, JRA-55, and ERA-40. For comparison, data were interpolated to common grids by taking the following steps: 1) Monthly- and zonal-mean data were computed from the original data. 2) Data on common pressure levels were extracted. These pressure levels consist of the following 26 levels: 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, 1, 0.7, 0.5, 0.3, 0.1 hPa. 3) All the common pressure level data were linearly interpolated onto the common latitudinal grid every 1.5 degree (*i.e.*, same as ERA-Interim). 4) The 1980 - 2001 common time period is used. This time period precludes large differences that would arise if years since 2004 were included, due to the assimilation of MLS into MERRA-2. Note that this common time period is different from that used in *Chapter 3* (1980 - 2010), hence anomalies may be different. Residual-mean meridional and vertical velocities are given by the Transformed Eulerian-Mean (TEM) formulation (*Andrews et al.*, 1987).

11.2.1 Seasonal zonal means

Here we provide a global "atlas" of seasonal (*i.e.*, December-January-February (DJF), March-April-May (MAM), June-July-August (JJA), and September-October-November (SON)) zonal means of temperature, Eulerian zonal and meridional winds, and Lagrangian TEM residual circulation velocities averaged over 1980 - 2001 for each of the six reanalyses listed above and their differences from MERRA (Reanalyses minus MERRA). MERRA was chosen as a reference because it covers the lower mesosphere above 1 hPa and has been widely used. However, this does not imply that MERRA should be considered "truth". In fact, if differences are similar among multiple reanalyses then this likely indicates that MERRA is biased.

Figure 11.7 shows latitude-pressure sections of seasonal zonal means of temperature. Here, data above 1 hPa are available only for MERRA and MERRA-2.

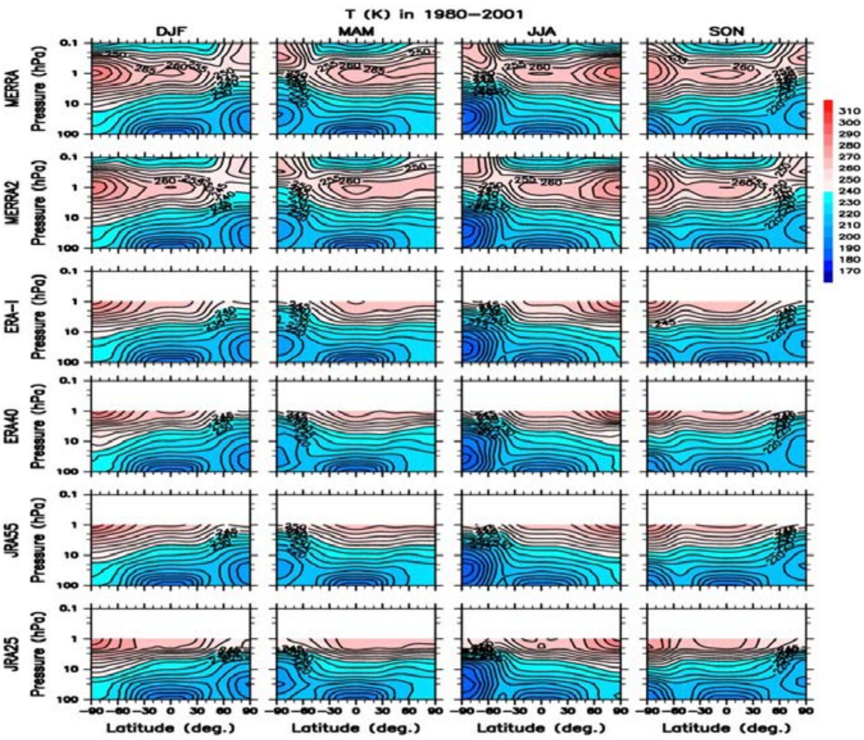


***Figure 11.7:** Latitude-pressure sections of seasonal zonal-mean temperature averaged over 1980 - 2001 from MERRA, MERRA-2, ERA-Interim, ERA-40, JRA-55, and JRA-25. Contour intervals are 5 K.*

Overall features of the temperature distributions are similar among all the available reanalyses except for JRA-25, in which an anomalous vertical temperature gradient is seen around 3 hPa. This is induced by systematic positive and negative temperature biases in the upper and middle stratosphere, respectively (see also **Fig. 11.8**), which result from problems in the radiative transfer model used in JRA-25 (*Onogi et al.*, 2007). The stratopause is located around 1 hPa both in MERRA and MERRA-2 except for polar autumn and winter, during which it is located above 1 hPa (*e.g.*, *Hitchman et al.*, 1989).

Temperature differences from MERRA are shown in **Figure 11.8**. While the latest reanalyses (*i.e.*, MERRA-2, ERA-Interim, and JRA-55) show small differences (*i.e.*, mostly < 1 K) below 10 hPa, ERA-40 and JRA-25 show large differences (*i.e.*, > 2 K) from the latest reanalyses even below 10 hPa. On the other hand, there can be seen vertically stacked structures of temperature differences with large magnitudes (*i.e.*, > 2 K) above 10 hPa even for the latest reanalyses. MERRA-2 has positive and negative temperature differences between 10 hPa and 3 hPa and between 3 hPa and 0.3 hPa, respectively, and the magnitude of these anomalies is larger in the polar regions. ERA-Interim has positive and negative temperature differences from MERRA at 10 hPa and 1 hPa, respectively, which are especially large in the SH winter extratropics. JRA-55 has negative temperature differences between 10 hPa and 1 hPa in the tropical and mid-latitude regions and positive differences in the polar regions in autumn and winter between 3 hPa and 1 hPa. In JRA-25, there are weak vertical temperature gradients above ~ 2 - 3 hPa, due to sponge layer effects, that give rise to large vertical gradients in the temperature differences between JRA-25 and MERRA (bottom row). To summarize, temperature differences (with respect to MERRA) are often substantial among older and newer reanalyses, in both the lower and upper stratosphere, and ranging from the tropics to the high latitudes.

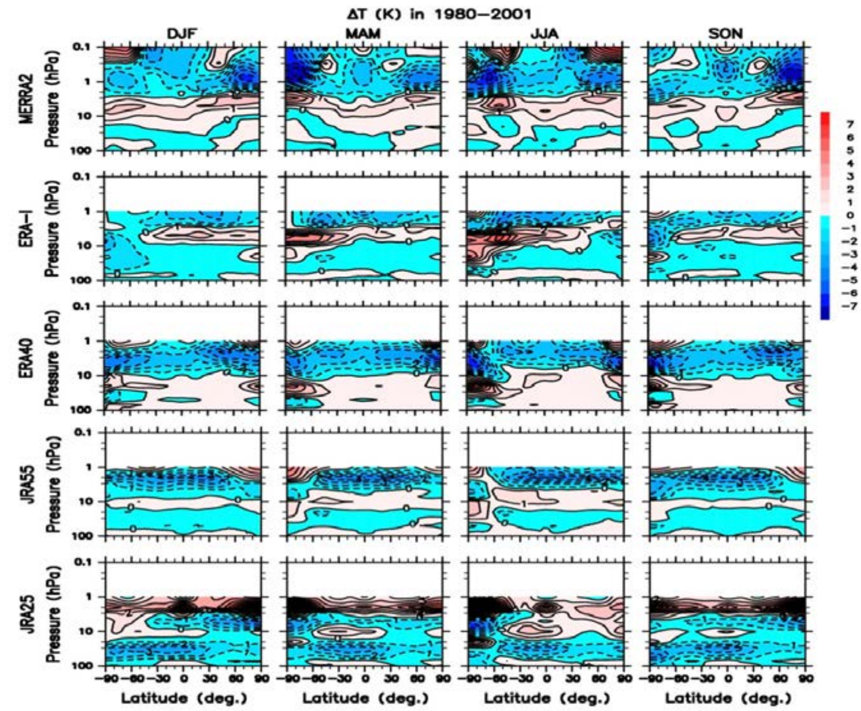


***Figure 11.8:** Same as **Figure 11.7** except for differences from MERRA. Contour intervals are 1 K.*

Figure 11.9 shows latitude-pressure sections of seasonal zonal means of zonal wind.

Westerly and easterly jets in the winter and summer stratosphere, respectively, are well reproduced in all the available reanalyses. An equatorward tilt of the polar-night westerly jet in autumn and winter is common to all the available reanalyses. Likewise, all reanalyses show a poleward-tilting summer easterly jet (except for the NH summer jet in MERRA, which tilts poleward below 1 hPa and equatorward above that level).

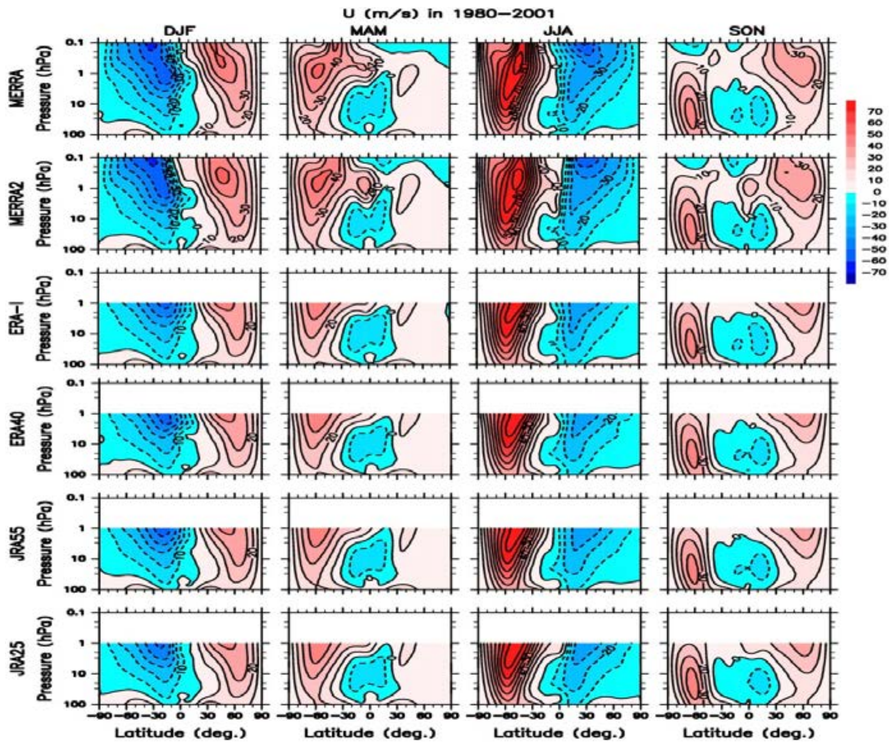


Figure 11.9: *Same as* ***Figure 11.7*** *except for zonal wind. Contour interval is 10 m s*$^{-1}$.

Zonal wind differences from MERRA are shown in **Figure 11.10**. Differences among reanalyses are largest (*i.e.*, > 4 m s^{-1}) in the tropical stratosphere (and mesosphere), which is probably due to the dearth of wind observations assimilated and because the available temperature observations do not provide a strong constraint on the wind in the deep Tropics. MERRA-2 shows a year-round positive (westerly) bias near the tropical stratopause and semi-annually oscillating differences in the tropical mesosphere compared to MERRA.

Figure 11.11 and **Figure 11.12** depict latitude-pressure sections of seasonal means of Eulerian-mean and residual-mean meridional winds, respectively. The residual-mean meridional (and vertical) winds were computed from Eulerian-mean winds and resolved wave fluxes on the primitive equation system (*cf., Andrews et al.*, 1987). Eulerian-mean meridional wind shows a strong equatorward flow in the polar-night jet regions especially in the NH, which does not appear in the residual-mean meridional flow. This feature is also seen in the SH spring (SON) because of the persistent SH polar vortex. Other features in the stratosphere hardly change between Eulerian-mean and residual-mean flows, suggesting that the divergence of the eddy heat flux of Rossby waves is small. MERRA and MERRA-2 show the Eulerian-mean equatorward flow up to 0.2 hPa in the polar-night jet region, above which sponge layer effects may obscure the upward propagation of planetary waves. Another interesting feature is that both Eulerian-mean and residual-mean meridional flows above 1 hPa in MERRA and MERRA-2 have a weak negative peak around 20 ° S in addition to a strong negative peak around

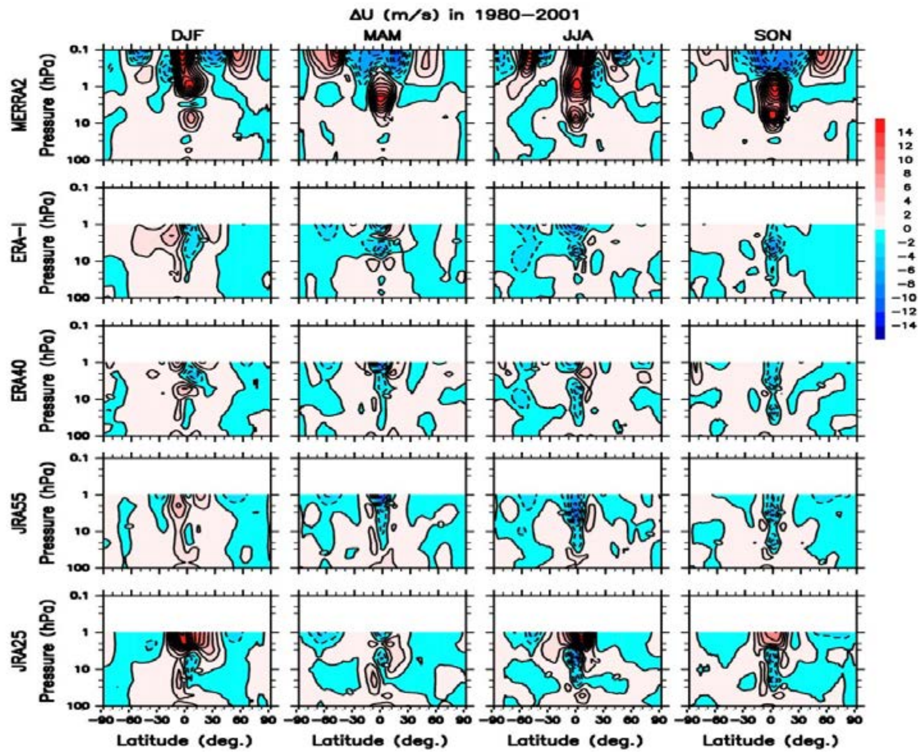


Figure 11.10: *Same as* ***Figure 11.8*** *except for zonal wind. Contour interval is 2 m s*$^{-1}$.

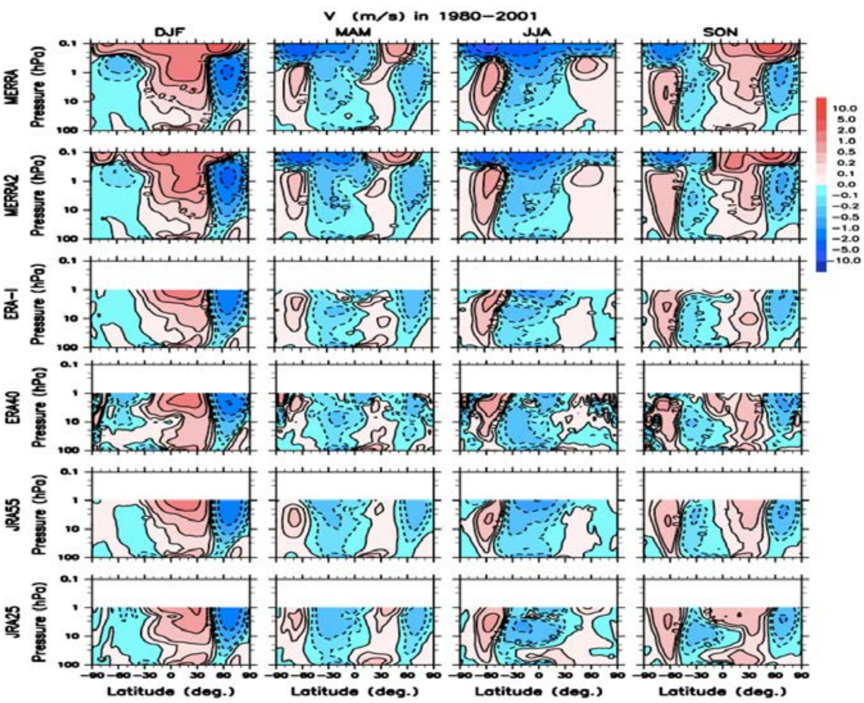


Figure 11.11: *Same as* ***Figure 11.7*** *except for meridional wind. Contour intervals are logarithmic to emphasize small speeds.*

60 ° S in MAM and JJA (*i.e.*, austral autumn and winter). Their extension down to 1 hPa is also partially seen in ERA-Interim and ERA-40, but not in JRA-55 and JRA-25 (see **Figure 11.12**). This meridional flow is likely driven by parameterized gravity wave drag in the forecast model. Gravity-wave resolving GCMs such as KANTO (*Watanabe et al.*, 2008) and KMCM (*Becker and Vadas*, 2018) do not show such separated peaks of gravity wave drag in the SH winter mesosphere. This suggests that meridional propagation of gravity waves neglected in the gravity wave drag parameterization is essential for the representation of the meridional circulation in the mesosphere (*cf.*, *Sato et al.*, 2009). It is also worth noting that both Eulerian-mean and residual-mean meridional flows in ERA-40 are noisier than those in the other reanalyses in the polar regions. This may be linked to the noisier vertical velocity in ERA-40 (*e.g.*, *Iwasaki et al.*, 2009; *Monge-Sanz et al.*, 2007; see also **Figs. 11.15**, **11.16**) through mass continuity; thus, science studies based on ERA-40 residual circulation velocities would likely generate noisier results.

Eulerian-mean and residual-mean meridional wind differences from MERRA (Reanalyses minus MERRA) are shown in **Figure 11.13** and **Figure 11.14**, respectively. ERA-Interim, ERA-40, JRA-55, and JRA-25 have negative anomalies around 1 hPa in the SH winter midlatitudes (for JRA-25) or polar regions in both Eulerian-mean and residual-mean flows, which indicates that MERRA and MERRA-2 have stronger Eulerian-mean equatorward flow induced by the parameterized gravity wave drag. It may be because non-orographic gravity wave drag dominant in the SH is implemented only for MERRA and MERRA-2 (*e.g.*, *Fujiwara et al.*, 2017). On the other hand, the wind differences in ERA-Interim and JRA-55 are positive for Eulerian-mean but negative for residual-mean around 1 hPa in the NH winter polar region. This suggests that the Stokes drift induced by resolved planetary waves is different among ERA-Interim/JRA-55 and the other reanalyses. Eulerian-mean and residual-mean meridional wind differences in JRA-25 change signs around 3 hPa in the winter polar regions, which may be a result of the anomalous vertical temperature gradient there. Note also that 1 hPa is in the sponge layer for the lower top models, so that is another factor that might contribute

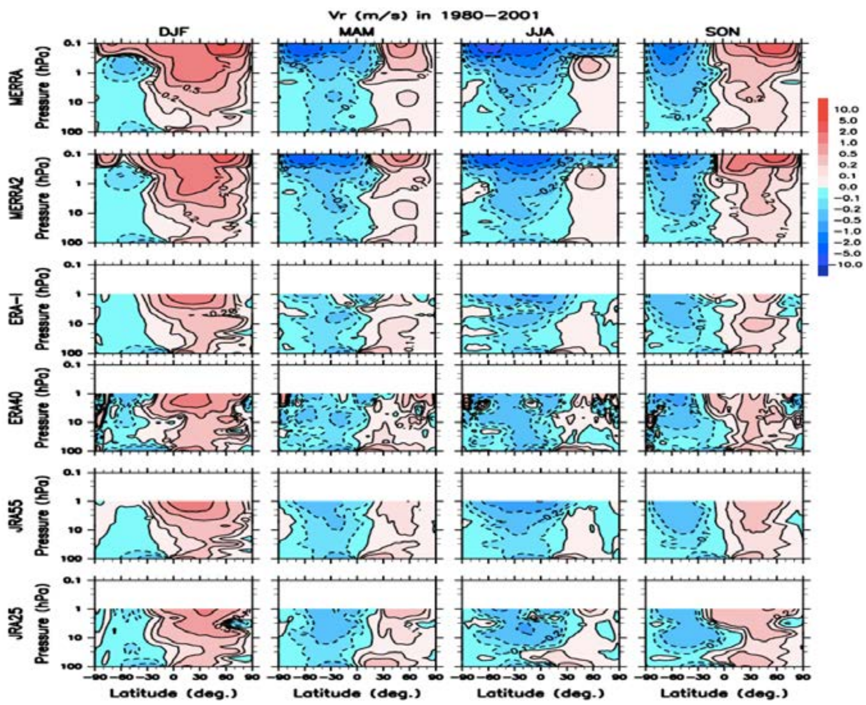


Figure 11.12: *Same as* ***Figure 11.7*** *except for the meridional component of the residual circulation. Contour intervals are logarithmic to emphasize small speeds.*

to the differences.

In the winter mesosphere, MERRA has stronger (Eulerian-mean and residual-mean) poleward flow than MERRA-2. Such a difference between MERRA and MERRA-2 could be attributable to the difference of their gravity wave drag schemes (*i.e.*, non-orographic gravity wave source and intermittency of drag (*Molod et al.*, 2015)). Keep in mind that differences in the USLM also suffer from effects of the sponge layers and these may be different between high-top (*i.e.*, MERRA and MERRA-2) and low-top models. Since enhanced diffusion in the sponge layer is expected to be induced at a lower height in low-top models than in high-top models, pseudomomentum carried by resolved waves might be weaker in low-top models, which may lead to weaker Eulerian-mean meridional flow.

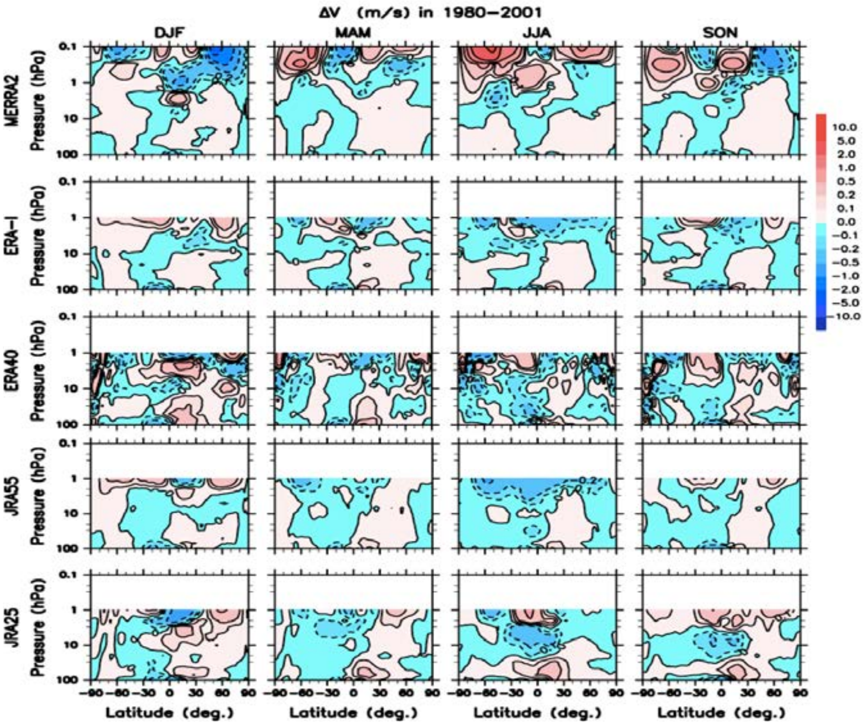


Figure 11.13: *Same as **Figure 11.8** except for meridional wind. Contour intervals are logarithmic to emphasize small differences.*

Figure 11.15 shows latitude-pressure sections of seasonal means of residual-mean vertical wind. A strong descending branch in the winter stratosphere can be seen in all the available reanalyses and is maximized around 75 °N in the NH and around 50 °S in the SH. Another weak descending branch is seen in the polar summer hemisphere below about 10 hPa, but only extends up to 1 hPa in JRA-25. A strong ascending branch in the stratosphere is maximized in the summer subtropics in all the available reanalyses. In the mesosphere from MERRA and MERRA-2, ascending and descending branches are maximized at the summer and winter poles, respectively. It should be noted that the residual-mean vertical and meridional winds in ERA-40 are much noisier than the other reanalysis especially in the polar regions. *Iwasaki et al.* (2009) reported a noisiness of residual-mean vertical wind in ERA-40 and attributed it to inconsistent dynamical noise induced by the assimilation process.

Residual-mean vertical wind differences from MERRA are shown in **Figure 11.16**. They clearly show that the differences in the winter polar stratosphere are positive, which indicates that the descending branch in the winter stratosphere is strongest in MERRA. In the mesosphere, the residual-mean vertical flow in MERRA is stronger than in MERRA-2 both at the summer and winter poles, which is consistent with the

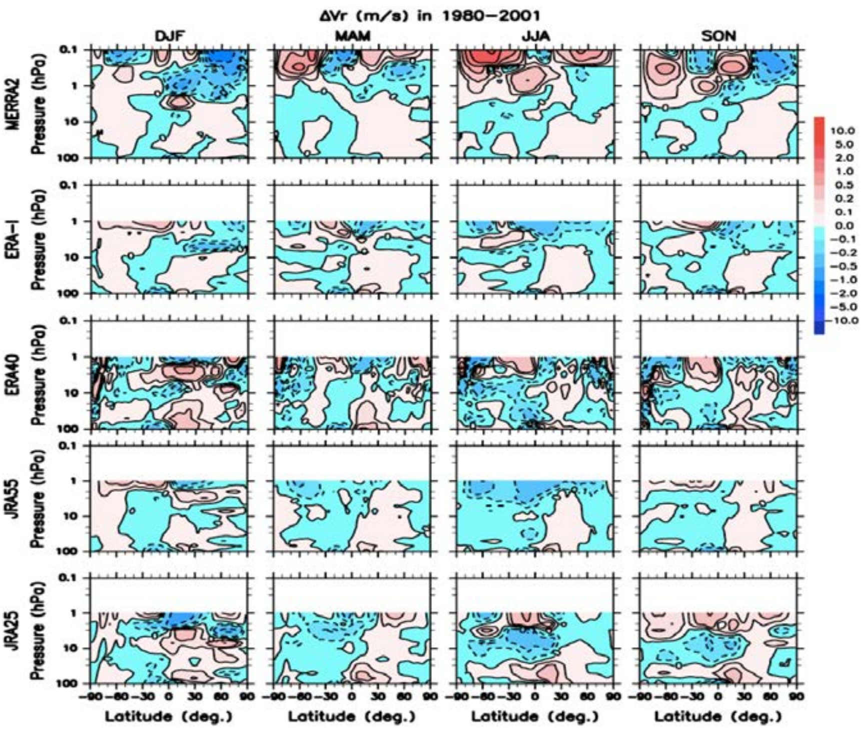


Figure 11.14: *Same as **Figure 11.8** except for the meridional component of the residual circulation. Contour intervals are logarithmic to emphasize small differences.*

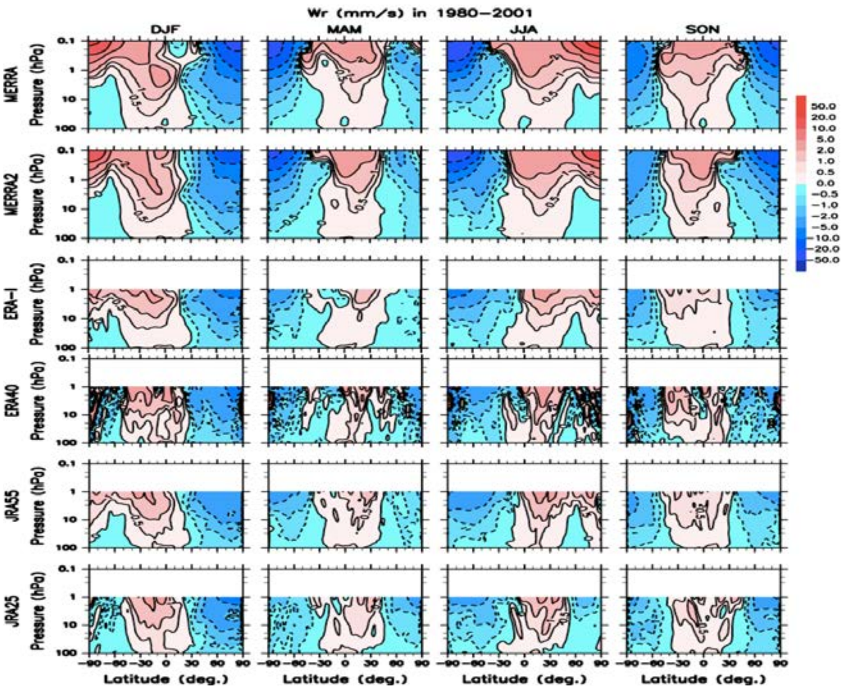


Figure 11.15: *Same as **Figure 11.7** except for the vertical component of the residual circulation. Contour intervals are logarithmic to emphasize small speeds.*

stronger summer-to-winter meridional flow in MERRA. Comparisons of residual-mean meridional and vertical winds among the reanalyses do not conclude which reanalysis gives the most realistic and reliable meridional circulation in the stratosphere and mesosphere, but do inform the choice of which reanalysis dataset to use for different science applications. Noisy meridional and vertical winds in ERA-40 can cause larger dispersion of air parcels, which leads to shorter age of air and a weaker subtropical barrier in the stratosphere (*e.g.*, *Diallo et al.*, 2012; *Schoeberl et al.*, 2003). JRA-25 showed the upward extension of the descending branch in the summer stratosphere and anomalous flow around 3 hPa in the winter stratosphere, unlike the other reanalyses. The strongest descending branch in the winter stratosphere in MERRA may give shorter ages of air, similar to ERA-40. This result is consistent with results prepared for *Chapter 5* (not shown; *Thomas Birner, personal communication*, 2021), despite weak lower stratospheric tropical upwelling in MERRA (**Fig. 5.8**). *Abalos et al.* (2015) also reported differences among MERRA, ERA-Interim, and JRA-55. Thus, it is concluded that the newer MERRA-2, ERA-Interim, and JRA-55 reanalyses should be used to study transport by the residual circulation and that the older MERRA, ERA-40, and JRA-25 reanalyses are unsuitable for this purpose.

11.2.2 Annual cycles

In this section, we show annual cycles of zonal-mean temperature, Eulerian zonal and meridional winds, and Lagrangian TEM residual circulation velocities averaged over 1980 - 2001 for each of six reanalyses and their differences from MERRA.

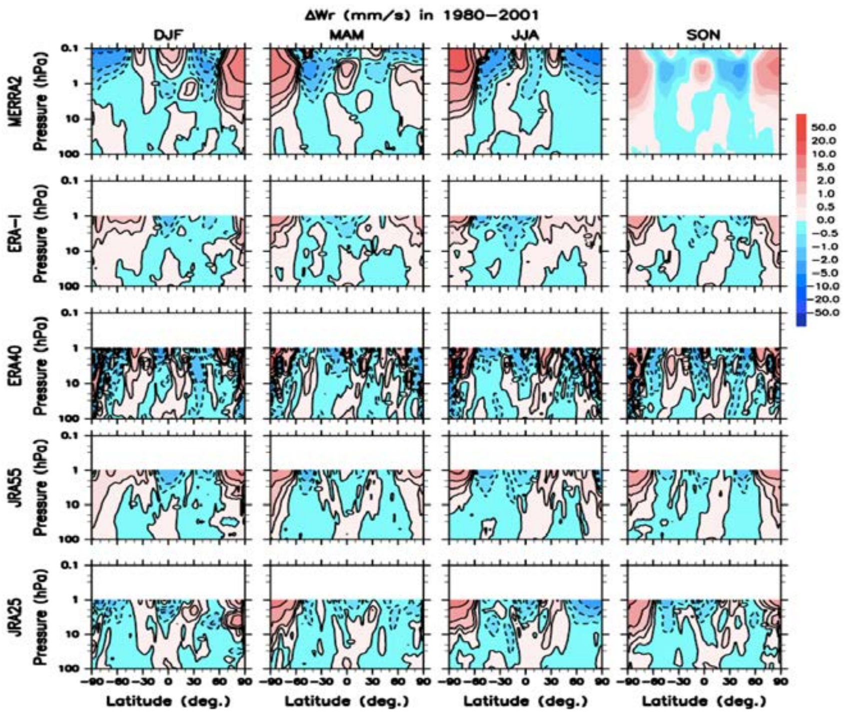


Figure 11.16: *Same as **Figure 11.8** except for the vertical component of the residual circulation. Contour intervals are logarithmic to emphasize small speeds.*

Figure 11.17 shows time-latitude sections of monthly zonal means of temperature at 1 hPa (left column) and their differences from MERRA (right column). The left column shows that the seasonal evolution of temperature near the stratopause (1 hPa) is similar among the reanalyses. While temperature is maximized in summer and minimized in the winter polar regions in both the NH and SH, its seasonal variation is smaller in the Tropics. Temperature differences from MERRA (right column) are larger in polar winter and spring and often exceed 4 K. The temperature in polar winter and spring is lower in MERRA-2, intermediate in MERRA, and higher in ERA-Interim, ERA-40, JRA-55, and JRA-25. The colder polar winter upper stratosphere in MERRA-2

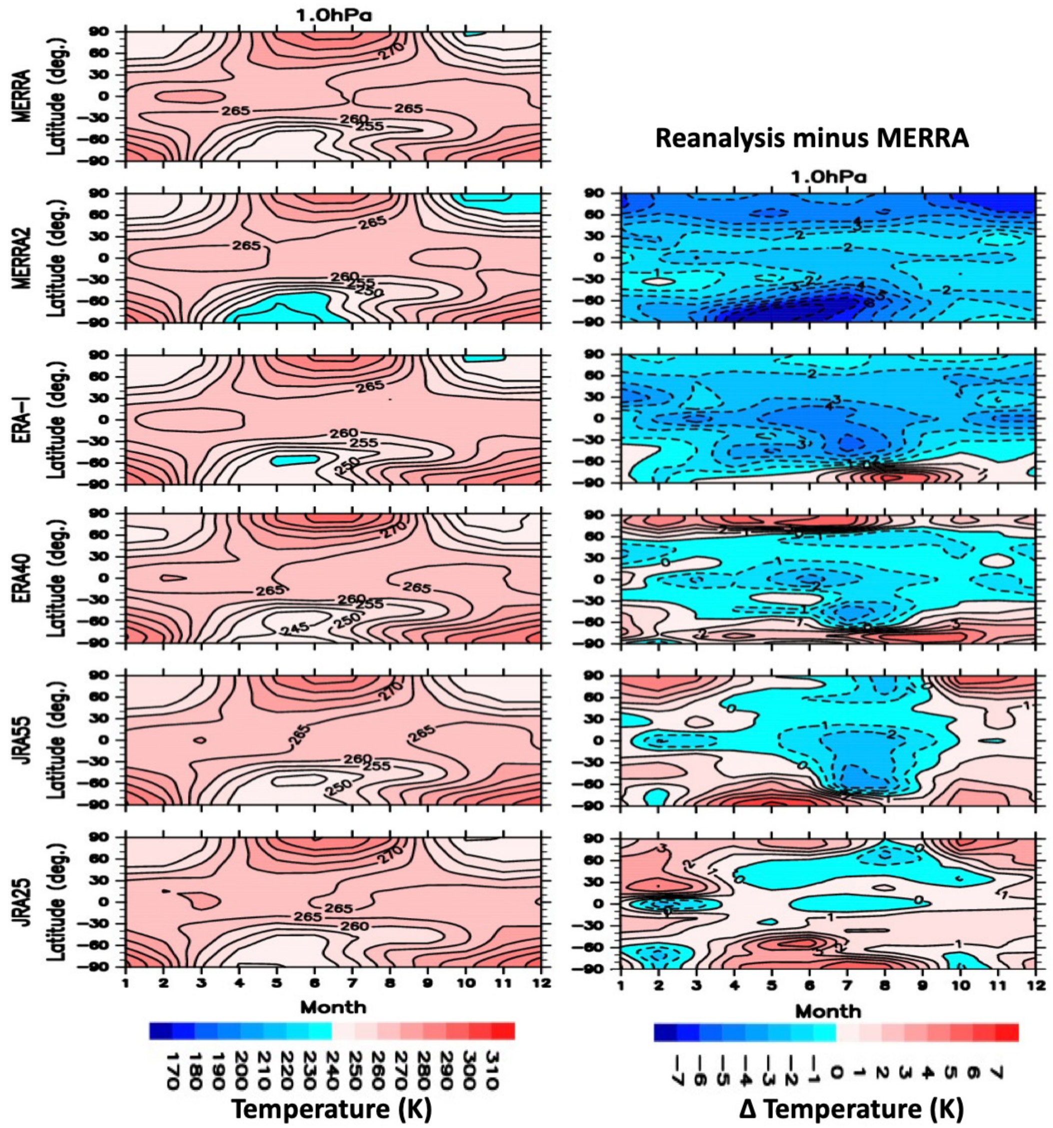


Figure 11.17: *Time-latitude sections of monthly-mean zonal-mean temperature (left column) at 1 hPa averaged over 1980 - 2001 from MERRA, MERRA-2, ERA-Interim, ERA-40, JRA-55, and JRA-25. Contour intervals are 5 K. The right column shows differences among each reanalysis on the left minus MERRA (top left). Contour interval is 1 K.*

is consistent with a slightly larger and longer-lived polar vortex, as will be shown in *Section 11.4.1*.

Figure 11.18 shows time-latitude sections of monthly zonal mean zonal wind at 1 hPa (left column) and their differences from MERRA (right column). Overall, the evolution of the summer and winter zonal jets near the stratopause is in good agreement among the reanalyses (left column). Zonal wind differences from MERRA (right column) show that the Antarctic polar-night jet at this altitude is latitudinally broader in MERRA and MERRA-2 than in the other reanalyses. This is evidenced by the two negative regions on the poleward and equatorward flanks of the jet in May, June, and July in the ERA and JRA systems (though less clear in JRA-25). Unlike temperature (which showed largest differences at high latitudes), largest zonal wind differences occur in the Tropics and can be attributed to differences in the SAO among the reanalyses.

Time-latitude sections of monthly-means of Eulerian-mean and residual-mean meridional winds at 1 hPa and their differences from MERRA are shown in **Figure 11.19** and **Figure 11.20**, respectively. Although a couplet of Eulerian-mean equatorward and poleward flow is seen at 1 hPa in NH winter in all the available reanalyses (left column of **Figure 11.19**), it is stronger in MERRA than in the other reanalyses (right column of **Figure 11.19**). This difference is mostly confined between 0° and 30°N in February and may be associated with differences in the SAO's secondary circulation. A similar feature is partially seen in SH winter, but much weaker than in the NH. On the other hand, the residual-mean meridional flow in winter is always poleward in both the NH and SH (left column of **Figure 11.20**). While the poleward residual-mean flow in NH is maximized from 0° - 30°N in December and January in all the reanalyses, the latitude and month of the strongest

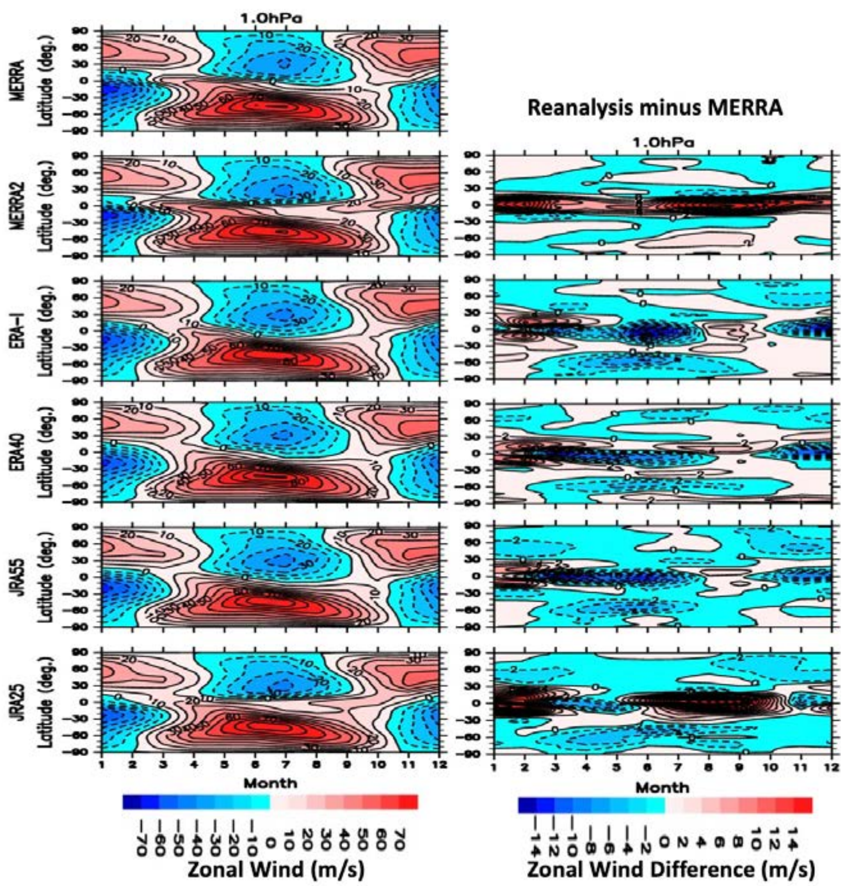


***Figure 11.18:** Same as **Figure 11.17** except for zonal wind. Contour intervals are 10 m s-1 for panels in the left column and 2 m s-1 for panels in the right column.*

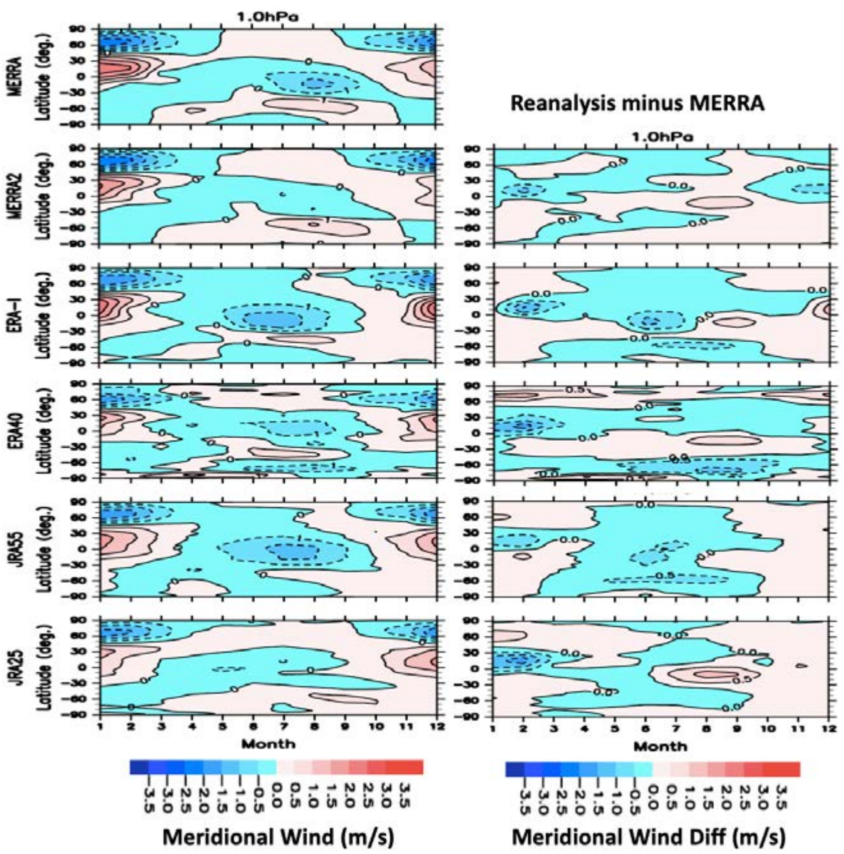


***Figure 11.19:** Same as **Figure 11.17** except for meridional wind. Contour intervals are 0.5 m s-1 for all panels.*

poleward residual-mean flow in the SH are variable among the reanalyses. Eulerian-mean and residual-mean meridional wind differences from MERRA are basically similar, so that the variable feature of residual-mean flow in SH could be attributable to the difference of parameterized gravity wave drag among the reanalyses as mentioned in *Section 11.2.1*.

Figure 11.21 shows time-latitude sections of monthly means of residual-mean vertical wind at 1 hPa (left column) and their differences from MERRA (right column). A strong descending branch is seen in the winter polar regions in all the reanalyses, but it is too noisy in ERA-40 as mentioned in *Section 11.2.1*. The descending branch in polar winter is strongest in MERRA. It indicates that the temperature differences among MERRA and the other reanalyses (except for MERRA-2) shown in **Figure 11.17** are not due to the difference of dynamical heating induced by the downward flow but more likely due to the difference in the radiation schemes and the assimilation process among the reanalyses.

Seasonal variations of residual-mean meridional wind in 30°-60°N and 30°-60°S at 1 hPa are shown in **Figure 11.22**. Poleward flow is maximized in December and January in NH and in August in SH (except for ERA-40), which is nearly coincident with the maxima of residual-mean downward flow from 60°-90°N and from 60°-90°S, respectively, at 1 hPa (see below). Seasonal variations of the residual-mean meridional wind are larger in NH than in SH because of the larger planetary wave activity in the NH winter. The residual-mean meridional flow in summer becomes equatorward both in NH and SH only for ERA-Interim and JRA-55, which was also seen in **Figure 11.20**. It looks consistent with the strongest upward residual flow in summer for ERA-Interim and JRA-55 as will be shown below. Since most of the planetary and orographic gravity waves are prohibited to propagate upward between the westerlies in the troposphere and the summer easterlies in the stratosphere because of the critical layer filtering, the differences in the meridional flow around the summer stratopause are likely due to the difference of non-orographic gravity wave drag among the reanalyses.

Seasonal variations in the residual-mean vertical velocities from 60°-90°N and from 60°-90°S at 1 hPa are presented in **Figure 11.23**. At this altitude the residual-mean downward flow in polar winter is maximized in December and January in NH and in August in SH (except for ERA-40). The different seasonal variations between the NH and SH (*i.e.*, not a 6-month shift) are consistent with their different seasonal marches of planetary wave activity (*e.g.*, *Shiotani and Hirota*, 1985). It is worth noting that seasonal marches of downward residual-mean flow at 10 hPa are delayed by 1 - 2 months compared to at the 1 hPa altitude level (not shown).

Turning specifically to the deep Tropics, the annual march of zonal-mean zonal winds and their variability are shown in **Figure 11.24**. Comparing four reanalyses to MLS gradient winds (panel **a**) show good agreement between the reanalyses and the observations time-averaged annual cycle of zonal winds from 10°S-10°N. The year-round westerly bias in MERRA-2 is apparent. While the mean annual cycle in zonal wind (panel **a**) in each reanalysis is similar the interannual variability (panel **b**) is quite different among the reanalyses. Future studies need to document and understand the cause of these year-to-year variations. In panel (**c**), we demonstrate that the standard deviation among the four reanalyses varies as a function of the time of year; it is smaller (6 - 7 m s^{-1}) during the westerly phase of the SAO around the equinoxes and larger (8 - 9 m s^{-1}) during the easterly phases in solstice seasons. Separating the tropical regions by hemisphere (10°-20°N and 10°-20°S) reveals a clear annual cycle. Namely, the variability among the reanalyses is

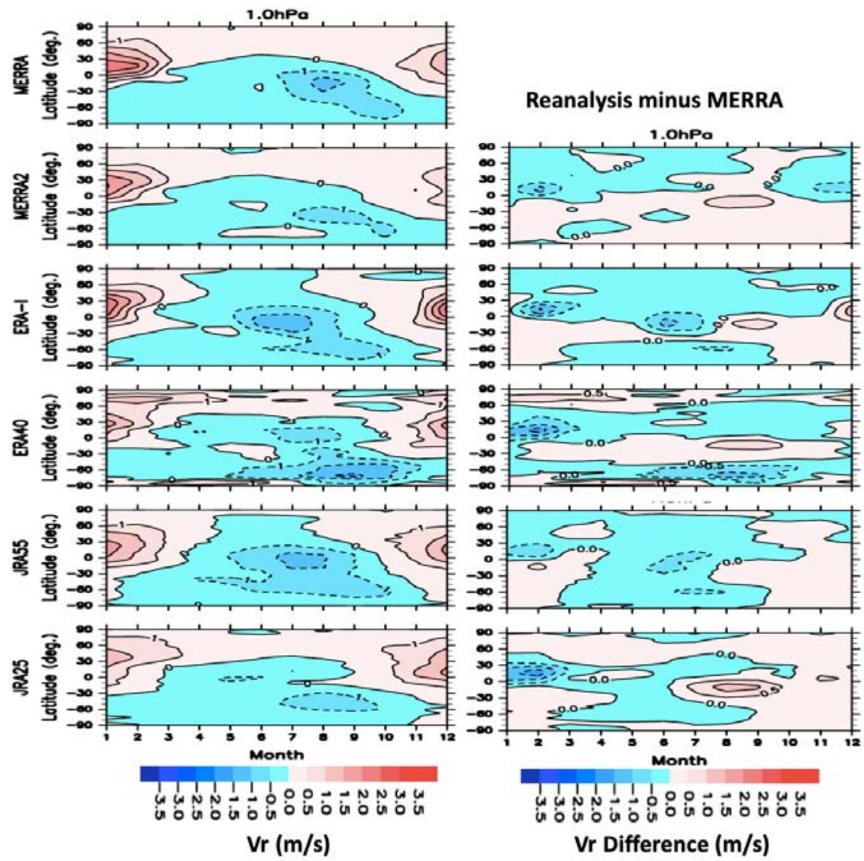


Figure 11.20: *Same as **Figure 11.17** except for the meridional component of the residual circulation. Contour intervals are 0.5 m s^{-1} for all panels.*

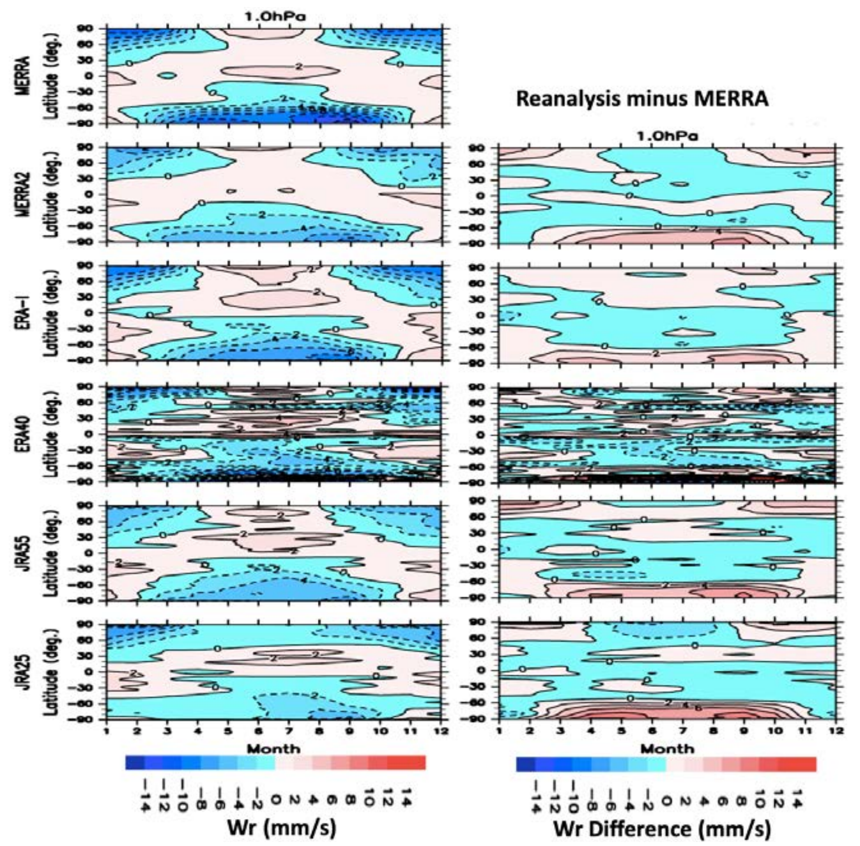


Figure 11.21: *Same as **Figure 11.17** except for the vertical component of the residual circulation. Contour intervals are 2 mm s^{-1} for all panels.*

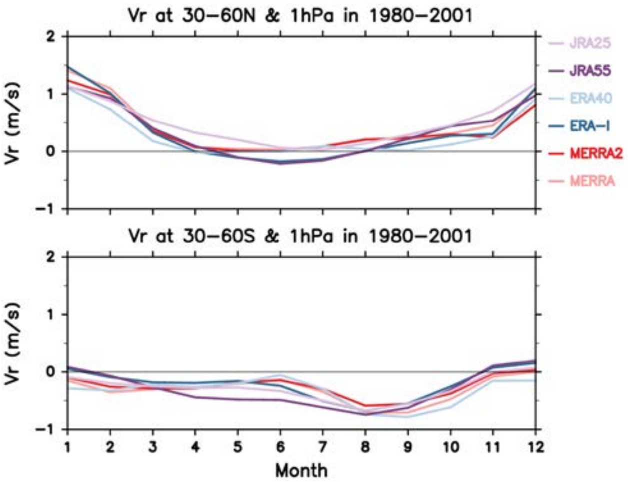


Figure 11.22: *Line plot showing the annual cycles of the meridional component of the residual circulation from 30°-60° latitude in the a) NH and b) SH at 1 hPa in 6 reanalysis datasets.*

twice as large (5 - 6 m s-1) in the winter than in summer (2 - 3 m s-1) (not shown).

Recall that temperature differences are large at all latitudes and increase with increasing altitude (shown in **Figure 11.4**). Here we explore whether those differences depend on the annual cycle. **Figure 11.25** depicts the annual cycle of zonal mean temperature at 60°N (top panels) and 60°S (bottom panels) at the stratopause (left panels) and in the lower mesosphere (right panels). There is excellent agreement among the reanalyses at 1 hPa (with the exception of CFSR which is ~10 K warmer in the winter). At this altitude the differences range from 3 - 8 K and are smallest in early winter. The right column shows that mid-latitude temperature differences grow rapidly in the lower mesosphere. At 0.3 hPa there are 10 - 15 K differences among the reanalyses year-round, with larger differences in the SH (bottom right) than in the NH (top right). It is not surprising that MERRA-2 (red contour) is in closest agreement with MLS (black contour) since the time period shown here is 2005 - 2015 when MERRA-2 assimilates MLS temperature data. Thus we conclude that reanalysis temperatures are suspect at 0.3 hPa in general, but are perhaps more believable in MERRA-2 because MERRA-2 assimilates some data at these altitudes.

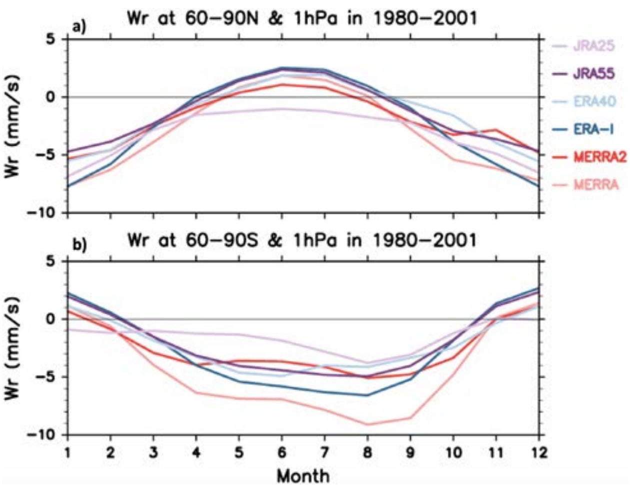


Figure 11.23: *Line plot showing the annual cycles of the vertical component of the residual circulation from 60°-90° latitude in the a) NH and b) SH at 1 hPa in 6 reanalysis datasets.*

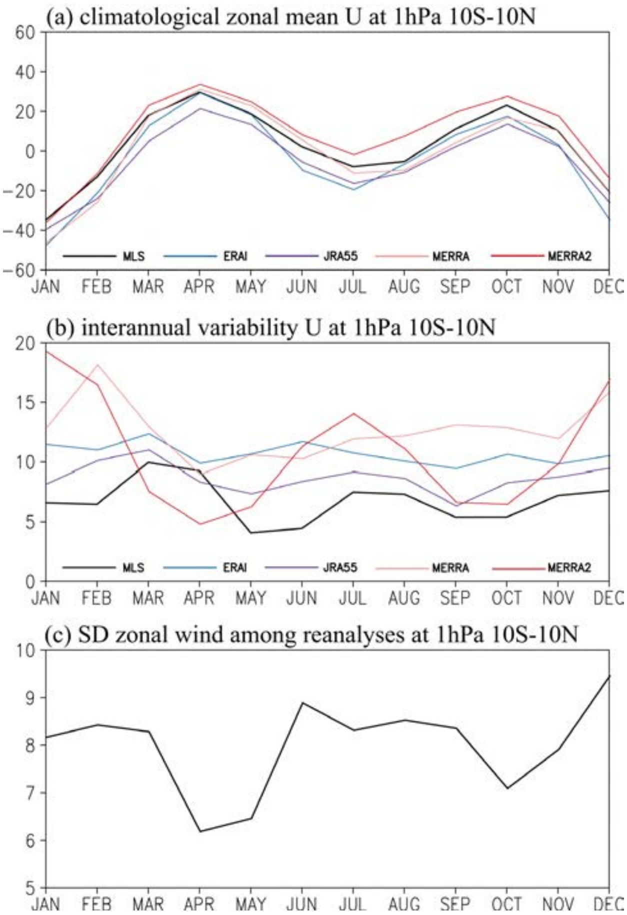


Figure 11.24: *Line plot of the 10°S - 10°N climatological annual marches of (a) zonal-mean zonal winds at 1 hPa for 1980 - 2010 for five reanalyses (color) and for the September 2004 to August 2014 average MLS gradient winds (black), (b) zonal wind interannual variability within each reanalysis dataset, and (c) zonal wind standard deviations showing variability among the five reanalyses. Wind is in m s-1. Figures 11.24a and 11.24b are modified from Kawatani et al. (2020).*

11.2.3 Long-term variability

The reanalyses also exhibit long-term variability due to different climate forcing mechanisms. A multi-linear regression (MLR) analysis has been performed to characterize USLM variability associated with the ENSO, the QBO, the 11-year solar cycle, and volcanic eruptions (while taking into account any trend associated with GHG changes – see *Crooks and Gray*, 2005). Results from the following datasets are shown and discussed

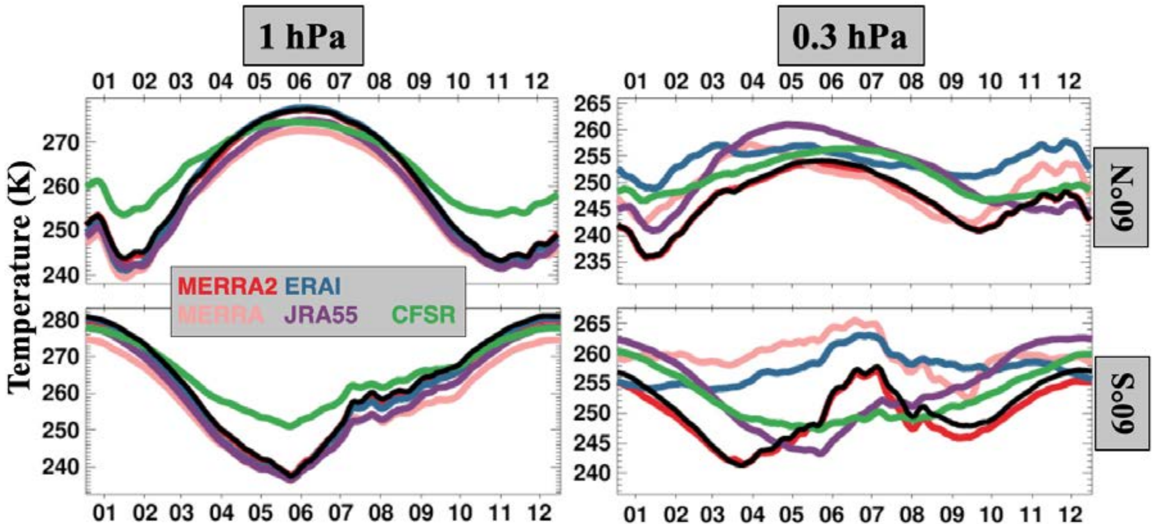


Figure 11.25: *Line plots of multi-year (2005 - 2015) annual cycles of temperature at 60° N (top) and 60° S (bottom) at 1 hPa (left) and 0.3 hPa (right) in five reanalyses (MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR) and MLS (black).*

here: JRA-55, MERRA-2 and ERA-Interim, since these are the most up-to-date reanalysis products with available data on several pressure levels above 1 hPa. Further comparisons including earlier reanalysis products and those that do not fully resolve the USLM can be found in *Mitchell et al.* (2015).

MLR Methodology and Indices

The ENSO index employed was the Nino3.4 time-series (5° S - 5° N; 120° W - 170° W) from the Extended Reconstructed Sea Surface Temperature (ERSST) dataset (*Smith and Reynolds*, 2003; **http://www.cpc.ncep.noaa.gov/data/indices/**). The volcanic eruption index was derived from the *Sato et al.* (1993) aerosol index. The 11-year solar cycle index was derived from an updated version of the Naval Research Laboratory model for Solar Spectral Irradiance (NRLSSI) time-series of total solar irradiance (*Wang et al.,* 2005) available at **http://solarisheppa.geomar.de/solarisheppa/cmip5**. QBO variability was expressed by a combination of two separate indices, comprised of the principal components of the 1st two terms from an Empirical Orthogonal Function (EOF) analysis of tropical winds averaged over the region 5° S - 5° N and 100 - 10 hPa (for more details see *Chapter 9*), in order to capture the time-variation at different heights associated with the gradually descending QBO phase. An autoregressive noise term was included (see *Crooks and Gray*, 2005) and a Student's t-test was employed to determine the probability that the regression coefficients are significantly different from the noise (light/dark shaded regions in all figures denote statistical significance at the 95 % / 99 % level). The regression coefficients have been rescaled in all figures to show the typical maximum response *e.g.*, between opposite QBO/ENSO phases or between periods of solar maximum-minimum conditions. For further details see *Mitchell et al.* (2015).

Temperature and Zonal Wind Variability in the USLM

Figure 11.26 and Figure **11.27** show variability in annual-mean, zonally-averaged temperatures and zonal winds associated with ENSO, volcanic eruptions, QBO and 11-year solar cycle covering the period 1979-2009 (1980 - 2009 in the case of MERRA-2). Responses at high latitudes arise primarily from the winter season in each hemisphere (particularly in the case of the zonal winds) and can therefore be interpreted as the winter response.

As expected, the impact of ENSO on tropical tropospheric temperatures and winds (top row) is clearly evident and highly statistically significant. There is additionally a highly statistically significant influence of ENSO in the USLM temperatures, particularly in the mid-to-high latitudes of the NH. Anomalous warming of up to 5 K is present in the NH polar US peaking around 10 hPa with cooling of up to 5 K in the LM above ~ 1 hPa, together with a corresponding (but less significant) easterly zonal wind anomaly in both US and LM, indicating a more disturbed winter circulation. These signals are consistent among the reanalyses, although there are some variations in latitudinal extent and amplitude of the LM temperature response among the datasets. These results support previous observational and modeling studies (*e.g.*, *Garcia-Herrera et al.*, 2006) that suggest the presence of increased wave forcing from the troposphere and hence a more disturbed stratosphere/mesosphere winter circulation associated with warm ENSO events.

The zonal wind and temperature QBO anomalies (3rd row; only variability associated with one of the EOFs is shown for brevity – see also *Chapter 9*) extend upward into the tropical USLM with the familiar pancake-like structure in the vertical (*Pascoe et al.*, 2005). The equatorial QBO temperature anomalies reach 3 - 4 K near ~ 3 hPa and the USLM zonal wind anomalies, while not

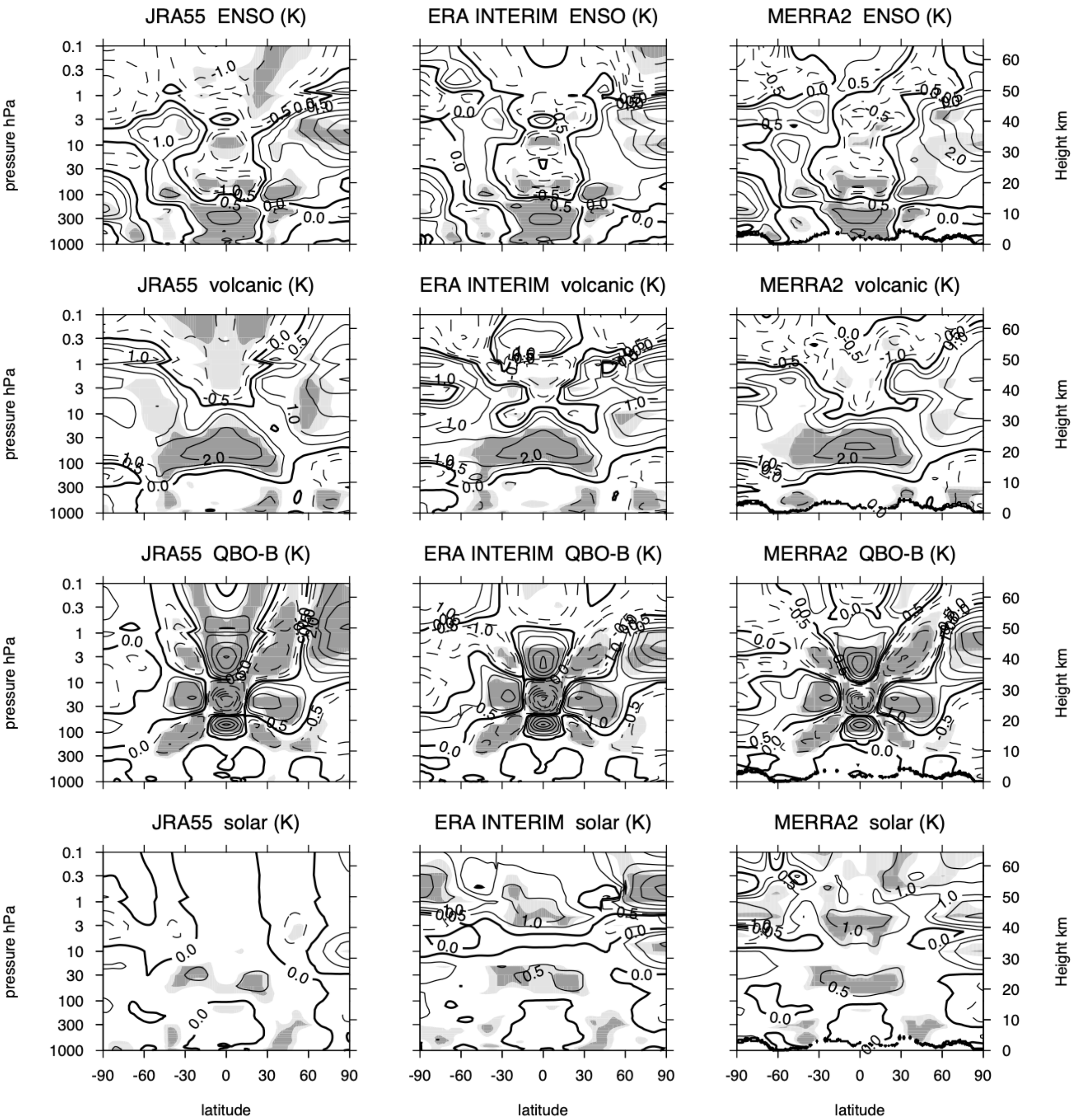


Figure 11.26: *The annual-mean temperature variability (K) associated with ENSO (top row), volcanic (2nd row), QBO (3rd row) and 11-year solar cycle (bottom row) for each of the 3 reanalysis datasets JRA-55, ERA-Interim and MERRA-2 from a multiple linear regression analysis. The regression coefficients have been multiplied to show the maximum temperature difference e.g., El Niño minus La Niña, QBO west minus east phase and Smax minus Smin. Contour intervals are 0.5, 1, 2, 3, 4, 5 K; the thick solid line denotes the zero contour, solid (dashed) contours denote positive (negative) responses. Light and dark gray shading indicates statistical significance at the 95 % and 99 % levels, respectively. Taken from Mitchell et al. (2015).*

as strong as in the lower stratosphere nevertheless exceed 5 m s^{-1}. The temperature anomalies of opposite sign in the subtropics associated with local QBO-induced secondary circulations extend to around 60 °N. While a mesospheric equatorial QBO has previously been identified (see *e.g.*, *Baldwin et al.*, 2001) it is not well characterized due to lack of observations and the reanalyses reflect this uncertainty, showing considerable variation in the structure, sign and statistical significance of the QBO signal in that region. At high latitudes there is a statistically significant response with warmer temperatures above ~ 3 hPa overlying cool anomalies and a stronger (westerly) zonal wind anomaly associated with QBO westerlies at 30 hPa, in good agreement with the Holton-Tan relationship (*Anstey and Shepherd*, 2014; *Holton and Tan*, 1982). Further discussion of the QBO response is provided in *Chapter 9*.

The 11-year solar cycle signal (bottom row) shows considerable differences among the reanalysis datasets. The primary radiative response to the 11-year solar cycle is in the mid-upper equatorial stratosphere, associated with both increased UV irradiance and ozone production (*Gray et al.*, 2010). Both ERA-Interim and MERRA-2 show a statistically significant warm anomaly of ~ 1.25 K from solar maximum to solar minimum (Smax-Smin) centered

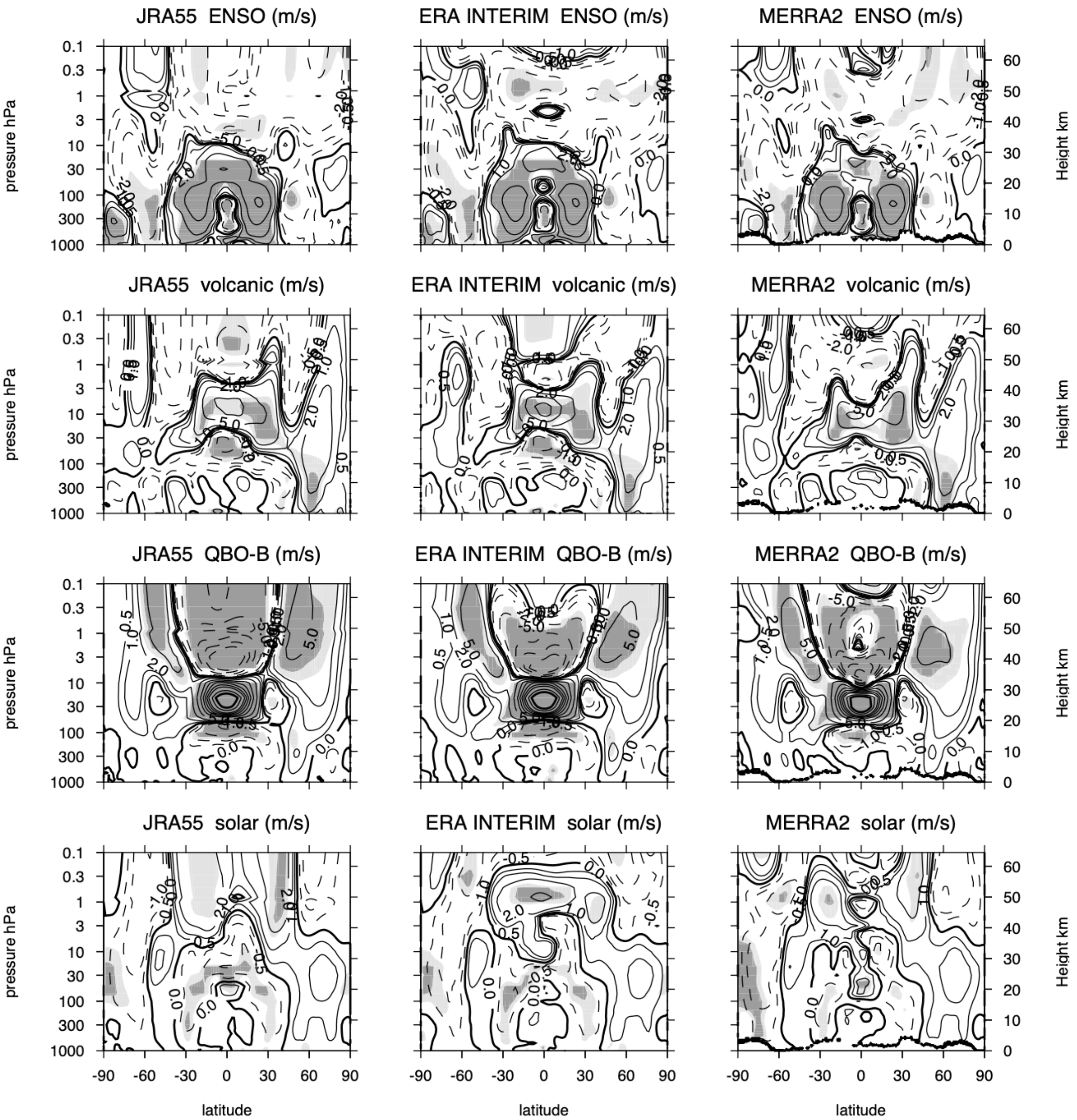


Figure 11.27: *The annual-mean zonal wind variability (ms-1) associated with ENSO (top row), volcanic (2nd row), QBO (3rd row) and 11-year solar cycle (bottom row) for each of the 3 reanalysis datasets JRA-55, ERA-Interim and MERRA-2 from a multiple linear regression analysis. The regression coefficients have been multiplied to show the maximum temperature difference e.g., El Niño minus La Niña, QBO west minus east phase and Smax minus Smin. Contour intervals are 1, 2, 3, 5, 10, 15, 20, 30, 40, 50 m s^{-1}; the thick solid line denotes the zero contour; solid (dashed) contours denote positive (negative) responses. Light and dark gray shading indicates statistical significance at the 95 % and 99 % levels, respectively. Taken from Mitchell et al. (2015).*

over the equatorial region near 1 hPa with a corresponding westerly (thermal) wind anomaly mainly in the subtropics. JRA-55 on the other hand shows a very weak temperature response with no statistical significance although the wind response agrees well with the other two reanalyses. This inconsistency in the USLM solar response among the reanalysis datasets is likely due to a combination of poor vertical resolution of the available satellite data and the difficulties of extracting an 11-year signal from datasets that assimilate observations from relatively short-lived satellite instruments. This lack of agreement is in contrast to the lower stratospheric solar response which is much more consistent between the reanalyses, presumably because this region is covered extensively by radiosonde observations. Further investigation of the weak USLM solar cycle response in JRA-55 (*Stergios Misios, personal communication*, 2018) indicates that this is primarily because of a difference in timing of the peak solar response in JRA-55: the maximum response (which is statistically significant) occurs a year or so leading up to solar maximum and is therefore not captured in **Figure 11.26** since no lag/lead has been applied to the regression analysis. This highlights possible differences in the treatment of solar irradiance changes in the underlying reanalysis model and also the importance of careful examination of lead/lag responses when performing regression analysis of the solar signal; results may also be sensitive to the solar index employed in the analysis (*e.g.*, total solar irradiance, sunspot number, solar magnetic flux) since the different solar flux proxies show some variation in their timing.

11.3 Tropical dynamics

11.3.1 Semi-annual oscillation

The SAO is a reversal of zonal winds from easterly to westerly at both the stratopause and at the mesopause (with the phase reversed) with a period of approximately six months (*Garcia et al.*, 1997). First discovered by Reed in 1962, the SAO is driven by a combination of planetary and gravity wave forcing as well as mean meridional advection (*Hamilton*, 1998), and exerts control over thermal and chemical transport processes at both the lower and upper regions of the USLM. It is crucial to examine this and other tropical processes in this chapter because differences in the reanalyses are largest in the tropical USLM. (See *Kawatani et al.* (2020) for the detailed reanalysis comparisons.)

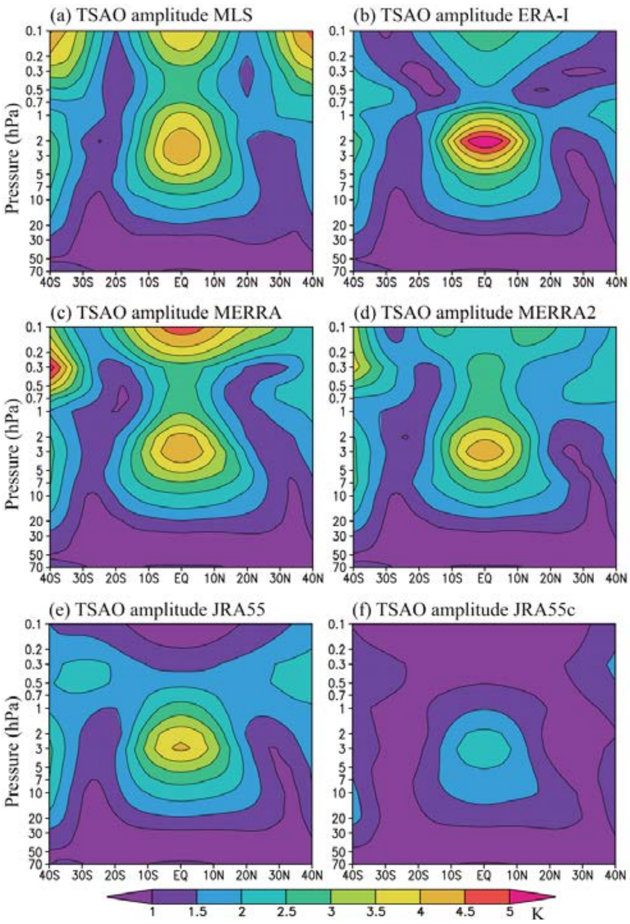


Figure 11.28: *Latitude-altitude distributions of the zonal mean temperature SAO component averaged over 2004 - 2014 for MLS observations (upper left) and 1980 - 2012 for the reanalyses. Temperature is in degrees K. Note that JRA-55C is a reanalysis without assimilated satellite observations. Modified from Kawatani et al. (2020).*

Semiannual components are extracted from monthly mean data by applying an Ormsby time filter and then calculating the amplitude ($\sqrt{2}\times\sqrt{\sum U_{SAO}^2/N}$, where USAO indicates SAO components of the zonal wind). **Figure 11.28** shows that the SAO in temperature is almost symmetric with respect to the Equator, and that the amplitude of the temperature SAO is larger in ERA-Interim compared to the other reanalyses at the tropical stratopause. In general, the structure of the SAO below 1 hPa is similar in the reanalyses, while differences among the reanalyses become large above that level. The SAO components in the mesosphere are very small in JRA-55 and MERRA-2. In order to see the effects of satellite observations, results from JRA-55C (in which only conventional data were assimilated; panel **f**) are shown alongside JRA-55 (panel **e**). The SAO amplitude in JRA-55C is severely underestimated both in the stratosphere and mesosphere. These results make clear that the physical parameterizations in the JRA-55 model apparently cannot simulate an SAO on their own.

Turning to the SAO in winds, **Figure 11.29** shows that the zonal-wind SAO is asymmetric with respect to the Equator; the maximum amplitude exists from 10°-20°S,

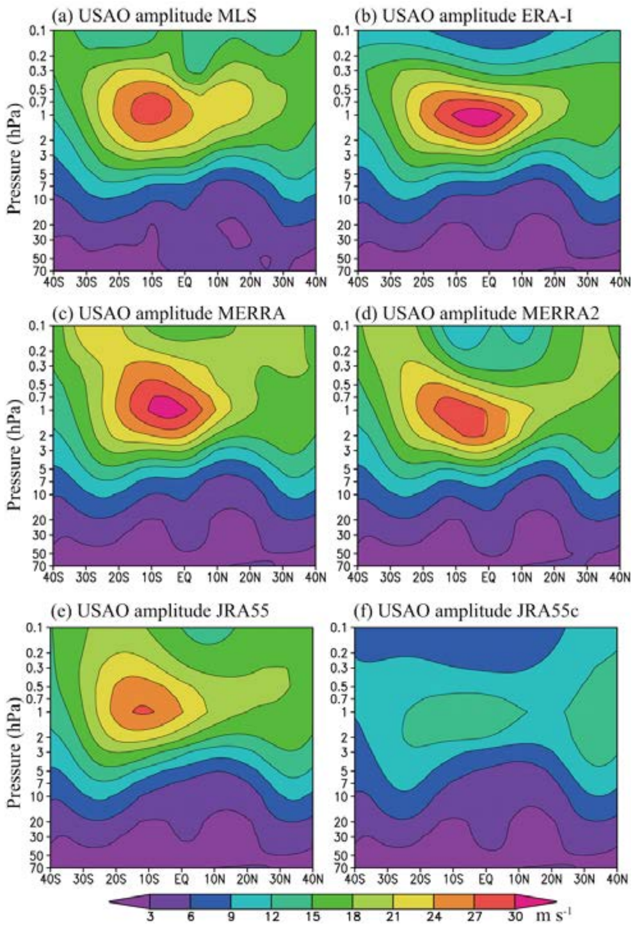


Figure 11.29: *Same as* ***Figure 11.28*** *but for the zonal mean zonal-wind SAO component in m s*$^{-1}$*. Modified from Kawatani et al. (2020).*

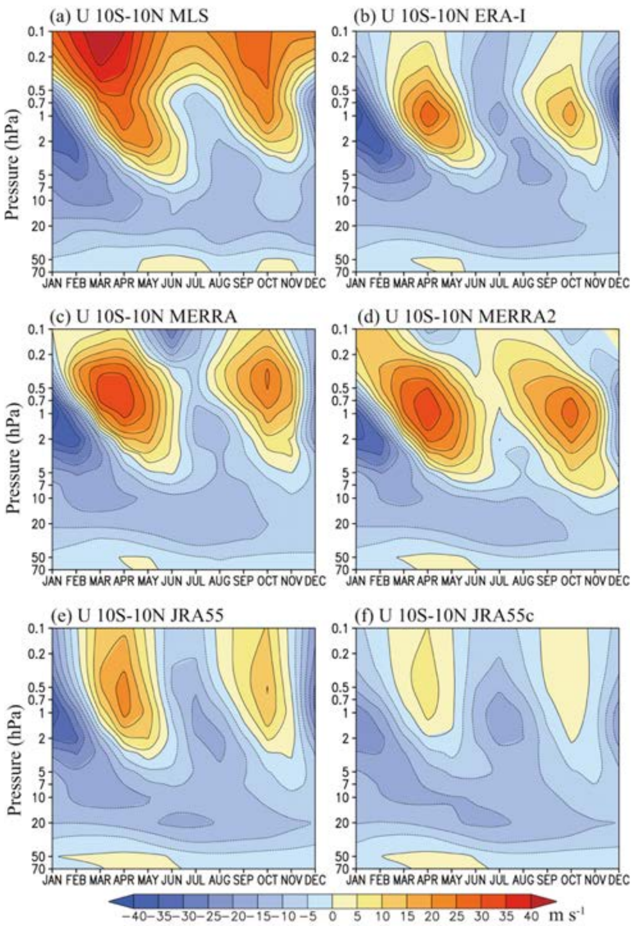


***Figure 11.30:** Time-height sections of climatological zonal-mean zonal winds (in m s^{-1}) averaged between 10 °S and 10°N for MLS derived gradient winds (top left) and five reanalyses. Modified from Kawatani et al. (2020).*

which is consistent with earlier rocketsonde observations (*e.g.*, *Hopkins*, 1975). The above asymmetry is likely due to asymmetric components in the temperature SAO. The amplitude of the zonal-wind SAO is larger in ERA-Interim and MERRA compared to MERRA-2, the JRA model versions, and the MLS observations. These results are consistent with *Das et al.* (2016) (see their **Figure 7d**) who showed the amplitude of the SAO to be larger in ERA-Interim and MERRA compared to rocket observations. Similarly, *Kishore Kumar et al.* (2015) (see their **Figure 4**) reported 30 % larger SAO amplitudes near the stratopause in MERRA compared to rocketsonde winds. These results suggest that the JRA-55 model requires upper-air data assimilation to capture the SAO. Another possible factor for the weak winds in JRA-55C is the sponge layer; the forecast model might generate winds even if there were no observations.

Time-height sections of the zonal-mean zonal wind in the deep Tropics (**Figure 11.30**) reveal large differences between the reanalyses and the observations in the mesosphere. MLS-derived gradient winds (upper left panel) are strong westerly year-round above 0.5 hPa. This wind regime is consistent with the rocketsonde climatology of tropical zonal winds at Thumba (8.5 °N, 77 °E) in the lower mesosphere shown by *Kishore Kumar et al.* (2015; see their **Figure 1**, upper left panel). Persistent westerlies are also consistent with SABER observations (*Smith et al.*, 2017; see their **Figure 1**). However, the zonal wind climatology based on the Horizontal Wind Model 07 (also shown by *Kishore Kumar et al.,* 2015; see their **Figure 1**, lower left panel) indicates the presence of summer easterlies in the tropical lower mesosphere at Thumba. The overall features in the reanalyses are similar to the MLS and rocketsonde observations between 5 hPa and 0.5 hPa but diverge at higher altitudes. In ERA-Interim and JRA-55 above 0.5 hPa the onset of the westerlies occurs at the same time instead of progressing downward as in MLS, MERRA, and MERRA-2. This is probably due to the influence of sponge layers in the models, where equatorial waves cannot propagate upward. In MERRA-2, the downward progression of zonal wind anomalies is more pronounced such that the easterlies in the lower mesosphere occur during the equinoxes instead of during the solstices. As expected, differences between JRA-55 and JRA-55C are relatively small up to ~ 10 hPa, while they become large above this level. The JRA-55C results make clear the need to assimilate satellite data in the equatorial USLM where relatively small-scale gravity waves and Kelvin waves are dominant. Users are advised that JRA-55C cannot be used in the USLM.

New results from the ERA5 also shed light on the performance of different reanalyses with regard to the SAO. As seen in **Figure 11.31**, while the zonal winds are in excellent agreement between ERA-Interim and ERA5 in the QBO altitude regime, the mesopause SAO in ERA5 is substantially different from the same feature in ERA-Interim (see *Shepherd et al.* (2018) and references therein). In the region from 0.5 hPa to 0.1 hPa, ERA5 westerlies are at least 30 m s^{-1} larger than in ERA-Interim; ERA5 lacks descending solstitial easterlies in this region as well. The strong and persistent mesospheric jet in ERA5 is evident in both the

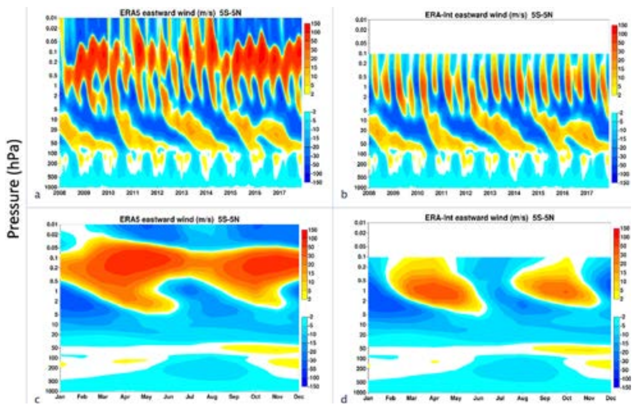


***Figure 11.31:** Time-altitude section of monthly mean zonal mean zonal wind averaged from 5 ° S - 5 ° N in ERA5 (left) and ERA-Interim (right). Top panels show interannual variability from 2008 to 2017. Bottom panels show the average annual cycle averaged between 2008 and 2017. Units are m s^{-1}. The vertical coordinate is the reference pressure of the model levels. Taken from* **https://confluence.ecmwf.int/display/CKB/ERA5%3A+The+QBO+and+SAO**. *©Copernicus Climate Change Service/ECMWF. Used with permission.*

average annual cycle of the zonal winds (bottom panels) and in individual years (top panels). Researchers at ECMWF note that the "predominance of westerlies in ERA5 is related to the spurious equatorial mesospheric jet that occurs in CY41R2 of the IFS and which peaks in the transition seasons" (*Shepherd et al.*, 2018; *Polichtchouk et al.*, 2017). While this behavior diverges from the other reanalyses presented here, it is consistent with the year-round westerly winds observed by rocketsondes and derived using satellite temperatures. ERA5 validation efforts should accompany the use of this (and any) reanalysis dataset with regard to tropical mesospheric dynamics.

11.3.2 Middle-atmosphere Hadley circulation

The Hadley cell circulation, most often associated with tropospheric dynamics, extends into the USLM. By way of explanation, a brief synopsis of the theory explaining the middle-atmosphere Hadley circulation is given below.

The residual-mean meridional circulation ($\equiv(0, \bar{v}^*, \bar{w}^*)$) satisfies the following zonal momentum equation in the transformed Eulerian-mean formalism:

$$\bar{u}_t + \bar{v}^*\left[(a\cos\phi)^{-1}(\bar{u}\cos\phi)_\phi - f\right] + \bar{w}^*\bar{u}_z = (\rho_0 a\cos\phi)^{-1}\boldsymbol{\nabla}\cdot\boldsymbol{F} + \bar{X} \quad (11.1),$$

where a is the Earth's radius, ϕ the latitude, f the Coriolis parameter, $\bar{X}$ the zonal-mean unresolved mechanical forcing, and $\boldsymbol{\nabla}\cdot\boldsymbol{F}$ the Eliassen-Palm (E-P) flux divergence (*cf.*, *Andrews et al.*, 1987). Its alternative form can be expressed using the zonal-mean absolute angular momentum ($M \equiv a\cos\phi(\bar{u} + \Omega a\cos\phi)$, where Ω is the Earth's rotation rate) as

$$M_t + \bar{v}^* a^{-1} M_\phi + \bar{w}^* M_z = {\rho_0}^{-1}\boldsymbol{\nabla}\cdot\boldsymbol{F} + \bar{X} a\cos\phi \quad (11.2).$$

This equation indicates that the residual mean flow cannot cross isopleths of M without non-zero E-P flux divergence under the assumption of $M_t = 0$ and $\bar{X} = 0$. This situation occurs in the extratropics for the monthly-mean (or longer timescales) residual circulation (which is driven by the E-P flux divergence due primarily to Rossby waves and explicitly resolved gravity waves). On the other hand, the steady state assumption does not hold in the Tropics because it takes a long time to achieve thermal wind balance due to the small Coriolis parameter there. This allows a meridional Hadley-type circulation to be thermally driven in the Tropics if there is an imbalance between the radiative equilibrium temperatures and the zonal wind distributions This circulation cancels the meridional gradients in ozone heating that form across the Equator in the USLM and those gradients are particularly large during the solstice seasons (*Semeniuk and Shepherd*, 2001a; 2001b; *Dunkerton*, 1989). Such a thermally-driven meridional circulation is referred to here as the middle-atmosphere or stratopause Hadley circulation.

The upwelling branch of the residual-mean meridional circulation in the tropical upper stratosphere and lower mesosphere is maximized in the summer subtropics (*e.g.*, *Eluszkiewicz et al.*, 1996). Planetary wave forcing in the winter extratropics can affect the summer subtropics (across the Equator). However, because the meridional gradient of absolute angular momentum in this region is small enough to neglect its advection (*Dunkerton*, 1989) it is unlikely that PW forcing explains latitudinal distributions and seasonal variations of upwelling in the tropical USLM. Instead, it is hypothesized that the existence of the middle-atmosphere Hadley circulation is required to explain these features (*Plumb and Eluszkiewicz*, 1999). It is also believed that the middle-atmosphere Hadley circulation plays a role in driving the easterly phase of the SAO through the absolute angular momentum transport because of strong nonlinearity around the tropical stratopause (*Dunkerton*, 1991). Although these features of the middle-atmosphere

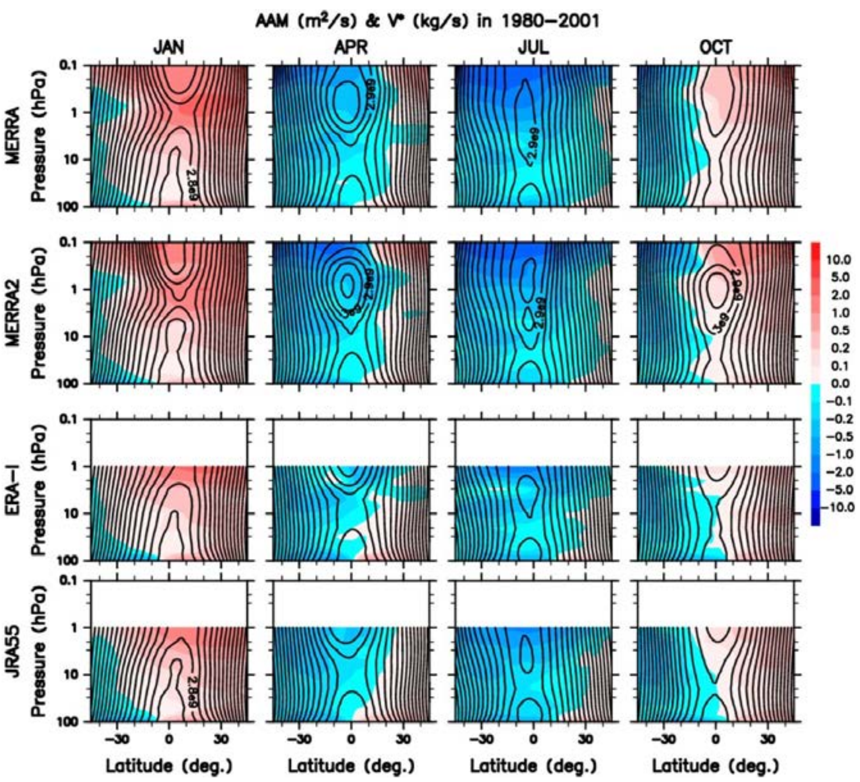


Figure 11.32: *Latitude-pressure sections of monthly- and zonal-mean absolute angular momentum (contours) and residual-mean meridional flow (colors) in (left to right) January, April, July, and October averaged over 1980-2001 from MERRA (top row), MERRA-2 (2nd row), ERA-Interim (3rd row), and JRA-55 (bottom row). Contour intervals are $10^8 m^2 s^{-1}$.*

Hadley circulation have been examined in some general circulation models (*e.g.*, *Semeniuk and Shepherd*, 2001a; 2001b), they have not yet been confirmed by observations and reanalyses. Upward extension and improved accuracy of the latest reanalysis data will facilitate quantitative examination of the middle-atmosphere Hadley circulation (*e.g.*, *Sato and Hirano*, 2019) and its relationship with the SAO.

Next we show annual cycles of the residual-mean meridional circulation in the Tropics averaged over 1980 - 2001 for each of four reanalyses (*i.e.*, MERRA, MERRA-2, ERA-Interim, and JRA-55). Since it is difficult to distinguish wave-driven and thermally-driven meridional circulation in the Tropics, we just show basic features of the meridional circulation in the tropical USLM.

Figure 11.32 shows latitude-pressure sections of monthly zonal means of absolute angular momentum (*M*) and residual-mean meridional wind in January, April, July, and October. Four reanalyses show good agreement in the spatial structure of residual-mean meridional flow. The cross-equatorial flow from the summer to the winter pole, especially in January, tends to be maximized slightly above the trough in *M*, where the *M* isopleths become horizontal. This is probably because air parcels can move meridionally without any wave forcing in the trough of *M* (*Tung and Kinnersley*, 2001; *Dunkerton*, 1989; *Hitchman and Leovy*, 1986). The trough of *M* during the solstice seasons corresponds to the easterly phase of SAO.

In order to show a relationship between the residual-mean meridional flow and the SAO, seasonal variations in the residual-mean meridional wind and zonal-mean zonal wind from 15 ° S - 15 ° N at 1 hPa are shown in **Figure 11.33**. The SAO's easterly phase is maximized

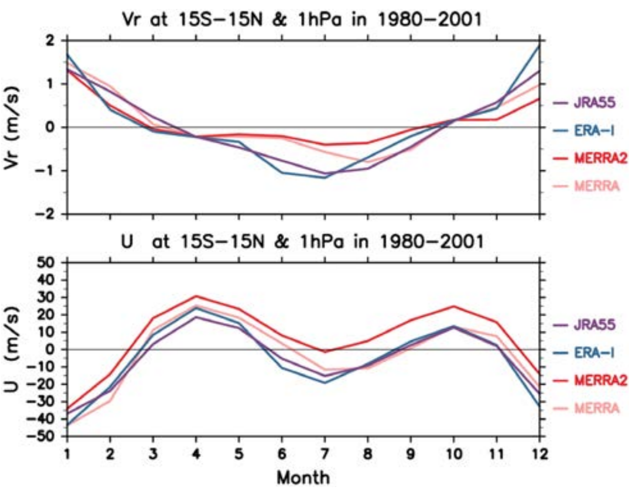


Figure 11.33: *Annual cycles averaged between 15° S and 15° N at 1 hPa of (top) the meridional component of the residual circulation and (bottom) the zonal mean zonal wind. Averages are over 1980 - 2001 and are given for JRA-55, ERA-Interim, MERRA, and MERRA-2.*

in January and July and stronger in January than in July for all the reanalyses, but MERRA-2 has a westerly bias compared to the other reanalyses throughout the year (see also *Section 11.3.1*). On the other hand, northward and southward residual-mean meridional flow in the Tropics is maximized in December and January and in July and August, respectively. In addition, maxima of northward and southward flow in MERRA-2 are smaller than those in the other reanalyses. These results are likely due to weaker planetary wave forcing in the winter subtropics (10 ° - 20 ° N) in MERRA-2, which is associated with weaker cross-equatorial flow and weaker transport of absolute angular momentum (*M*). In addition, while weaker cross-equatorial flow induces weaker SAO easterlies through the weaker transport (*e.g.*, *Tomikawa et al.*, 2008), the weaker SAO easterly phase in MERRA-2 cannot create horizontally aligned isopleths of *M* and this suppresses cross-equatorial meridional flow due to "sideways control." Thus, the weaker cross-equatorial meridional circulation in MERRA-2 could not only be induced by the weaker subtropical wave forcing but also through the interaction between the SAO easterly phase and cross-equatorial flow. In the tropical USLM, wave-driven and thermally-driven (*i.e.*, Hadley) circulations as well as "sideways control" each contribute to driving the meridional circulation; it is beyond the scope of this report to quantify individual contributions.

11.3.3 Inertial instability

Inertial instability is a hydrodynamic instability caused by an imbalance between the pressure gradient force and the centrifugal force. For zonally symmetric flow in the Earth's atmosphere, it is equivalent to the increase of the absolute angular momentum at latitudes moving away from the Equator. This condition is satisfied when a latitudinal shear of the zonal wind exists at the Equator, so that the inertial instability easily occurs in the Tropics. It creates vertically-stacked temperature structures (*i.e.*, pancake structures) induced by a local meridional circulation in the inertially unstable region (*cf.*, *Dunkerton*, 1981). An important role of inertial instability is to transport and homogenize the absolute angular momentum in the Tropics through the local meridional circulation, which partly contributes to an easterly phase of the SAO. A criterion of the inertial instability in zonally asymmetric flow is not yet established, but its analogue in zonally symmetric flow has been used in previous studies (*cf.*, *Knox*, 2003). Thus, we use $fq < 0$ as the criterion, where f is Coriolis parameter and q is Ertel's potential vorticity.

In this section, we show frequency of occurrence distributions of $fq < 0$, used here as a proxy for inertial instability. Ertel's PV (q) at 00 UT on each day was computed from MERRA, MERRA-2, ERA-Interim, and JRA-55, and used for the calculation. A horizontal resolution

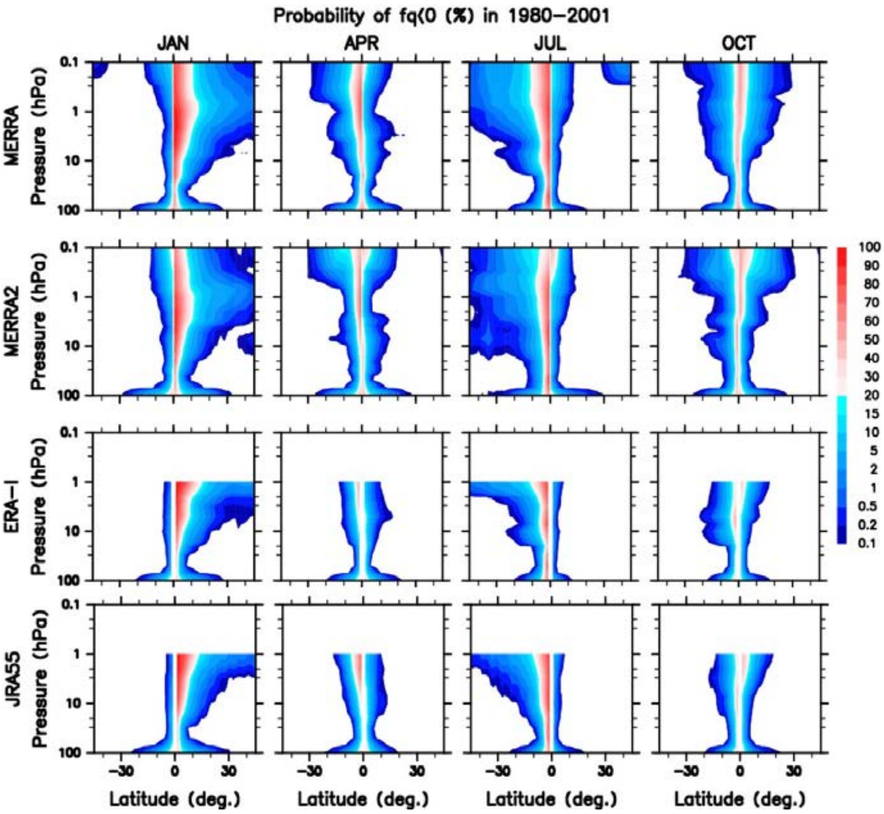


Figure 11.34: *Latitude-pressure sections of inertial instability frequency of occurrence rates in January, April, July, and October averaged over 1980-2001 for (from top to bottom) MERRA, MERRA-2, ERA-Interim, and JRA-55.*

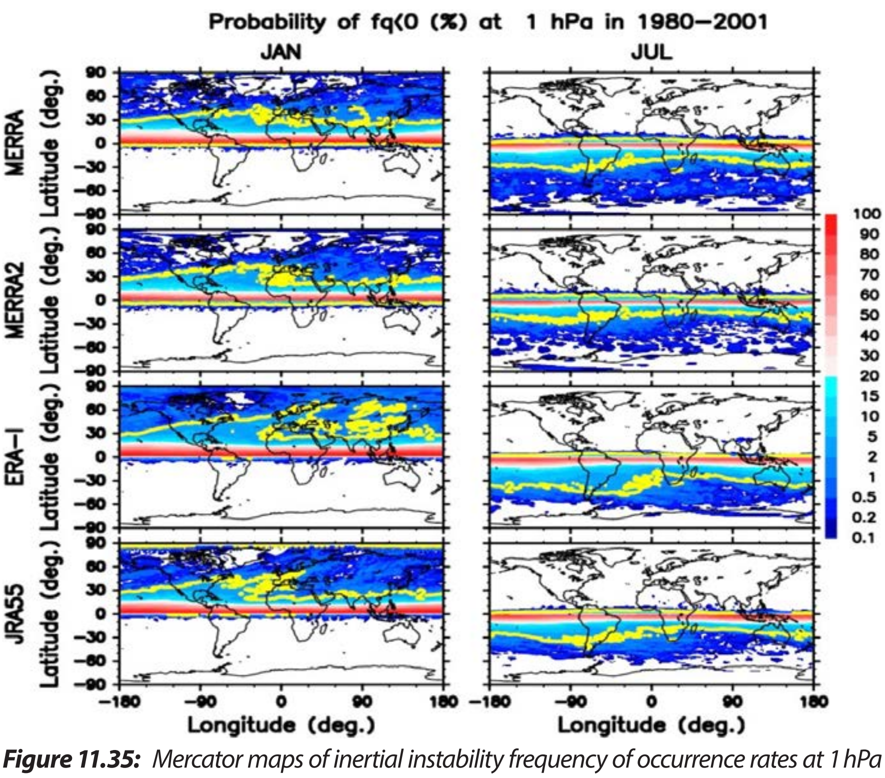


Figure 11.35: *Mercator maps of inertial instability frequency of occurrence rates at 1 hPa in January (left) and July (right) averaged over 1980-2001 for (from top to bottom) MERRA, MERRA-2, ERA-Interim, and JRA-55. The yellow contour indicates values of 2 %.*

of the reanalysis data is 1.25° longitude and 1.25° latitude for MERRA and JRA-55, 0.625° and 0.5° for MERRA-2, and 1.5° and 1.5° for ERA-Interim. A missing value is assigned at the Equator for ERA-Interim and JRA-55 because of $f = 0$ there. Here, inertial instability frequency of occurrence rates are given as the percent of the time that a given longitude, latitude, and pressure grid point satisfies $fq < 0$.

Figure 11.34 shows latitude-pressure sections of the inertial instability frequency of occurrence rates during January, April, July, and October in MERRA, MERRA-2, ERA-Interim, and JRA-55. Although the magnitude of the frequencies is different among the four reanalyses because of the different horizontal resolution of the data used, they show qualitatively good agreement in their latitudinal and vertical distributions. The inertial instability frequency is larger in winter than in summer and is maximized around the winter stratopause. This feature is consistent with stronger planetary wave driving in the winter hemisphere, and with the fact that the SAO's easterly phase is maximized in the summer (*i.e.*, the absolute angular momentum is maximized in the winter). Isolated regions of 1% occurrence rates near 40° latitude in MERRA in the summer near 0.1 hPa are due to noisier horizontal wind fields (and derived potential vorticity) compared to MERRA-2 (not shown).

Figure 11.35 shows Mercator maps of the inertial instability frequency at 1 hPa in January and July for each reanalysis. A tongue of high frequencies (highlighted by the yellow 2% frequency contour) stretches poleward in the winter western hemisphere in both the NH (*i.e.*, January) and the SH (*i.e.*, July). The higher inertial instability frequencies in the western hemisphere are consistent with the results of *Knox and Harvey* (2005), but its magnitude is smaller here. Poleward elongation of the region of higher frequencies is larger in ERA-Interim and JRA-55 than in MERRA and MERRA-2. Since the zonally asymmetric inertial instability pattern is due primarily to the planetary wave breaking process, the differences in inertial instability frequency among the reanalyses may be related to the differences in planetary wave activity among the reanalyses mentioned in *Section 11.2.1*.

11.4 Polar dynamics

11.4.1 Polar vortices

The circulation in the polar winter middle atmosphere is dominated by a large circumpolar vortex that forms as a result of decreased solar insolation (*Schoeberl and Hartmann*, 1991). These "polar vortices" are hemispheric in scale and persist throughout the winter in both hemispheres (*e.g.*, *Waugh and Polvani*, 2010; *Harvey et al.*, 2002, and references therein). They extend from the tropopause to the mesopause and they act to vertically couple the atmosphere-ionosphere system. For example, SSW disturbances to the polar vortex (*Butler et al.*, 2017; *Charlton and Polvani*, 2007, and references therein) are linked to weather patterns at the surface (*e.g.*, *Baldwin and Dunkerton*, 2001), mesospheric cooling (*Siskind et al.*, 2005; *Labitzke*, 1972), thermospheric warming (*Liu and Roble*, 2002; *Walterscheid et al.*, 2000), and anomalies in the ionosphere (*Goncharenko et al.*, 2010) both at high and low latitudes (*Pedatella et al.*, 2018). In the mesosphere-lower-thermosphere (MLT), descent in the vortex is required to transport reactive odd nitrogen produced by energetic particle precipitation from the thermosphere to the stratosphere (*Randall et al.*, 2015 and references therein). Throughout the stratosphere and mesosphere the shape and strength of the jet stream at the vortex edge affects vertical wave filtering (*Smith*, 1996; 1997). Thus, the polar vortices play an important role in coupling the atmosphere-ionosphere system.

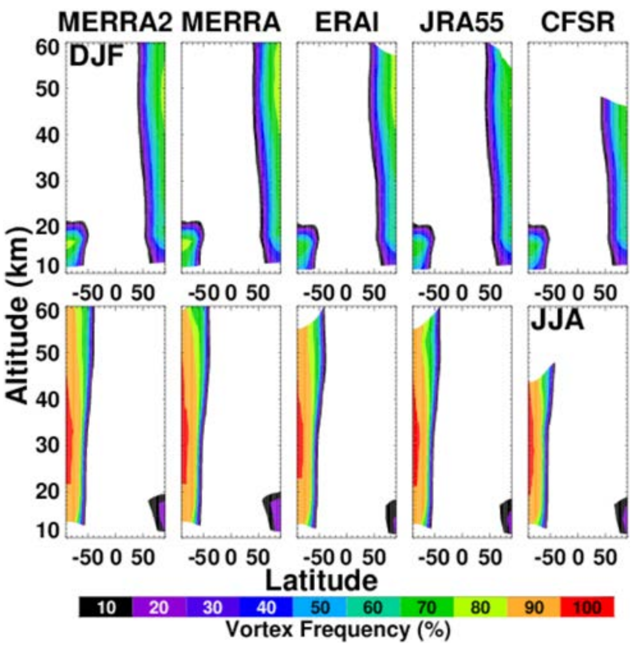


Figure 11.36: *Latitude-height plots of multi-year (2005 - 2015) average DJF (top) and JJA (bottom) polar vortex frequency as a function of latitude and altitude in five reanalyses (left to right, MERRA-2, MERRA, ERA-Interim, JRA-55, and CFSR).*

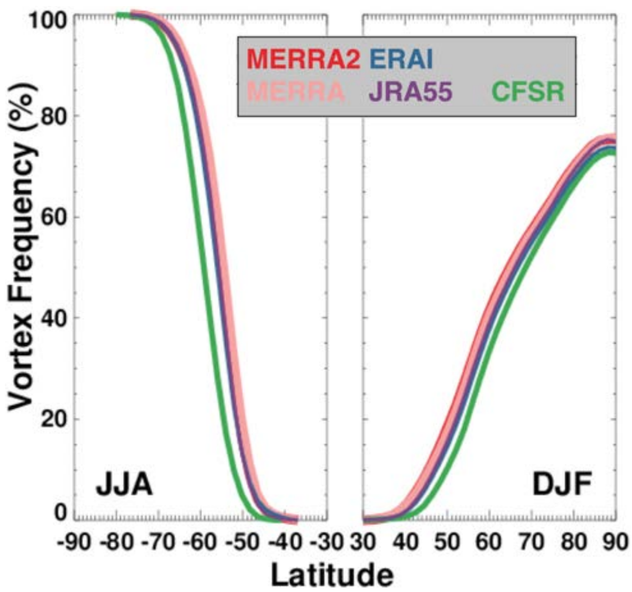


Figure 11.37: *Multi-year (2005 - 2015) average polar vortex frequency at one altitude in the upper stratosphere (1000 K; 2 hPa; 45 km) for JJA (left) and DJF (right) as a function of latitude.*

It is therefore of interest to evaluate the degree to which the reanalyses agree in terms of vortex structure and frequency of occurrence. In this work we identify the polar vortices using the streamfunction (ψ)-based algorithm described by *Harvey et al.* (2002). This vortex identification method is applied to MERRA-2, MERRA, ERA-Interim, JRA-55, CFSR/CFSv2 for the 11 years 2005 - 2015. In each reanalysis, on each day, for each altitude level and in each hemisphere, the polar vortex edge is defined and grid points inside (outside) the vortex are assigned a value of 1 (0). Thus, on each day a 3-D binary field of in-vortex points and exterior points is generated. Hereafter, vortex frequency of occurrence is defined as the percent of time a given grid point is located inside the vortex. As seen in **Figure 11.36**, all five reanalyses are in excellent agreement with respect to the latitude extension and magnitude of mean polar vortex frequency of occurrence rates during the winter months. These results demonstrate that all of these reanalysis datasets capture the primary multi-year winter mean vortex characteristics in both hemispheres.

In order to compare the reanalyses in more detail, we next look at depictions of polar vortex frequency in the upper stratosphere (**Figure 11.37**). This perspective reveals differences in vortex frequency that were obscured in the previous figure. At this altitude, multi-year mean polar vortex frequencies among MERRA, MERRA-2, ERA-Interim, and JRA-55 are in excellent agreement at all latitudes. CFSR wintertime vortex frequencies are 10 - 20 % lower than the other four reanalyses in the 50 ° to 70 ° latitude bands in both hemispheres. This could be attributed to the higher polar temperatures in the CFSR (see *Chapter 3*, **Figures 3.6** and **3.7**), hence a weaker polar night jet and lower vortex frequencies.

Figure 11.38 shows multi-year winter mean vortex frequency distributions in the longitude-altitude plane. All five reanalyses compared here are in excellent agreement with respect to their zonally asymmetric vortex frequency distributions. All five reanalyses contain a polar vortex that tilts westward with height from 15 km to stratopause altitudes. Overall, Antarctic vortex frequencies in CFSR (lower right panel) are lower compared to the other four datasets.

Finally, we examine the annual cycle in both the Arctic and Antarctic polar vortices near the stratopause in MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR. **Figure 11.39** shows multi-year (2005 - 2015) mean polar vortex frequencies (colored contours) as a function of geographic latitude and day of year at 1 hPa in the (**a**) NH and (**b**) SH in the five reanalysis datasets. MERRA-2 zonal mean winds are contoured in the background using thin black lines to provide dynamical context and illustrate how, at this altitude, 50 % frequency contours tend to coincide with maximum wind speeds. There is excellent agreement among the reanalyses in the vortex formation date in both hemispheres. In all five reanalysis datasets the Arctic vortex typically forms on 9 Sept and the Antarctic vortex forms on 6 March at this altitude.

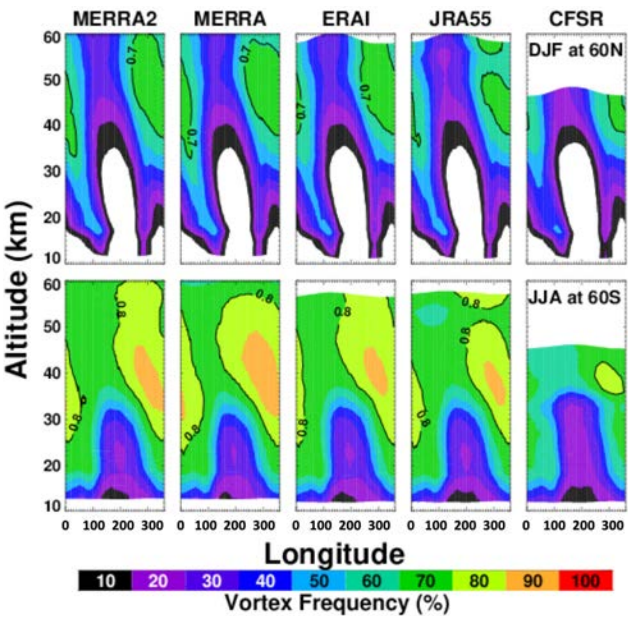


Figure 11.38: *Multi-year (2005 - 2015) average DJF (top) and JJA (bottom) polar vortex frequency as a function of longitude and altitude showing PW1 zonal asymmetry and a westward phase tilt with height.*

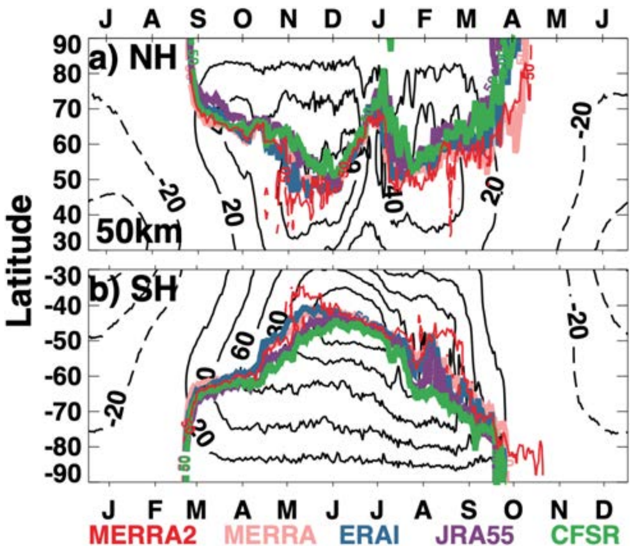


Figure 11.39: *Multi-year (2005 - 2015) mean polar vortex frequency (colored lines) as a function of geographic latitude and day of year at 1 hPa (~ 50 km) in the (a) NH and (b) SH in different 5 reanalysis datasets: MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2. Polar vortex frequencies correspond to the vortex being present 50 % of the time. Months in the NH are shifted such that the winter is in the middle of both panels. MERRA-2 zonal mean winds contoured every 20 m s^{-1} with thin black contour lines in the background. X-tick labels are on the 15th of each month. X-ticks appear on the top for the NH and on the bottom for the SH. Taken from Harvey et al. (2018). ©American Geophysical Union. Used with permission.*

There is also good overall agreement between the 5 reanalyses in the evolution of vortex latitudinal extent and duration. The polar vortices tend to be most confined to polar latitudes in CFSR/CFSv2 and present at lower latitudes in MERRA-2, and these differences are consistent with slight (1 ° - 2 °) differences in the mean latitude of the PNJ (not shown). The MERRA-2 (red) 50 % vortex frequency contour extends 5 - 10 % further equatorward than in the other four reanalyses, which reflects the sensitivity of the identification algorithm to slight differences in the horizontal winds. There is a 5-day spread in Arctic vortex breakup date and a 3-day spread in Antarctic vortex breakup date (when vortex frequencies go to zero, not shown). However, there is a significant amount of interannual variability in vortex breakup date, so small differences shown here are not necessarily representative of agreement in vortex longevity on a year-to-year basis. These results show that all of these reanalysis datasets sufficiently capture the multi-year mean seasonal evolution of the vortex at the stratopause during 2005 - 2015. However, users need to bear in mind that these multi-year averaged comparisons do not quantify the extent to which the reanalyses differ during individual years.

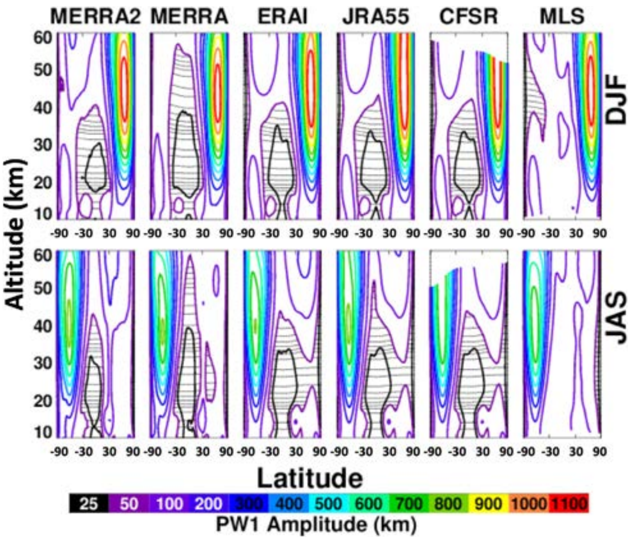


***Figure 11.40:** Latitude-altitude plots of multi-year (2005 - 2015) mean PW-1 amplitudes based on MERRA, MERRA-2, ERA-Interim, JRA-55, CFSR, and MLS GPH for DJF (top) and JAS (bottom). Black symbols denote tropical regions where PW-1 amplitudes are smaller than 50 m.*

11.4.2 Planetary waves

According to *Charney and Drazin* (1961), low-frequency planetary waves (PWs) propagate upward from the troposphere during winter and grow to have large amplitudes in the stratosphere and mesosphere. PWs are supported by a northward potential vorticity gradient and can arise from tropospheric forcing due to zonal variability in solar heating (land-sea thermal contrasts), flow over large-scale orographic features, as well as through the growth of normal modes due to local instability, leading to both stationary and traveling PWs (*e.g.*, *Andrews et al.*, 1987). PWs deposit their momentum in the surf zone where winds are weak (*e.g.*, *Sassi et al.*, 2002). It is well known that PWs contribute significantly to the momentum budget of the stratosphere and mesosphere, act to redistribute trace species both meridionally and vertically (*e.g.*, *Kouker and Brasseur*, 1986), and create zonal asymmetries in temperature, winds, and gravity wave propagation (*Lieberman et al.*, 2013; *Smith*, 1996; 1997). Quasi-stationary planetary wave-1 (PW-1) structures have been documented extensively in observations of temperature, winds, and trace gas distributions (*e.g.*, *Demirhan Bari et al.*, 2013; *Ialongo et al.*, 2012; *Gabriel et al.*, 2011; *Offermann et al.*, 2003; *Allen et al.*, 2000; *Barnett and Labitzke*, 1990; *Hirota and Barnett*, 1977). In this section we compare the representation of quasi-stationary PW-1 patterns among MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 during the 11 years spanning 2005 - 2015.

First, decadal average PW-1 latitude-altitude structures are compared in **Figure 11.40** during solstice. There is excellent agreement among the reanalyses and with MLS observations (far right column) in the magnitude and latitude-height structure of PW-1 amplitudes during the winter season. There are subtle differences in regions of the atmosphere where PW amplitudes are small; these occur in the Tropics and in the summer hemisphere. PW-2 results are similar (not shown).

Figure 11.41 shows the annual cycle of PW-1 amplitudes at 1 hPa near the stratopause at 60° latitude (left column) where amplitudes are largest and at 30° latitude (right column) where differences are large. Results indicate that there is remarkable agreement in winter PW-1 amplitudes among the reanalyses at 60° latitude in both hemispheres. In fact, the most apparent differences occur in the subtropics during summertime when PW-1 amplitudes are smallest (right column, note different y-axis range). It is interesting that all reanalyses except MERRA overestimate PW-1 amplitudes in the summer, in both hemispheres at this altitude level. This result is puzzling and requires further analysis.

11.4.3 Elevated stratopause events

Despite the absence of sunlight and the corresponding heating from shortwave absorption by ozone, the stratopause in polar night remains a well-defined feature of the general circulation. The temperature maximum is formed instead by adiabatic descent associated with the enhanced breaking of gravity waves which are also responsible for closing off the strong westerlies that form in the wintertime stratosphere (*Hitchman et al.*, 1989). The altitude of the stratopause is thus subject to dynamical variability, which is demonstrated most spectacularly over the Arctic in what are known as elevated stratopause

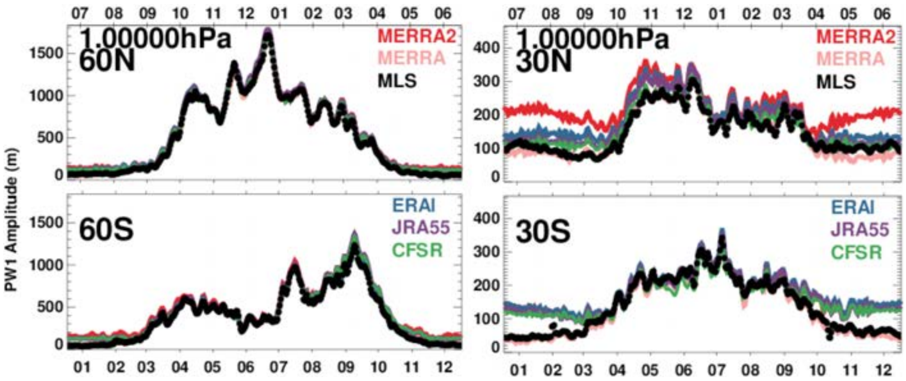


***Figure 11.41:** Line plots of multi-year average (2005-2015) annual cycles of PW-1 amplitude at 30° and 60° N/S at 1 hPa near the stratopause. Note different y-axis ranges at the 60° vs. at 30° latitude. The months are shifted in the NH such that winter is in the middle of all panels.*

(ES) events (*Chandran et al.*, 2013; *Limpasuvan et al.*, 2012; *Tomikawa et al.*, 2012; *Manney et al.*, 2008; *Siskind et al.*, 2007; 2010). These often occur during major SSWs (as first noted by *Labitzke*, 1972), when the persistently weakened lower stratospheric westerly jet and super-recovery of the upper stratospheric winds strongly modify the spectrum of gravity waves propagating up to the USLM (*Hitchcock and Shepherd*, 2013; *Tomikawa et al.*, 2012). They are characterized by a rapid (one to two week) initial descent or even disappearance of the polar stratopause, followed by the reformation of a temperature maximum at pressure levels as high as 0.03 hPa which, over the subsequent month or two, descends gradually to its climatological altitude near 1 hPa.

The initiation of the stratopause reformation as well as the slower descent of the mesospheric temperature anomalies are associated with an anomalous mean meridional circulation driven by planetary-scale Rossby waves likely generated within the middle atmosphere, as well as by small-scale gravity waves (*Hitchcock and Shepherd*, 2013; *Limpasuvan et al.*, 2012; *Tomikawa et al.*, 2012). The mechanism is related to but distinct from that responsible for the descent of tropical zonal jets in the QBO (*Hitchcock and Shepherd*, 2013). The stratopause reformation and descent is of particular importance for transporting trace species from the lower thermosphere into the mesosphere, as well as into the polar stratosphere (*Orsolini et al.*, 2017; *Siskind et al.*, 2010; *Manney et al.*, 2009; *Randall et al.*, 2006). The strong perturbation of the winds and temperatures throughout the stratospheric column and the interaction between stratospheric and mesospheric levels makes these events a challenging test for the representation of the gravity wave field in a forecast model. Given the general scarcity of observations to assimilate in the USLM, the forecast models of reanalyses are generally left to fend for themselves.

Figure 11.42 shows the evolution of the polar cap averaged temperature field plotted against time and pressure during

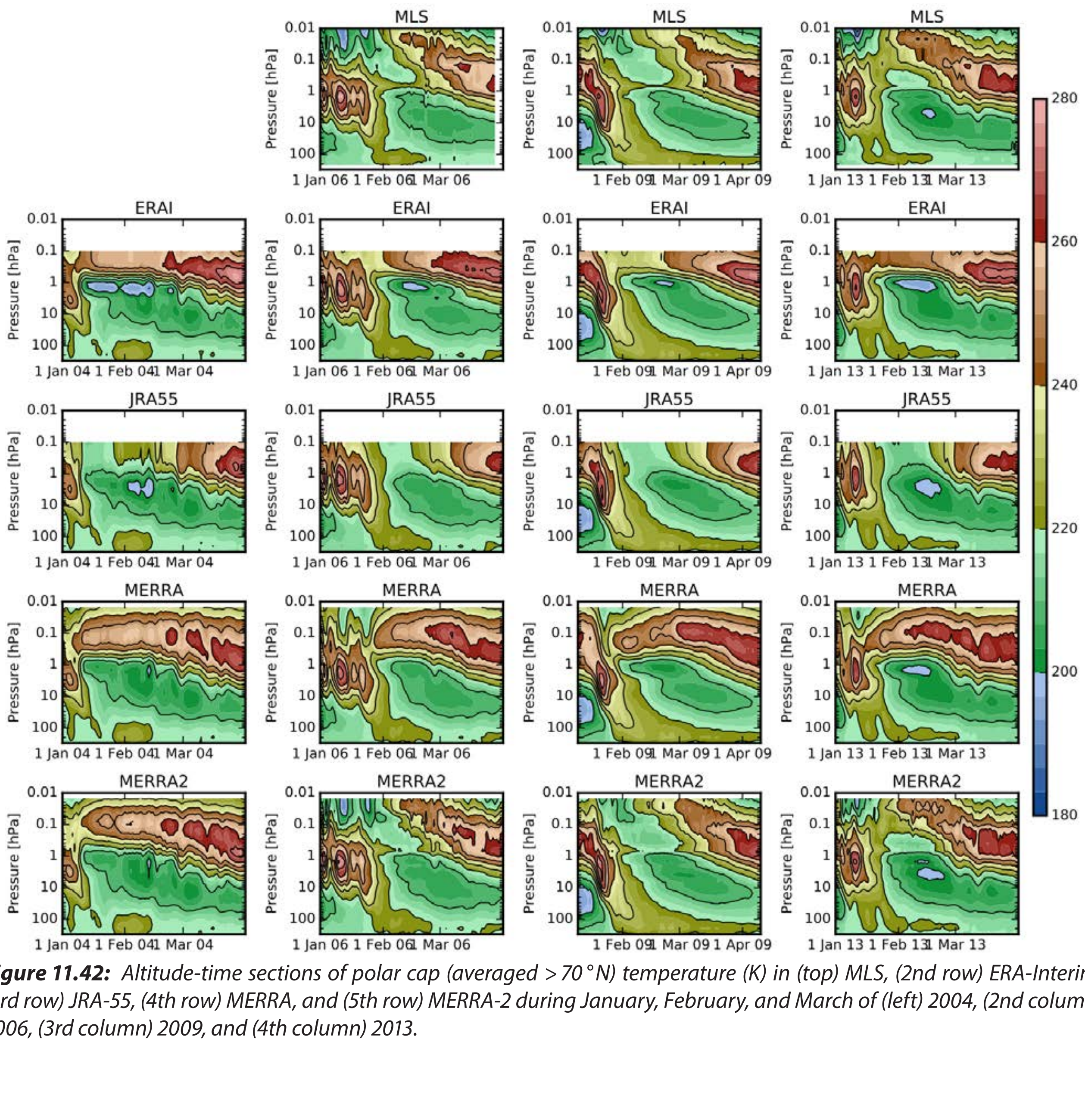


Figure 11.42: *Altitude-time sections of polar cap (averaged > 70 °N) temperature (K) in (top) MLS, (2nd row) ERA-Interim, (3rd row) JRA-55, (4th row) MERRA, and (5th row) MERRA-2 during January, February, and March of (left) 2004, (2nd column) 2006, (3rd column) 2009, and (4th column) 2013.*

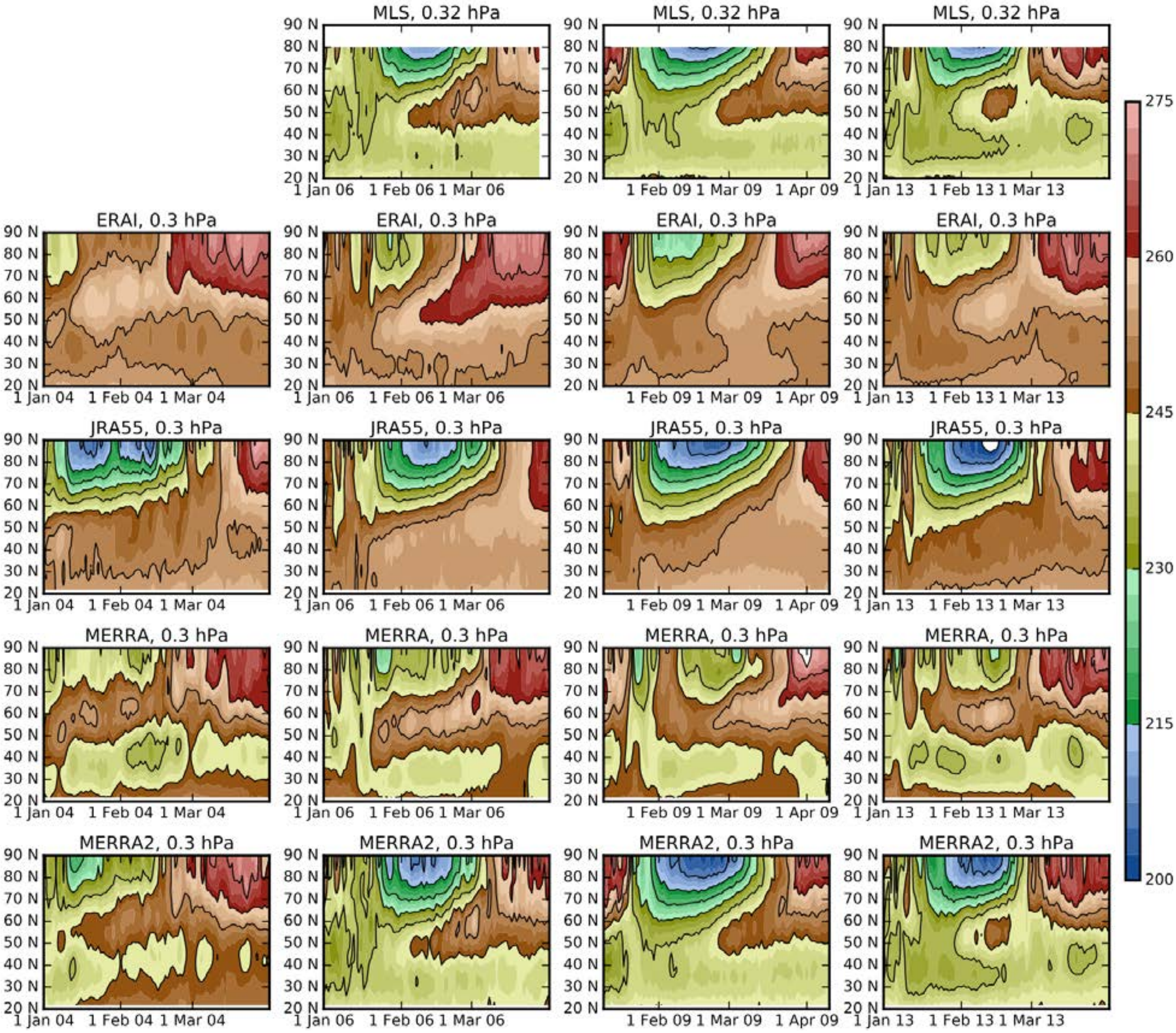


Figure 11.43: *Same as* ***Figure 11.42*** *but latitude-time plots at 0.3 hPa (~ 60 km).*

four recent ES events that occurred in the Januaries of the years 2004, 2006, 2009, and 2013. These events coincided with major SSWs with 10 hPa, 60 °N zonal mean zonal wind reversals that occurred on 5 Jan 2004, 21 Jan 2006, 24 Jan 2009, and 6 Jan 2013. Temperatures retrieved from the MLS instrument aboard the Aura satellite are available for the later three events and serve as a good observational reference (top row of **Fig. 11.42**). The general evolution of the stratopause described above is evident in each case. There is some inter-event variability in the temperature of the elevated stratopause immediately following the disappearance of the climatological stratopause, as well as in the cold anomaly near 6 hPa that forms as the stratopause descends, but the overall similarity among the events is remarkable and apparent even in these three events.

The corresponding temperatures from each of the four reanalyses are shown in the subsequent four rows. Three of the four reanalyses (ERA-Interim, MERRA, and MERRA-2) produce spuriously high temperatures near 0.1 hPa during the initial phase of the 2004 event. Similar biases are seen during the other three events in ERA-Interim and MERRA. MLS temperatures are assimilated in MERRA-2 when available; the effects of this assimilation on these events are immediately apparent. This may be associated with model lid or sponge layer effects in these three cases; however it is interesting that this temperature maximum occurs well below the model top (0.01 hPa) in the case of MERRA and MERRA-2. Also apparent is the formation of a shallow cold layer just below 1 hPa that is most pronounced in ERA-Interim but is also apparent in MERRA in all four events and in MERRA-2 during the 2004 event. The overall similarity between MERRA and MERRA-2 during the 2004 event suggests that the underlying forecast model in the two reanalyses treat these events quite similarly. The model lid of JRA-55 is at 0.1 hPa and it is therefore also unable to capture the elevated stratopause at its maximum altitude near 0.03 hPa. However, the JRA-55 reanalysis does not exhibit the temperature dipoles apparent in the other reanalyses, instead exhibiting nearly isothermal layers above 1 hPa that correspond to some extent with the descending stratopause seen in the observations. *McLandress et al.* (2013) studied the 2006 and 2009 events with the nudged Canadian Middle Atmosphere Model, which reproduced the MLS results almost perfectly (see their **Figures 3** and **4**). However the agreement degraded considerably when either the orographic or the non-orographic gravity wave drag, or both, were turned off (**Figures 5** and **6** of *McLandress et al.,* 2013), which shows the critical role of parameterized gravity wave drag in driving the mesospheric response to SSWs when there are no observations being assimilated in the mesosphere, and only the stratospheric state is being constrained.

Figure 11.43 shows a similar set of figures but now displays temperatures as a function of time and latitude at 0.3 hPa, an altitude just above the climatological stratopause. In the observed events, the cold anomaly is strongest at latitudes poleward of 60 ° N. Higher temperatures occur at lower latitudes first, giving rise about one month after the disappearance of the climatological stratopause to regions warmer than 245 K near 50 ° N. These spread to the pole as the elevated stratopause descends. Latitudinal structure similar to this is seen in ERA-Interim, MERRA, and MERRA-2 even prior to the assimilation of MLS data, although overall temperatures are biased considerably high at this level, consistent with what was seen in the previous figure. In contrast, while the lower temperatures at higher latitudes in JRA-55 agree well with observations, there are no relatively warm regions to the south.

The temperature structures in ERA-Interim, MERRA, and MERRA-2 (when not constrained by the assimilation of MLS temperatures) are consistent with a strong sponge layer feedback caused by the presence of artificial momentum damping within the USLM, which, in these reanalyses, is acting to reduce the zonal mean zonal wind above 0.3 hPa (not shown). This feedback produces a spurious anomalous meridional circulation cell between the sponge layer and the region of anomalously low wave driving in the mid-stratosphere as described by *Shepherd et al.* (1996); see in particular their **Figure 1c**, though in this case the sign of the effective forcing is in the opposite sense. The different behavior of JRA-55 is consistent with the use of thermal damping towards the layer-averaged global mean temperature and the lack of momentum damping, unlike the other three reanalyses which use some form of momentum dissipation in their sponge layers (*Chapter 2*; *Fujiwara et al.*, 2017). Near a sponge layer, the presence or lack of a non-orographic gravity wave parameterization would seem to be of lesser direct relevance given that MERRA and MERRA-2 include such parameterizations while ERA-Interim and JRA-55 do not (*Chapter 2*; *Fujiwara et al.*, 2017). However, if the model lid is sufficiently high and the sponge is not applied to the zonal mean flow, *McLandress et al.* (2013) showed that non-orographic gravity waves can be important.

It is important to note that in almost all cases the representation of ES events in the USLM is not constrained by the assimilation of observations, and in all cases the effects of the forecast model sponge layers are directly felt and likely will be associated with artificial meridional circulations (*Shepherd et al.*, 1996). These inferred biases in the meridional circulations will have an effect on the inferred tracer transport as well. The assimilation of MLS temperatures into MERRA-2 results in close agreement with MLS temperatures during the period where observations are available, but events prior to this should be treated separately, and the presence of the sponge layer implies that the corresponding meridional circulation should not be trusted even during the assimilation period. The use of thermal dissipation as the sponge layer in JRA-55 avoids the strong spurious circulations that affect MERRA, ERA-Interim, and MERRA-2 prior to the assimilation of MLS temperatures, but the physical circulation evident in the observed temperature structure is also missing. Studies of ES events that make use of reanalyses must be aware of these shortcomings.

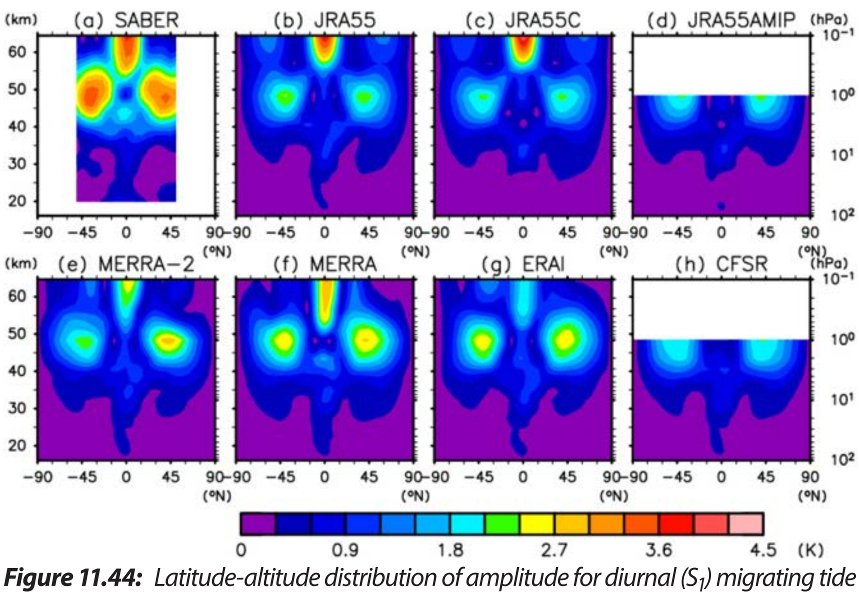


***Figure 11.44:** Latitude-altitude distribution of amplitude for diurnal (S_1) migrating tide in temperature (K), as derived from (a) SABER, (b) JRA-55, (c) JRA-55-C, (d) JRA-55-AMIP, (d) MERRA-2, (e) MERRA, (f) ERA-Interim, and (h) CFSR. Taken from Sakazaki et al. (2018).*

11.5 Tides and normal modes

11.5.1 Tides

Atmospheric solar tides are global-scale inertia-gravity waves with periods that are integer fractions of a solar day (*Chapman and Lindzen*, 1970). They are primarily driven by diurnally varying diabatic heating, such as the absorption of solar radiation by tropospheric water and stratospheric ozone, and the latent heat release associated with tropical

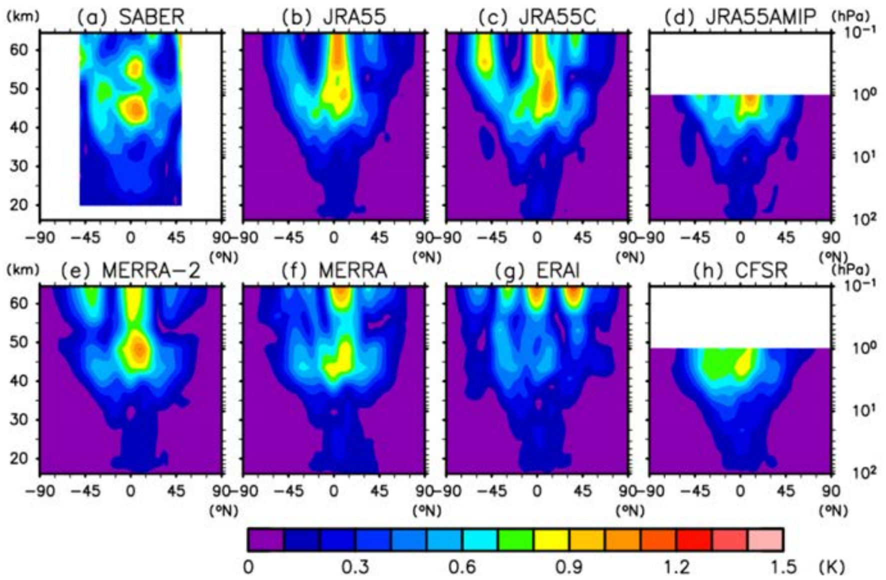


***Figure 11.45:** Same as **Fig. 11.44** but for semidiurnal (S_2) migrating tide. Taken from Sakazaki et al. (2018).*

convection (*Hagan and Forbes*, 2002; *Hagan et al.*, 1995). The diurnal (S_1) and semidiurnal (S_2) variations around the globe can be decomposed into zonal harmonics with the "migrating" (Sun-synchronous) components for the S_1/S_2 tides represented by westward propagating wave number one/two. The remainder of the tidal zonal harmonics are "non-migrating components" and are excited mainly by zonally-asymmetric variations in (local time) heat sources or topography. Tides in reanalysis data provide an important "lower boundary condition" for driving so-called whole atmosphere models (*e.g.*, *Pedatella et al.*, 2014). They are also used for correcting the diurnal anomaly or drift seen in Sun-synchronous satellite measurements (*Zou et al.*, 2014).

We analyze and compare data from five recent global reanalyses (MERRA-2, MERRA, JRA-55, ERA-Interim and CFSR) as well as SABER (not assimilated in any reanalyses) and MLS satellite measurements (only assimilated in MERRA-2) during the 7-year period 2006-2012. For JRA-55, the two other "family" members, JRA-55C and JRA-55AMIP, are analyzed to examine the effects of data assimilation on the representation of the solar tides. We will not consider here the JRA-25, ERA-40 and NCEP1/2 reanalyses. *Sakazaki et al.* (2012) showed that the global structure and seasonality of the S_1 migrating tide represented in JRA-25 or NCEP1/2 were less consistent with available observations than were the newer reanalyses data sets.

In this report, the (1) diurnal (S_1) migrating tide, (2) semidiurnal (S_2) migrating tide, and (3) nonmigrating tides are extracted and diagnosed individually. For SABER, a composite analysis is made at each longitude-latitude bin after a 60-day running mean that is regarded as daily-mean is subtracted from the original time series (see *Sakazaki et al.* 2018 for details). For reanalyses, first, 3- or 6-hourly diurnal variations in universal time (UT) are extracted at each grid point with a composite analysis after the subtraction of the daily-mean. Next, by averaging data at the same local time (LT) for each latitude band, migrating tides that are a function of LT are calculated; for example, for 6-hourly reanalyses, data at 0000 LT is the average of data points at (0000 UT, 0°E) (0600 UT, 90°E) (1200 UT, 180°E) (1800 UT, 270°E), while data at 01:00 LT is the average of data points at 00:00 UT (15°E), 06:00 UT (105°E), 12:00 UT (195°E) and 18:00 UT (285°E). Then, the harmonic fitting is performed for the diurnal variations in LT to extract the migrating S_1 and S_2 components. Note that the 6-hourly data (ERA-Interim, JRA-55 and CFSR) cannot resolve S_2 at each grid point; but the 'migrating component' of S_2 can be extracted by using data at grid points on the same latitude belt as noted above. Finally, we diagnose S_1 nonmigrating tides by applying the zonal wavenumber decomposition for the S_1 component (*Dai and Wang*, 1999). See *Sakazaki et al.* (2018) for the comparison of nonmigrating tides in physical space.

Figure 11.44 shows the latitude-altitude distribution of amplitude for annual-mean S_1 migrating temperature tides computed from SABER data (upper left panel) and from the various reanalyses from 2006-2012. Both reanalyses and observations show that tidal amplitudes increase with altitude in the Tropics; tides based on SABER observations reach ~4 K in the tropical lower mesosphere. This feature is underestimated by 30-50% in the various reanalyses. The tidal maxima in the midlatitude upper stratosphere are similarly underestimated by the reanalysis systems by 20-30% compared to SABER. Notably the JRA-55C and JRA-55AMIP results are close together and differ from the JRA-55 results, indicating that satellite measurements improve the tidal representation in reanalyses.

Figure 11.45 shows the latitude-altitude distribution of amplitude for annual-mean S_2 migrating tides in temperature. Observations and reanalyses indicate that amplitudes are largest in the Tropics, with a local maximum around at 40-45 km (up to ~1.2 K), *i.e.*, close to the location of maximum in ozone heating. Note that the ERA-Interim overall shows a smaller amplitude in the stratosphere (reduced by up to ~50% compared to SABER and the other reanalyses).

Figure 11.46 shows the zonal wavenumber dependence for the annual-mean S1 (24 hour) harmonic of non-migrating tides for each symmetric and anti-symmetric component with respect to the Equator (migrating component,

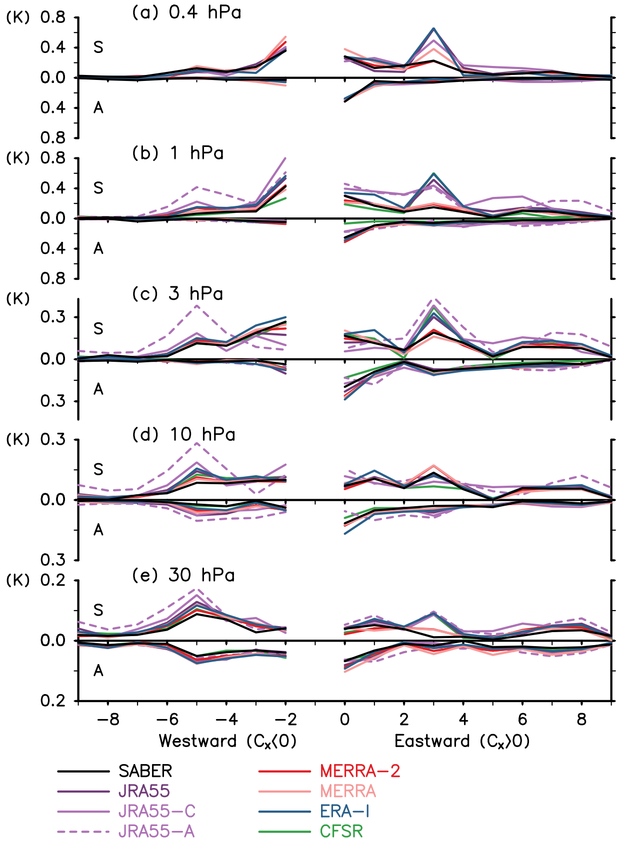


Figure 11.46: *Amplitudes for each zonal wavenumber component of diurnal (S_1) nonmigrating tides for the region between 10°S and 10°N, at (a) 0.4 hPa, (b) 1 hPa, (c) 3 hPa, (d) 10 hPa and (e) 30 hPa. Top and bottom half in each panel shows the results of symmetric and anti-symmetric components, respectively. Positive and negative wavenumbers are for the eastward and westward travelling waves, respectively. The S1 migrating tide (westward wavenumber 1) is not shown. Modified from Sakazaki et al. (2018).*

westward zonal wavenumber 1, is not shown). All data sets show that zonal wavenumber 0 (so-called D0; particularly for anti-symmetric components), westward zonal wavenumbers 5 and 2 (DW5 and DW2), and eastward zonal wavenumber 3 (DE3) are dominant, being consistent with previous studies (*Sakazaki et al.*, 2015; *Forbes and Wu*, 2006; *Zhang et al.*, 2006). Although the dominant wavenumbers agree among the data sets, their magnitudes display some differences. The biggest outliers are JRA-55C and JRA-55AMIP and those two datasets display somewhat larger amplitudes than the full-input reanalyses (that assimilate both conventional and upper air observations). Another marked difference is seen for DE3; the MERRA and MERRA-2 results are close to the SABER but the other reanalyses have larger amplitudes than SABER above the middle stratosphere (pressures less than 3 hPa).

To summarize, the latest reanalyses agree reasonably well with each other and with the satellite observations for both migrating and nonmigrating components including their vertical and meridional structure. However, the agreement among reanalyses is better in the lower stratosphere and differences increase in the USLM. The diurnal migrating tides are weaker in the reanalyses compared to SABER, although such differences are less clear between MLS and the reanalyses (not shown). Reanalyses are a very useful tool to investigate the global structure of tides and its temporal variability. At the same time, one should note that the representation of tides is significantly affected by assimilated satellite data so that the present intercomparison results during 2006-2012 do not necessarily apply to other periods with different data assimilated (especially before 2000 when AMSU was not assimilated; see **Fig. 11.46.** See also *Sakazaki et al.*, 2018 for more details.

11.5.2 Quasi-2-day wave

The quasi-2-day wave (QTDW) is a well-documented feature of upper stratospheric and mesospheric dynamics that consists primarily of a westward propagating zonal wave number 3 that moves around a latitude circle in approximately 6 days. With a wave-3, this leads to local oscillations of 2 days, giving it the name "2-day wave". Early evidence of the QTDW was found in wind observations and radiances from meteor radar, satellite, and rocket-borne instruments (*e.g.*, *Burks and Leovy*, 1986; *Rodgers and Prata*, 1981; *Coy*, 1979; *Muller and Nelson*, 1978). Subsequent analyses of wind, temperature, and constituent observations from a plethora of ground-based and satellite-based instruments have shown the QTDW to be a major, recurring dynamical feature in the mesosphere and lower thermosphere (MLT) that is most prominent in the extratropical summer hemisphere (*Gu et al.*, 2013; *Tunbridge et al.*, 2011; *Pancheva*, 2006; *Garcia et al.*, 2005; *Limpasuvan and Wu*, 2003; *Lieberman*, 1999; *Harris*, 1994; *Wu et al.*, 1993).

Detailed studies of the upper stratospheric QTDW using operational meteorological analyses (*e.g.*, *Orsolini et al.*, 1997; *Randel*, 1994) have analyzed wavenumber-frequency spectra and potential vorticity-based diagnostics from daily wind and temperature fields. The results of these studies supported earlier theoretical results indicating that the QTDW originates primarily from regions of baroclinic instability in the easterly mesospheric summer jet (*Pfister*, 1985; *Plumb*, 1983), but also from regions of barotropic instability of the easterly jet in the subtropical upper stratosphere (*e.g.*, *Manney and Nathan*, 1990; *Burks and Leovy*, 1986), itself triggered by inertial instability (*Orsolini et al.*, 1997). The QTDW also projects onto a global zonal wavenumber 3 normal mode. Furthermore, these studies also clearly demonstrated the utility of stratospheric analyses for providing a fully self-consistent set of meteorological variables needed to describe the physical mechanisms that drive the QTDW and other key circulation features related to normal modes in the stratosphere, mesosphere, and lower thermosphere.

One aspect of the QTDW that is not yet well understood is the cause of its intraseasonal and interannual variability, which can have a wide-ranging effect on, *e.g.*, summer polar mesopause temperatures (*France et al.*, 2018; *Siskind and McCormack*, 2014), thermospheric neutral winds (*Chang et al.*, 2011), and ionospheric electron content (*Yue et al.*, 2012). Modeling and observational studies have shown that underlying variations in the background zonal wind field throughout the tropical and extratropical stratosphere and lower mesosphere that promote both baroclinic and barotropic instability are likely a key source of observed intraseasonal and interannual variability in the QTDW (*McCormack et al.*, 2014; *Rojas and Norton*, 2007; *Limpasuvan et al.,* 2000; *Norton and Thuburn*, 1999).

Reanalysis data sets extending into the USLM can now provide a more comprehensive understanding of the dynamical origins of QTDW variability. To this end, it is necessary to first understand how the QTDW is represented in current reanalysis data sets. In this section, the characteristics of the QTDW are compared using three reanalysis temperature data sets extending into the mesosphere: MERRA-2, JRA-55, and ERA-Interim. The comparisons are performed using data from 2010, focusing on the seasonal variability in the QTDW.

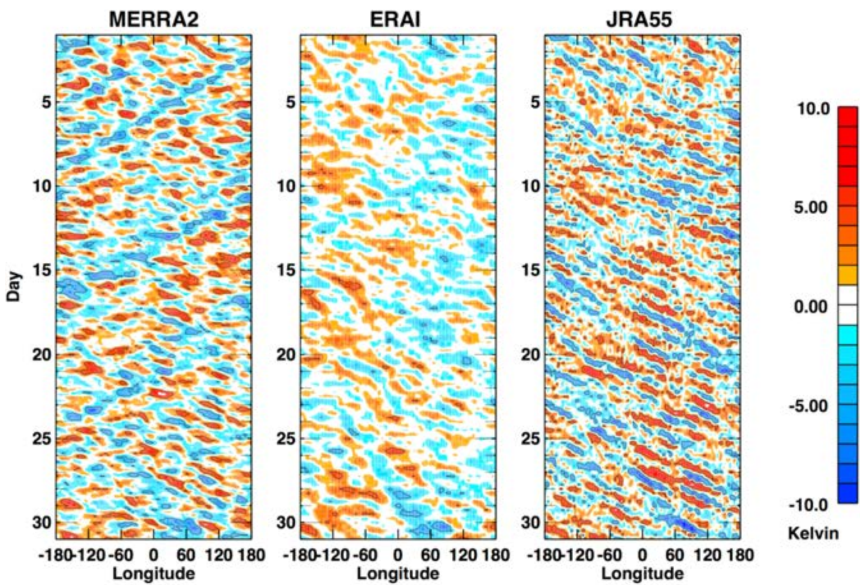


Figure 11.47: *Hovmöller diagrams at 30 °S and 0.3 hPa of temperature anomalies (minus the zonal and time mean) in MERRA-2 (left), ERA-Interim (middle), and in JRA-55 (right) during January 2010. Solid contours drawn at ± 4 K.*

The representation of the QTDW, and any other planetary scale normal modes, in reanalyses of the USLM will depend on a variety of factors. These factors include the vertical domain of the analysis system, the physical parameterizations used in the atmospheric model component, and the type of observations (if any) that provide information within this altitude region. In comparing the QTDW among the three reanalysis data sets, we note that there are two important features that distinguish MERRA-2 from JRA-55 and ERA-Interim. First, MERRA-2 extends to higher altitudes than ERA-Interim and JRA-55; the top pressure levels used for this comparison are 0.015 hPa for MERRA-2 and 0.1 hPa for both ERA-Interim and JRA-55. Second, only MERRA-2 assimilates temperature observations from the MLS instrument in the USLM.

To illustrate how these and other differences impact the reanalysis, **Figure 11.47** shows Hovmöller plots of temperature anomalies (zonal and time mean subtracted at each grid point to remove stationary wave components) for 30 °S at the 0.3 hPa level for January 2010 from (left) MERRA-2, (center) ERA-Interim, and (right) JRA-55. During this time period, prominent westward- and eastward-propagating temperature anomalies in the range of ±5 K can be seen in MERRA-2. The corresponding ERA-Interim temperature anomalies are weaker, typically in the range of ±2 K, and show some eastward propagation but little to no westward propagation. The JRA-55 temperature anomalies at this latitude and pressure level exhibit higher frequency eastward propagating features than either ERA-Interim or MERRA-2 up to ± 10 K. We note that 0.1 hPa is the top reported level of the JRA-55 data set, and so the reanalysis may be influenced by model upper boundary effects. Although a more detailed comparison is needed to conclusively identify the reasons for the differences among the three reanalyses shown in **Figure 11.47**, this initial comparison illustrates that all reanalysis data sets (even MERRA-2) must be used with caution in the USLM. Ideally, any studies using reanalyses at these upper levels should include validation with independent observations whenever possible.

With this caveat in mind, we compare the representation of the QTDW in MERRA-2, ERA-Interim, and JRA-55 during 2010. To describe the characteristics of the QTDW, a two-dimensional fast Fourier transform (2DFFT) (*Hayashi*, 1971) in longitude and time is applied to the reanalysis temperature anomaly fields (see **Fig. 11.47**) at a given latitude and pressure level. Following the procedure described in *McCormack et al.* (2009), daily zonal means are subtracted from each 3-hourly (MERRA-2) or 6-hourly (JRA-55 and ERA-Interim) longitude-time field and then a cosine taper is applied to the first and last 10 % of each record in time. The resulting power spectrum describes the amount of variance at each frequency and zonal wave number. Variance associated with the QTDW is isolated by reconstructing the longitude-time fields using the inverse 2DFFT with a bandpass filter for a given zonal wavenumber at frequencies from 0.45 - 0.6 cycles per day. This frequency range was determined by examining individual wavenumber-frequency spectra from the reanalysis temperature data throughout the year at latitudes in the lower mesosphere where the QTDW signal is largest. The 2DFFT is applied to reanalysis temperature fields on a monthly basis, producing a mean amplitude of the QTDW over the month-long analysis interval. Observational studies of the global QTDW structure have found evidence of prominent westward zonal wavenumber 3 and wavenumber 4 components (*e.g.*, *McCormack et al.*, 2014; *Gu et al.*, 2013; *Tunbridge et al.*, 2011). Here we

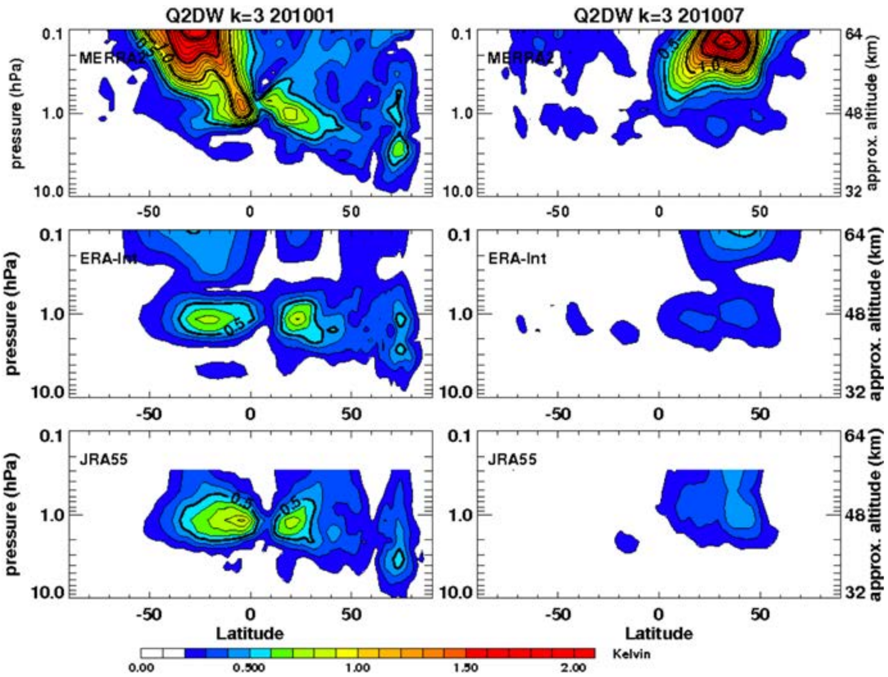


Figure 11.48: Latitude-altitude sections of quasi-2-day wave amplitudes for westward zonal wavenumber 3 for January (left) and July (right) in 2010 from MERRA-2 (top), ERA-Interim (middle) and JRA-55 (bottom). Thin contours drawn every 0.1 K starting at 0.2 K, thick contours drawn every 0.5 K.

examine both components, focusing on the months of January and July when the QTDW amplitudes are found to be largest in the respective summer hemispheres. For these comparisons, we limit our attention to the region of the USLM between 0.1 - 10 hPa.

Figure 11.48 plots the altitude and latitude dependence of mean QTDW wavenumber 3 amplitudes for January 2010 (left column) and July 2010 (right column) from MERRA-2 (top), ERA-Interim (middle), and JRA-55 (bottom). All three reanalyses show qualitatively similar latitudinal structure in the QTDW, but there are large quantitative differences. MERRA-2 shows the largest amplitude in the mid-latitude mesosphere and in a narrow subtropical tongue extending down to the stratopause level, likely tied to the aforementioned regions of jet instability. In both months, peak QTDW amplitudes from MERRA-2 range from 1.5 - 2 K from 10 ° - 50 ° latitude in the summer hemisphere above the 1 hPa level. These values are much larger than the peak amplitudes of 0.7 - 0.9 K seen in ERA-Interim and JRA-55, which are limited to the region between 1 - 2 hPa. This is to be expected since the MERRA-2 reanalysis extends higher in altitude and includes mesospheric temperature observations. All three reanalyses show similar seasonal behavior in the upper stratosphere, in that the zonal wavenumber 3 QTDW amplitudes are larger in SH summer (January) than in NH summer (July) near the 1 hPa level.

Figure 11.49 shows monthly mean amplitudes of the QTDW for zonal wavenumber 4 during January and July 2010 from the three different reanalyses. As with the wavenumber 3 case, here all three reanalysis data sets show similar latitude structure in the peak QTDW amplitudes. In contrast to the zonal wavenumber 3 component, the zonal wavenumber 4 QTDW is largest during NH summer throughout the stratosphere and lower mesosphere in MERRA-2, ERA-Interim, and JRA-55. Again, the peak values of the QTDW amplitudes are much larger in MERRA-2 temperatures than in JRA-55 and ERA-Interim, and only MERRA-2 exhibits a strong QTDW

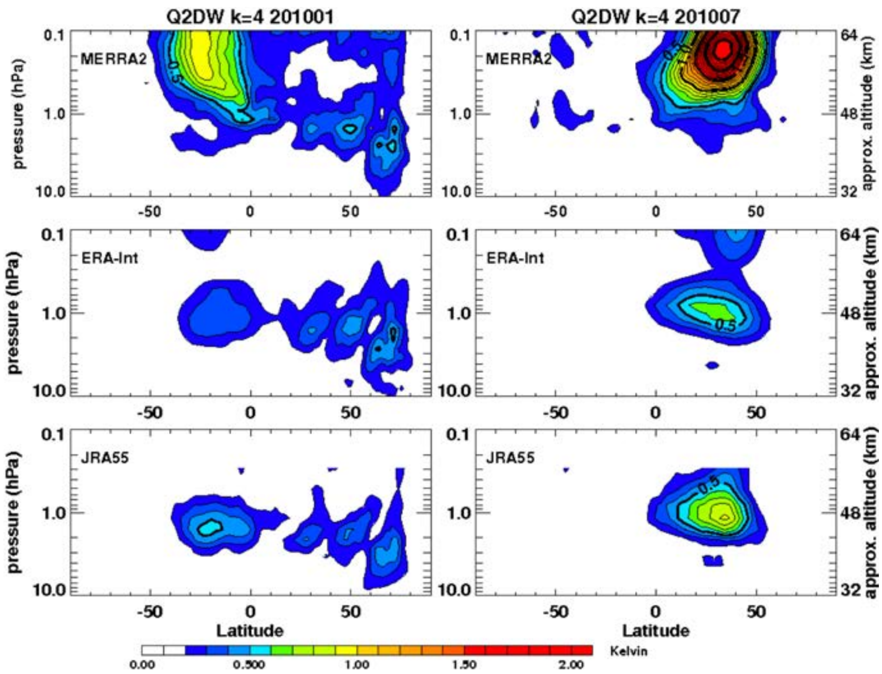


Figure 11.49: Latitude-altitude sections of quasi-2-day wave amplitudes for planetary wavenumber 4 in January (left) and July (right) in 2010 in MERRA-2 (top), ERA-Interim (middle) and JRA-55 (bottom). Thin contours drawn every 0.1 K starting at 0.2 K, thick contours drawn every 0.5 K.

signal above the 1 hPa level.

Since the QTDW is mainly a mesospheric phenomenon, the extended vertical domain and additional mesospheric observations from MLS allow the MERRA-2 reanalysis to better capture the main features of the QTDW in USLM temperatures as compared to ERA-Interim and JRA-55. Reanalysis systems with lower tops that lack mesospheric observations will not capture the main features of the QTDW. This reinforces the concept that the origin and propagation of the QTDW is mainly controlled by mesospheric dynamical variability. Reanalysis systems need to reproduce this variability in order to properly diagnose the physical mechanisms controlling both intraseasonal and interannual variability of the QTDW.

11.5.3 Quasi-5-day wave

The quasi-5-day wave (QFDW) consists of a westward propagating zonal wavenumber 1 disturbance that is related to the first symmetric normal (Rossby) mode. It has been observed in surface pressure observations (see, *e.g.*, *Madden and Julian*, 1972, and references therein), and is routinely found in reanalysis data sets throughout the tropical and extratropical troposphere and stratosphere, having a period ranging from ~ 4 - 7 days. The QFDW has also been observed in the mesosphere (*e.g.*, *Iimura et al.*, 2015; *Talaat et al.*, 2001, 2002). The origins of the QFDW in the stratosphere and lower mesosphere are complex, and may involve different mechanisms, including latent heat release in the tropical upper troposphere (*Miyoshi and Hirooka*, 2003), possible nonlinear interactions in the stratosphere between extratropical planetary scale waves propagating upward from the troposphere (*Talaat et al.*, 2002), and amplification via baroclinic instability in the mesosphere (*Lieberman et al.*, 2003). The QFDW plays a prominent role in the dynamics of the MLT region, particularly in the occurrence of polar mesospheric clouds (PMCs) at high latitudes in summer (*Nielsen et al.*, 2010).

Given the complex dynamical interactions throughout the troposphere, stratosphere, and mesosphere that give rise to the QFDW, capturing the key characteristics of the global circulation feature is a good test for reanalysis systems extending into the USLM region. In this section we compare the QFDW in temperature during 2010 from the MERRA-2, ERA-Interim, and JRA-55 reanalyses using the 2DFFT method described in the previous section. For this comparison, the 2DFFT is applied to temperature anomaly fields using a bandpass for westward zonal wavenumber 1 and 0.16 - 0.25 cycles per day (periods of 4.25 - 6 days). This frequency range was determined by examining individual wavenumber-frequency spectra from the reanalysis temperature data throughout the year at latitudes in the stratosphere where the QFDW signal is largest.

Figure 11.50 shows the latitude and altitude dependence of the mean QFDW temperature amplitudes from 10 - 0.1 hPa during January 2010 (left column) and August 2010 (right column) from MERRA-2 (top), ERA-Interim (middle), and JRA-55 (bottom). The January 2010 results from all three reanalyses show remarkably consistent results, both qualitatively and quantitatively. In January, two distinct patterns emerge. The first pattern is relatively weak (0.6 - 1 K) and hemispherically symmetric with maxima near 40 °N and 40 °S from 1 - 3 hPa. This pattern is qualitatively consistent with the theoretical structure of the first symmetric normal mode. The second pattern is much stronger (2 - 3 K) and is present throughout the stratosphere and lower mesosphere at high Northern latitudes, in contrast to the expected theoretical structure of the first symmetric normal mode. There is hemispheric asymmetry in this second QFDW pattern in the sense that amplitudes in January are substantial poleward of 60 °N, but are extremely small in the SH polar regions. This high northern latitude signal extends above the 1 hPa level in the MERRA-2 results, consistent with the

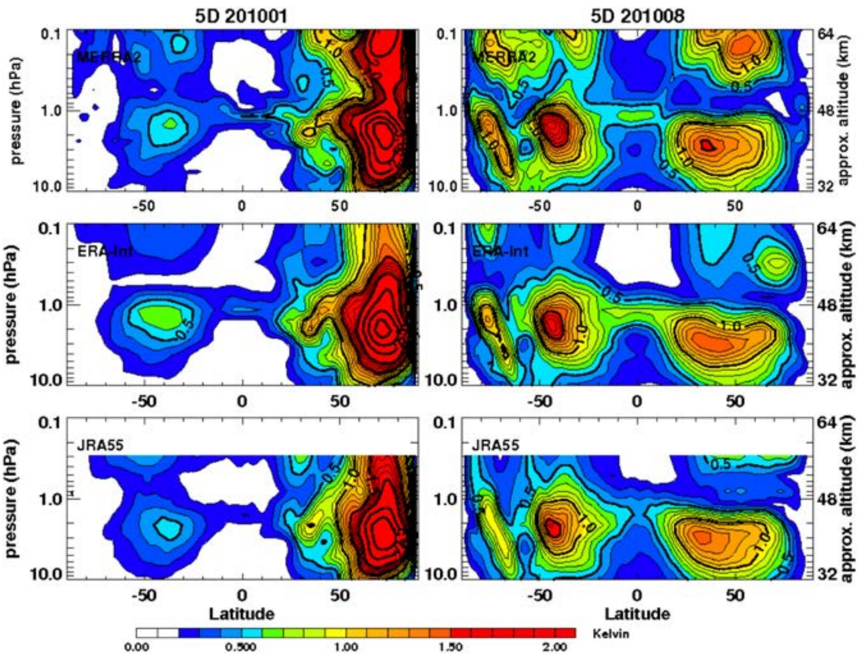


Figure 11.50: Latitude-altitude sections of quasi-5-day wave amplitudes for January (left) and August (right) in 2010 in MERRA-2 (top), ERA-Interim (middle) and JRA-55 (bottom). Thin contours drawn every 0.1 K starting at 0.2 K, thick contours drawn every 0.5 K.

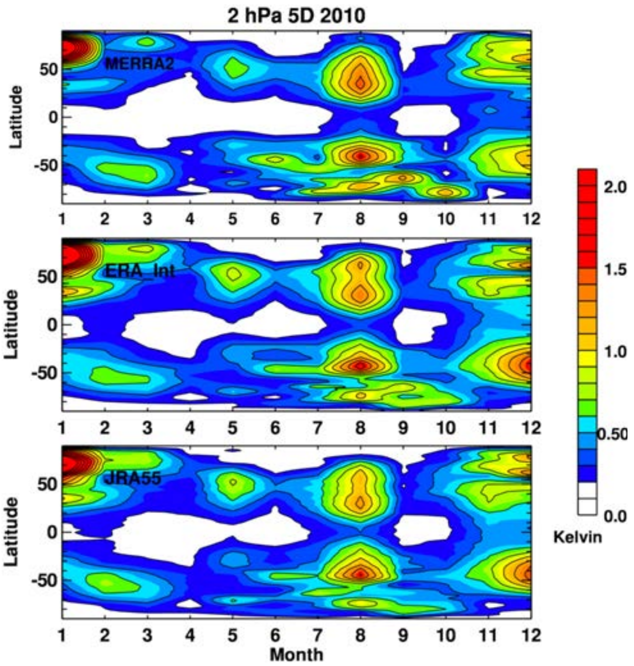


Figure 11.51: *Latitude-time plots of quasi-5-day wave amplitudes at 2 hPa during 2010 in MERRA-2 (top), ERA-Interim (middle) and JRA-55 (bottom). Contours every 0.2 K.*

system's higher top and inclusion of MLS mesospheric temperature observations, as discussed in the previous section. Between 10 hPa and 1 hPa, the amplitudes of the high-latitude QFDW signal in all three reanalyses are in good agreement. As discussed below, these two distinct patterns in the QFDW structure shown in **Figure 11.50** suggest that different processes may be involved in producing the QFDW signal within different latitude regions.

During August 2010 (**Figure 11.50**, right column), all three reanalyses again show very good qualitative and quantitative agreement between 10 hPa and 1 hPa. The dominant pattern is a hemispherically symmetric feature with peak amplitudes of ~ 2 K near 2 hPa between 20 ° and 50 ° latitude that is qualitatively consistent with the expected structure of the first normal mode. The amplitude of this hemispherically symmetric feature is roughly twice as large as a similar pattern seen in the upper stratosphere between 30 ° - 50 ° latitude during January (**Figure 11.50**, left column). In contrast to the very strong 2 - 3 K QFDW signal seen at high Northern latitudes in January, the corresponding high latitude feature in the SH during August is a much weaker ~1 K signal over a narrower latitude region and smaller altitude range. In addition, while in January (left panels) the QFDW amplitudes largely maximize at high northern (winter) latitudes, in the SH winter (right panels) there is a double maximum, with large amplitudes poleward of the jet core that are distinct from the maximum in mid-latitudes. MERRA-2 and ERA-Interim results both show a separate maximum in the QFDW signal in the lower mesosphere from 0.3 - 0.1 hPa. There is also evidence of a separate QFDW amplitude maximum in the NH extratropical lower mesosphere in all three reanalyses from 40 ° - 70 ° N.

The seasonal variation of the monthly mean QFDW amplitudes at 2 hPa during 2010 is shown in **Figure 11.51**. This pressure level is near the peak of the upper stratospheric, hemispherically symmetric feature seen in all three reanalyses (**Figure 11.50**). The largest QFDW amplitudes (> 2 K) are seen consistently at high Northern latitudes in January 2010 among all three data sets. During August 2010, there is a broad region of QFDW amplitudes ranging from 1 - 1.5 K between 30 ° N and 60 ° N, and a somewhat narrower region of slightly larger magnitude from 30 ° - 50 ° S that is sharply cut off poleward of 50 ° S. During November and December 2010, there is evidence in all three data sets of increasing QFDW amplitudes from 20 ° - 60 ° latitude in each hemisphere, with larger amplitudes and more coherent latitude structure in the SH. The evolution of the QFDW in 2010 is consistent with the seasonal cycle observed during 2002 - 2016 using SABER temperatures (*Huang et al.*, 2017). Their results show a consistent peak near the stratopause of 1 - 4 K from 30 ° - 50 ° N in December and January and from 30 ° - 50 ° S from June through September. The symmetry shown across the Equator in all three reanalyses during August in **Figure 11.50** is consistent with a QFDW signal related to the hemispherically symmetric normal mode (e.g., *Lieberman et al.*, 2003).

The very good agreement seen in the latitude, altitude, and seasonal variations in QFDW amplitudes for 2010 among the MERRA-2, ERA-Interim, and JRA-55 reanalyses suggests that much of the origin and propagation of the QFDW largely involves stratospheric processes that are well represented in these reanalyses. The two distinct patterns of QFDW amplitudes seen here further suggest that different processes are affecting the development and propagation of the QFDW. The first process, related to the apparent hemispherically symmetric mode, could be related to the interaction of the 5-day normal mode and tropospheric processes such as convective latent heat release (*Garcia and Salby*, 1987). The second process, related to the high Northern latitude maximum in January, could be related to growth through baroclinic/barotropic instability, leading to what is commonly referred to as the 6.5-day wave in the mesosphere and lower thermosphere (*Lieberman et al.*, 2003; *Talaat et al.*, 2002). Subsequent investigations of the QFDW in these and other reanalysis products over longer time periods, focusing on the periodicities of QFDW signals in the mid-latitude and polar latitudes would be helpful to further elucidate possible mechanisms for the origin and propagation of this signal throughout the stratosphere and mesosphere.

11.6 Summary, key findings, and recommendations

In summary, differences in the USLM among reanalyses are smaller in recent years (*vs*. in the 1980s and 1990s), increase with altitude, and increase nearer to the Equator. Improvements in observational data and in data assimilation explain much of the increasing agreement since 1998. The differences increase with altitude because of differences in model top altitude, the characteristics of the sponge layers near the model tops, and the differences in the gravity wave drag parameterizations in this region. The tropical USLM provides a particular challenge because the sparse observations leave key dynamical phenomena in this region weakly constrained in the models. As a result, large differences in USLM features can arise among the different reanalysis systems due to the different physical parameterizations in their respective model components. The intercomparisons presented in this chapter demonstrate that, while no one reanalysis system is clearly better in representing all aspects of the USLM, higher-top systems such as MERRA, MERRA-2, and ERA5 are essential for capturing the mesospheric circulation features such as the SAO and the QTDW, for example. However, we recommend that researchers interested in exploring a particular phenomenon within the USLM see the appropriate section of this chapter before choosing any one reanalysis for use in such research.

We also emphasize here that it is critical to compare reanalyses to independent observations (*e.g.*, radar winds, rocket winds, SABER temperatures). This process is imperative for data users to determine the optimum reanalysis dataset to use and to quantify the differences between a reanalysis dataset and observations prior to each scientific study. To the extent that such independent observations are lacking, we encourage additional observational campaigns to help establish which reanalyses perform best - as well as operational observational platforms to help tether models to observations within the reanalyses. With no firm plans to replace the aging satellites that are currently relied upon for temperature and constituent observations of the middle atmosphere, reanalysis systems could soon lack the key measurements needed to constrain models in the USLM region. Thus, it is imperative that plans are formulated and are executed to continue space-based global observations of the middle atmosphere to ensure the future of accurate simulation of middle atmosphere processes known to impact tropospheric weather forecasting.

Key findings

- Differences among the reanalyses 1) decrease with time due to improvements in assimilated observational data, 2) increase with altitude due to differences in model top, sponge layers, and gravity wave drag treatments, and 3) increase nearer the Equator where sparse observations leave key dynamical phenomena largely unconstrained.
- Although no single reanalysis system is clearly better in representing all aspects of the USLM, higher-top systems such as MERRA and MERRA-2 are essential for capturing mesospheric circulation features such as the SAO and the QTDW.
- Different satellite data assimilated into reanalyses as a function of time introduces discontinuities in both basic state variables and higher order diagnostics and this precludes trend studies based on a single reanalysis system.
- Differences in temperature among the reanalyses increase with height into the mesosphere at all latitudes. Likewise the inter-reanalysis differences in zonal wind increase with height especially in the equatorial region.
- Seasonal mean temperature differences (defined here to be with respect to MERRA) are larger in older reanalyses (ERA-40 and JRA-25) and smaller in newer reanalyses (MERRA-2, ERA-Interim, and JRA-55).
- Westerly and easterly jets in the winter and summer stratosphere, respectively, are well reproduced in MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2.
- The descending branch of the residual circulation in the winter stratosphere is strongest in MERRA, consistent with results prepared for *Chapter 5* (not shown; *Thomas Birner, personal communication*, 2021).
- There are anomalous vertical temperature gradients around 3 hPa in JRA-25 that lead to anomalous flow in the winter stratosphere and these are not observed in the other reanalyses.
- Noisy meridional and vertical winds in ERA-40 can cause larger dispersion of air parcels, which leads to "younger" age of air values and a weaker subtropical barrier in the stratosphere.
- Throughout the year, MERRA-2 has weaker cross-equatorial flow, a weaker middle-atmosphere Hadley circulation, and a westerly bias in the tropical USLM compared to ERA-Interim, JRA-55, and MERRA.

- Signatures of long-term variability due to the ENSO, the QBO, the 11-year solar cycle, and volcanic eruptions are shown in JRA-55, MERRA-2 and ERA-Interim; there are substantial differences among the reanalyses in the USLM, especially at equatorial latitudes.

- The mean SAO amplitude is reasonable in ERA-I, JRA-55, MERRA and MERRA-2; comparison between JRA-55 and JRA-55C highlights the inability for the free-running model to capture the SAO and the crucial role of assimilating satellite temperatures into this reanalysis system in order to accurately represent the SAO.

- The spatial patterns and magnitudes of inertial instability frequency are in good agreement among MERRA, MERRA-2, ERA-Interim, and JRA-55.

- MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 all capture multi-year winter mean polar vortex characteristics in both hemispheres; CFSR wintertime vortex frequencies are 10 - 20 % lower than the other four reanalyses in the 50 ° to 70 ° latitude bands in both hemispheres.

- MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 sufficiently capture the multi-year mean seasonal evolution of the polar vortex at the stratopause during 2005 - 2015; interannual variability is not assessed here.

- Quasi-stationary PW-1 amplitudes show remarkable agreement among the reanalyses and with MLS observations in the extratropics during winter; larger differences are seen at lower latitudes and during the summer.

- ES events are generally unconstrained by observations (with the exception of MERRA-2 which assimilates temperatures from Aura MLS after 2004). Their representation in reanalyses depends strongly on the nature of the sponge layer in the underlying forecast model of each reanalysis, and thus reanalyses cannot be regarded as trustworthy to study these phenomena.

- While reanalyses reproduce the global patterns in the diurnal and semi-diurnal migrating tides, their amplitudes are underestimated by 20 - 50 % compared to SABER.

- The representation of the quasi-2-day wave is qualitatively similar in MERRA-2, ERA-Interim, and JRA-55, but there are 50 % differences in amplitude.

- There is excellent agreement in the representation of the quasi-5-day wave among the MERRA-2, ERA-Interim, and JRA-55 reanalyses suggesting that much of its origin and propagation involves stratospheric processes that are well represented in these systems.

Recommendations

- Scientific studies using reanalyses in the USLM should make every effort to also include comparisons with independent observations.

- Large discontinuities that occur due to differences in the data assimilation process preclude trend studies based on a single reanalysis system.

- There are large temperature and wind differences among the reanalyses in the tropical USLM. Using two or more reanalysis datasets to study phenomena (*e.g.*, the SAO, the diurnal tide) in this region of the atmosphere is recommended to increase confidence.

- There are large uncertainties in MERRA-2 zonal winds in the Tropics prior to 1998 as it shows westerly biases in excess of 10 m s^{-1} compared to MERRA, ERA-Interim, and JRA-55 between 10 hPa and 1 hPa.

- There are large uncertainties in "older" reanalysis datasets in the USLM; the meridional circulation in the stratosphere and mesosphere is more realistic in MERRA-2, ERA-Interim, and JRA-55 than in MERRA, ERA-40, and JRA-25.

- Both Eulerian-mean and residual-mean meridional flows in ERA-40 are noisier than those in the other reanalyses, thus, science studies based on ERA-40 residual circulation velocities would likely generate noisier results.

- JRA-55C is not suitable for studies of the SAO.

- Low polar vortex frequency biases in CFSR/CFSv2 (due to high polar temperatures and a weak polar night jet) render this reanalysis dataset less suitable for polar vortex studies compared to MERRA, MERRA-2, ERA-Interim, or JRA-55.

- MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 are all suitable to study quasi-stationary PW-1 patterns in the winter extratropics but care should be exercised if the focus is in the subtropics or during the summer.

- Reanalyses should not be relied upon for studying ES events. Even for MERRA-2, the underlying forecast model does not capture the evolution of ES events correctly and so derived quantities (other than temperatures that are directly assimilated) should be treated with caution.

- Older reanalyses such as ERA-40 or JRA-25 are not suitable for tidal studies.

- Tidal results should not be extrapolated from one year to another as the representation of tides is sensitive to the satellite data assimilation.

- There are large uncertainties in using reanalysis data to study 5-day and 2-day wave normal modes; different reanalyses may yield different results.

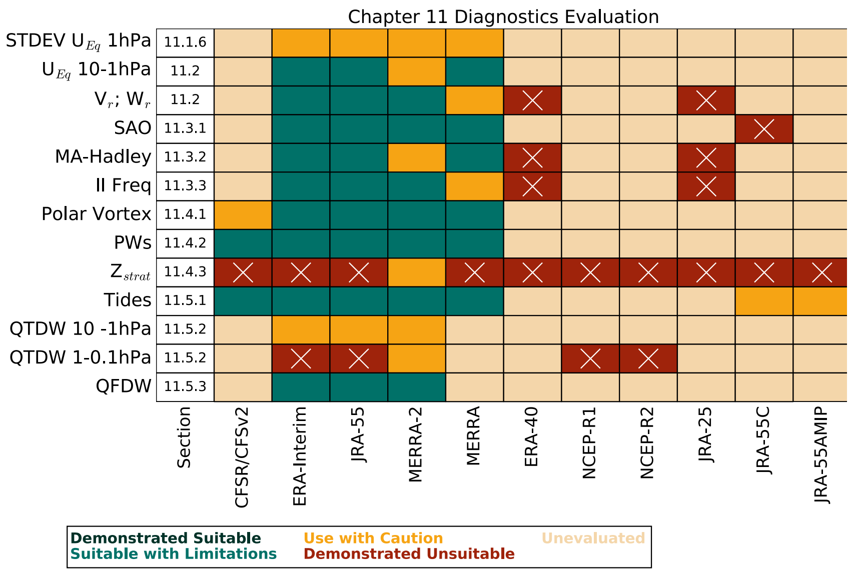


Figure 11.52: *Evaluation table of diagnostics relevant to Chapter 11 topics, listed along the y-axis. Different reanalyses are listed along the x-axis. The corresponding chapter and section numbers are given in the far left column. STDEV is the standard deviation, U_{Eq} is the zonal wind at the Equator, V_r and W_r are residual circulation meridional and vertical velocities, respectively, SAO is the Semi-Annual Oscillation, MA-Hadley is the middle-atmosphere Hadley circulation, II Freq is the inertial instability frequency of occurrence, PWs is planetary waves, Z_{strat} is the height of the stratopause with emphasis on elevated stratopause events, QTDW is the quasi-2-day wave, and QFDW is the quasi-5-day wave.*

Data availability

All of the reanalysis data included in this chapter are publicly accessible. JRA-55, ERA-Interim, and CFSR data are available at the Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory at **https://rda.ucar.edu/**. ERA-Interim data are also available at **http://apps.ecmwf.int/datasets/**. CFSR/CFSv2 data were also made available by Sean Davis at **ftpshare.al.noaa.gov**. The processing of CFSR/CFsv2 output was funded by the NOAA HPC grant “Climate Forecast System Reanalysis products for reanalysis validation and intercomparisons” to NOAA ESRL CSD with the bulk of the work performed by Sean Davis, Jeremiah Sjoberg and H. Leroy Miller. MERRA and MERRA-2 data are available at the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) at **https://gmao.gsfc.nasa.gov/reanalysis/**. TS analyzed diurnal monthly reanalysis data provided under the framework of the Data Integration and Analysis System (DIAS) funded by the Japan Ministry of Education, Culture, Sports, Science and Technology (MEXT). MLS v4.2 data are available from the NASA Goddard Space Flight Center for Earth Sciences DISC at **https://mls.jpl.nasa.gov**. SABER data are available from **http://saber.gats-inc.com/**.

Acknowledgements

VLH acknowledges support from NASA grants NNX-14AH54G, NNX17AB80G, 80NSSC181046, and 80NS-SC20K0628, and travel support from the Stratospheric Processes and their Role in Climate (SPARC) Reanalysis Intercomparison Project (S-RIP). JAF acknowledges support from NASA Small Explorer Program contract NAS5-03132 and NASA HGI grant 80NSSC18K0051. MF acknowledges support from the Japan Society for the Promotion of Science (JSPS) through Grants-in-Aid for Scientific Research (16K05548). LJG was supported by the U.K. Natural Environment Research Council (NERC) through the National Centre for Atmospheric Science (NCAS). TH was supported by JSPS KAKENHI Grant Number JP16H04052. PH acknowledges support from the European Research Council through the ACCI project (Grant 267760) and the NASA GNSS Remote Sensing Science Team Grant NNX16AK37G. MH was supported by NSF grant AGS-1555851. GLM was supported by the NASA MeaSUREs MUSTARD project via a JPL subcontract (#1483517). YK was supported by JSPS KAKENHI Grant Number JP15KK0178 and 18H01286. YK also acknowledges Anne Smith for her cubic spline codes used in the calculation of MLS gradient winds. JPM acknowledges support from NASA LWS award NN-H13AV95I. YOR was supported by the Research Council of Norway grant SOLENA #255276. TS acknowledges support from the JSPS through Grants-in-Aid for Scientific Research (15K17761). YT was supported by JSPS KAKENHI grant number JP15HP8027. The authors acknowledge useful conversations with Jonathon Wright, Sean Davis, Craig Long, Krzysztof Wargan, Rossana Dragani, Yayoi Harada, and the S-RIP Team.

References

Abalos, M., B. Legras, F. Ploeger, and W. J. Randel, 2015: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979 - 2012. *J. Geophys. Res. Atmos.*, **120**, 7534 - 7554, doi: 10.1002/2015JD023182.

Allen, D.R., *et al.*, 2000: Antarctic polar descent and planetary wave activity observed in ISAMS CO from April to July 1992. *Geophys. Res. Lett.*, **27**, 665 - 668, doi: 10.1029/1999GL010888.

Andrews, D.G., J.R. Holton, and C.B. Leovy, 1987: Middle Atmosphere Dynamics. *Academic Press*, San Diego, xi + 489 pp., doi: 10.1002/qj.49711548612.

Anstey, J.A., and T.G. Shepherd, 2014: High latitude influence of the quasi-biennial oscillation. *Q. J. Roy. Met. Soc.*, **140**, 1 - 21, doi: 10.1002/qj.2132.

Baldwin, M., et al., 2001: The quasi biennial oscillation. *Rev. Geophys.*, **39**, 179 - 229, doi: 10.1029/1999RG000073.

Baldwin, M.P., and T.J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581 - 584, doi: 10.1126/science.1063315.

Baker, D., *et al.*, 2019: The Scientific Foundation of Space Weather. 587 pp, *Springer Netherlands*, Dordrecht, Netherlands, **https://www.springer.com/gp/book/9789402415872**.

Barnett, J.J., and K. Labitzke, 1990: Climatological distribution of planetary waves in the middle atmosphere. *Adv. Space Res.*, **10**, 63 - 91, doi: 10.1016/0273-1177(90)90387-F.

Becker E., and S. L. Vadas, 2018: Secondary gravity waves in the winter mesosphere: Results from a high-resolution global circulation model. *J. Geophys. Res.*, **123**, 2605 - 2627, doi: 10.1002/2017JD027460.

Beig, G., J. Scheer, M.G. Mlynczak, and P. Keckhut, 2008: Overview of the temperature response in the mesosphere and lower thermosphere to solar activity. *Rev. Geophys.*, **46**, RG3002, doi: 10.1029/2007RG000236.

Bosilovich, M., *et al.*, 2015: MERRA-2: Initial evaluation of the climate. *NASA Tech. Rep. Series on Global Modeling and Data Assimilation*, *NASA/TM–2015-104606*, Vol. **43**, 139 pp.

Burks, D., and C. Leovy, 1986: Planetary waves near the mesospheric easterly jet. *Geophys. Res. Lett.*, **13**, 193 - 196, doi: 10.1029/GL013i003p00193.

Butler, A.H., J.P. Sjoberg, D.J. Seidel, and K.H. Rosenlof, 2017: A sudden stratospheric warming compendium. *Earth Syst. Sci. Data*, **9**, 63 - 76, doi: 10.5194/essd-9-63-2017.

Chandran, A., *et al.*, 2013: A climatology of elevated stratopause events in the whole atmosphere community climate model. *J. Geophys. Res. Atmos.*, **118**, 1234 - 1246, doi: 10.1002/jgrd.50123.

Chang, L.C., S.E. Palo, and H.-L. Liu, 2011: Short-term variability in the migrating diurnal tide caused by interactions with the quasi 2-day wave. *J. Geophys. Res.*, **116**, D12112, doi:10.1029/2010JD014996.

Charlton, A.J., and L.M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I. Climatology and modeling benchmarks. *J. Climate*, **20**, 449 - 469, doi: 10.1175/JCLI3994.1.

Charney, J.G., and P.G. Drazin, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83 - 109, doi: 10.1029/JZ066i001p00083.

Chapman, S., and R.S. Lindzen, 1970: Atmospheric Tides, *D. Reidel Publishing Company*, Dordrecht, Holland, 200 pp, doi: 10.1007/978-94-010-3399-2.

Chun, H.-Y. and J.-J. Baik, 1998: Momentum flux by thermally induced internal gravity waves and its approximation for large-scale models. *J. Atmos. Sci.*, **55**, 3299 - 3310, doi: 10.1175/1520-0469(1998)055<3299:MFBTII>2.0.CO;2.

Coy, L., 1979: A possible 2-day oscillation near the tropical stratopause. *J. Atmos. Sci.*, **36**, 1615 - 1618, doi: 10.1175/1520-0469(1979)036<1615:APDONT>2.0.CO;2.

Coy, L., *et al.*, 2016: Structure and Dynamics of the Quasi-Biennial Oscillation in MERRA-2. *J. Climate*, **29**, 5339 - 5354, doi: 10.1175/JCLI-D-15-0809.1.

Crooks S.A., and L.J. Gray, 2005: Characterisation of the 11-year solar signal using a multiple regression analysis of the ERA-40 data set. *J. Climate*, **18**, 996 - 1015, doi: 10.1175/JCLI-3308.1.

Dai, A., and J. Wang, J., 1999: Diurnal and semidiurnal tides in global surface pressure fields. *J. Atmos. Sci.*, **56**, 3874 - 3891, doi: 10.1175/1520-0469(1999)056<3874:DASTIG>2.0.CO;2.

Das, S.S., *et al.*, 2016: Four-decadal climatological intercomparison of rocketsonde and radiosonde with different reanalysis data: results from Thumba equatorial Station. *Q. J. Roy. Meteorol. Soc.*, **142**, 91 - 101, doi: 10.1002/qj.2632.

Dee, D.P., *et al.*, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Met. Soc.*, **137**, 553 - 597, doi: 10.1002/qj.828.

Demirhan Bari, D., A. Gabriel, H. Kornich, and D.W.H. Peters, 2013: The effect of zonal asymmetries in the Brewer-Dobson circulation on ozone and water vapor distributions in the northern middle atmosphere. *J. Geophys. Res.*, **118**, 3447 - 3466, doi: 10.1029/2012JD017709.

Diallo, M., B. Legras, and A. Chédin, 2012: Age of stratospheric air in the ERA Interim. *Atmos. Chem. Phys.*, **12**, 12,133 - 12,154, doi: 10.5194/acp-12-12133-2012.

Dunkerton, T.J., 1981: On the inertial stability of the equatorial middle atmosphere. *J. Atmos. Sci.*, **38**, 2354 - 2364, doi: 10.1175/1520-0469(1981)038<2354:OTISFT>2.0.CO;2.

Dunkerton, T. J., 1989: Nonlinear Hadley circulation driven by asymmetric differential heating. *J. Atmos. Sci.*, **46**, 956 - 974, doi: 10.1175/1520-0469(1989)046<0956:NHCDBA>2.0.CO;2.

Dunkerton, T. J., 1991: Nonlinear propagation of zonal winds in an atmosphere with Newtonian cooling and equatorial wave driving. *J. Atmos. Sci.*, **48**, 236 - 263, doi: 10.1175/1520-0469(1991)048<0236:NPOZWI>2.0.CO;2.

Eluszkiewicz, J., *et al.*, 1996: Residual Circulation in the Stratosphere and Lower Mesosphere as Diagnosed from Microwave Limb Sounder Data. *J. Atmos. Sci.*, **53**, 217 - 240, doi: 10.1175/1520-0469(1996)053<0217:RCITSA>2.0.CO;2.

Forbes, J.M., and D. Wu, 2006: Solar tides as revealed by measurements of mesosphere temperature by the MLS experiment on UARS. *J. Atmos. Sci.*, **63**, 1776 - 1797, doi: 10.1175/JAS3724.1.

France, J.A., *et al.*, 2018: Local and remote planetary wave effects on polar mesospheric clouds in the Northern Hemisphere in 2014. *J. Geophys. Res. Atmos.*, **123**, 5149 - 5162, doi: 10.1029/2017JD028224.

Fritts, D.C., and M.J. Alexander, 2003: Gravity wave effects and dynamics in the middle atmosphere. *Rev. Geophysics*, **41**, 1003, doi: 10.1029/2001RG000106.

Fujiwara, M., *et al.,* 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp-17-1417-2017.

Gabriel, A., *et al.*, 2011: Zonal asymmetries in middle atmospheric ozone and water vapour derived from Odin satellite data 2001 - 2010. *Atmos. Chem. Phys.*, **11**, 9865 - 9885, doi: 10.5194/acp-11-9865-2011.

Garcia, R.R. and M.L. Salby, 1987: Transient Response to Localized Episodic Heating in the Tropics. Part II: Far-Field Behavior. *J. Atmos. Sci.*, **44**, 499 - 532, doi: 10.1175/1520-0469(1987)044<0499:TRTLEH>2.0.CO;2.

Garcia, R.R., and B.A. Boville, 1994: Downward control of the mean meridional circulation and temperature distribution of the polar winter stratosphere. *J. Atmos. Sci.*, **51**, 2238 - 2245, doi: 10.1175/1520-0469(1994)051<2238:COTMMC>2.0.CO;2.

Garcia, R.R., T.J. Dunkerton, R.S. Lieberman, and R.A. Vincent, 1997: Climatology of the semiannual oscillation of the tropical middle atmosphere. *J. Geophys. Res.*, **102**, D22, 26019 - 26032, doi: 10.1029/97JD00207.

Garcia, R.R., R. Lieberman, J.M. Russell, and M.G. Mlynczak, 2005: Large-scale waves in the mesosphere and lower thermosphere observed by SABER. *J. Atmos. Sci.*, **62**, 4384 - 4399, doi: 10.1175/JAS3612.1.

Garcia-Herrera, R., N. Calvo, R.R. Garcia, and M.A. Giorgetta, 2006: Propagation of ENSO temperature signals into the middle atmosphere: A comparison of two general circulation models and ERA-40 reanalysis data. *J. Geophys. Res.*, **111**, D06101, doi:10.1029/2005JD006061.

Garfinkel, C.I., D.W. Waugh, and L.M. Polvani, 2015: Recent Hadley cell expansion: The role of internal atmospheric variability in reconciling modeled and observed trends. *Geophys. Res. Lett.*, **42**, 10,824 - 10,831, doi: 10.1002/2015GL066942.

Gelaro, R., *et al.*, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate*, **30**, 5419 - 5454, doi: 10.1175/JCLI-D-16-0758.1.

Gerrard A.J., *et al.*, 2002: Synoptic scale study of the Arctic polar vortex's influence on the middle atmosphere, 1, Observations. *J. Geophys. Res.*, **111**, D06107, doi: 10.1029/2001JD000681.

Global Modeling and Assimilation Office (GMAO), 2015a: MERRA-2 inst3_3d_asm_Np: 3d, 3-Hourly, Instantaneous, Pressure-Level, Assimilation, Assimilated Meteorological Fields V5.12.4. Greenbelt, MD, USA, *Goddard Earth Sciences Data and Information Services Center (GES DISC)*, Accessed in 2017, doi: 10.5067/QBZ6MG944HW0.

Global Modeling and Assimilation Office (GMAO), 2015b: MERRA-2 inst6_3d_ana_Nv: 3d, 6-Hourly, Instantaneous, Model-Level, Analysis, Analyzed Meteorological Fields V5.12.4. Greenbelt, MD, USA, *Goddard Earth Sciences Data and Information Services Center (GES DISC)*, Accessed in 2017, doi: 10.5067/IUUF4WB9FT4W.

Global Modeling and Assimilation Office (GMAO), 2015c: MERRA-2 inst3_3d_asm_Nv: 3d, 3-Hourly, Instantaneous, Model-Level, Assimilation, Assimilated Meteorological Fields V5.12.4. Greenbelt, MD, USA, *Goddard Earth Sciences Data and Information Services Center (GES DISC)*, Accessed in 2017, doi: 10.5067/WWQSXQ8IVFW8.

Global Modeling and Assimilation Office (GMAO), 2017: Use of MERRA-2 for Atmospheric Chemistry and Transport Studies. **https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/docs/ANAvsASM.pdf**.

Goncharenko, L.P., J.L. Chau, H.-L. Liu, and A.J. Coster, 2010: Unexpected connections between the stratosphere and ionosphere. *Geophys. Res. Lett.*, **37**, L10101, doi: 10.1029/2010GL043125.

Gu, S.-Y, *et al.*, 2013: Observations of quasi-two-day wave by TIMED/SABER and TIMED/TIDI. *J. Geophys. Res. Atmos.*, **118**, 1624 - 1639, doi: 10.1002/jgrd.50191.

Gray, L.J., *et al.*, 2010: Solar influences on climate. *Rev. Geophys.*, **48**, RG4001, doi: 10.1029/2009RG000282.

Hagan, M.E., J.M. Forbes, and F. Vial, 1995: On modeling migrating solar tides. *Geophys. Res. Lett.*, **22**, 893 - 896, doi: 10.1029/95GL00783.

Hagan, M.E., and J.M. Forbes, 2002: Migrating and nonmigrating diurnal tides in the middle and upper atmosphere excited by tropospheric latent heat release. *J. Geophys. Res.*, **107**, 4754, doi: 10.1029/2001JD001236.

Hagan, M.E., and J.M. Forbes, 2003: Migrating and nonmigrating semidiurnal tides in the upper atmosphere excited by tropospheric latent heat release. *J. Geophys. Res.*, **108**, doi: 10.1029/2002JA009466.

Hamilton, K., 1998: Dynamics of the tropical middle atmosphere: A tutorial review. *Atmosphere-Ocean*, **36**, 319 - 354, doi: 10.1080/07055900.1998.9649616.

Harris, T.J., 1994: A long-term study of the quasi-two-day wave in the middle atmosphere. *J. Atmos. Terr. Phys.*, **56**, 569 - 579, doi: 10.1016/0021-9169(94)90098-1.

Harvey, V.L., and M.H. Hitchman, 1996: A Climatology of the Aleutian High. *J. Atmos. Sci.*, **53**, 2088 - 2101, doi:10.1175/1520-0469(1996)053<2088:ACOTAH>2.0.CO;2.

Harvey, V.L., R.B. Pierce, T.D. Fairlie, and M.H. Hitchman, 2002: A climatology of stratospheric polar vortices and anticyclones. *J. Geophys. Res.*, **107**, 4442, doi: 10.1029/2001JD001471.

Harvey, V.L., et al., 2018: On the upward extension of the polar vortices into the mesosphere. *J. Geophys. Res. Atmos.*, **123**, 9171 - 9191, doi: 10.1029/2018JD028815.

Hayashi, Y., 1971: A generalized method of resolving disturbances into progressive and retrogressive waves by space Fourier and time cross-spectral analyses. *J. Meteorol. Soc. Jpn.*, **49**, 125 - 128, doi: 10.2151/jmsj1965.49.2_125.

Hersbach, H., and D. Dee, 2016: ERA5 reanalysis is in production. *ECMWF Newsletter*, **147**, 7.

Hirota, I., and J.J. Barnett, 1977: Planetary waves in the winter mesosphere - preliminary analysis of nimbus 6 PMR results. *Quart. J. Roy. Meteorol. Soc.*, **103**, 487 - 498. doi: 10.1002/qj.49710343709.

Hitchcock, P., and T.G. Shepherd, 2013: Zonal-mean dynamics of extended recoveries from stratospheric sudden warmings. *J. Atmos. Sci.*, **70**, 688 - 707, doi: 10.1175/JAS-D-12-0111.1.

Hitchman, M.H., and C.B. Leovy, 1986: Evolution of the zonal mean state in the equatorial middle atmosphere during October 1978 - May 1979. *J. Atmos. Sci.*, **43**, 3159 - 3176, doi: 10.1175/1520-0469(1986)043<3159:EOTZMS>2.0.CO;2.

Hitchman, M.H., J.C. Gille, C.D. Rodgers, and G. Brasseur, 1989: The separated polar winter stratopause: A gravity wave driven climatological feature. *J. Atmos. Sci.*, **46**, 410 - 422, doi: 10.1175/1520-0469(1989) 046<0410:TSPWSA>2.0.CO;2.

Holton, J., and H. Tan, 1982: The quasi-biennial oscillation in the Northern Hemisphere lower stratosphere. *J. Meteorol. Soc. Jpn.*, **60**, 140 - 148, doi: 10.2151/jmsj1965.60.1_140.

Hopkins, R.H., 1975: Evidence of polar-tropical coupling in upper stratosphere zonal wind anomalies. *J. Atmos. Sci.*, **32**, 712 - 719, doi: 10.1175/1520-0469(1975)032<0712:EOPTCI>2.0.CO;2.

Huang, Y.Y., *et al.*, 2017: Annual and interannual variations in global 6.5DWs from 20 to 110 km during 2002 - 2016 observed by TIMED/SABER. *J. Geophys. Res. Space Physics*, **122**, 8985 - 9002, doi: 10.1002/2017JA023886.

Ialongo, I., *et al.*, 2012: Ozone zonal asymmetry and planetary wave characterization during Antarctic spring. *Atmos. Chem. Phys.*, **12**, 2603 - 2614, doi: 10.5194/acp-12-2603-2012.

Iimura, H., *et al.*, 2015: Interhemispheric structure and variability of the 5-day planetary wave from meteor radar wind measurements. *Ann. Geophys.*, **33**, 1349 - 1359, doi: 10.5194/angeo-33-1349-2015.

Iwasaki, T., S. Yamada, and K. Tada, 1989a: A parameterization scheme of orographic gravity wave drag with two different vertical partitionings, Part I: Impact on medium range forecasts. *J. Meteorol. Soc. Jpn.*, **67**, 11 - 27, doi: 10.2151/jmsj1965.67.1_11.

Iwasaki, T., S. Yamada, and K. Tada, 1989b: A parameterization scheme of orographic gravity wave drag with two different vertical partitionings, Part II: Zonally averaged budget analyses based on transformed Eulerian-mean method. *J. Meteorol. Soc. Jpn.*, **67**, 29 - 41, doi: 10.2151/jmsj1965.67.1_29.

Iwasaki, T., H. Hamada, and K. Miyazaki, 2009: Comparisons of Brewer-Dobson circulations diagnosed from reanalysis. *J. Meteorol. Soc. Jpn.*, **87**, 997 - 1006, doi: 10.2151/jmsj.87.997.

Kawatani, Y., *et al.*, 2016: Representation of the tropical stratospheric zonal wind in global atmospheric reanalysis. *Atmos. Chem. Phys.*, **16**, 6681 - 6699, doi: 10.5194/acp-16-6681-2016.

Kawatani, Y., *et al.*, 2020: Representation of the equatorial stratopause semiannual oscillation in global atmospheric reanalyses. *Atmos. Chem. Phys.*, **20**, 9115 - 9133, doi: 10.5194/acp-20-9115-2020.

Kim, Y.-J. and A. Arakawa, 1995: Improvement of orographic gravity wave parameterization using a mesoscale gravity wave model. *J. Atmos. Sci.*, **83**, 1875 - 1902, doi: 10.1175/1520-0469(1995)052<1875:IOOGWP>2.0.CO;2.

Kishore Kumar, G., K. Kishore Kumar, G. Baumgarten, and G. Ramkumar, 2015: Validation of MERRA reanalysis upper-level winds over low latitudes with independent rocket sounding data. *J. Atmos. Sol. Terr. Phys.*, **123**, 48 - 54, doi: 10.1016/j.jastp.2014.12.001.

Knox, J.A., 2003: Inertial instability. In *Encyclopedia of the Atmospheric Sciences*, edited by J. R. Holton, J. Pyle, and J. A. Curry, pp. 1004 - 1013, Elsevier, New York.

Knox, J.A., and V.L. Harvey, 2005: Global climatology of inertial instability and Rossby wave breaking in the stratosphere. *J. Geophys. Res.*, **110**, D06108, doi: 10.1029/2004JD005068.

Kobayashi, S., *et al.*, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteorol. Soc. Jpn.*, **93**, 5 - 48, doi: 10.2151/jmsj.2015-001.

Kouker, W., and G. Brasseur, 1986: Transport of atmospheric tracers by planetary waves during a winter stratospheric warming event: A three-dimensional model simulation. *J. Geophys. Res.*, **91**, 13167 - 13185, doi: 10.1029/JD091iD12p13167.

Labitzke, K., 1972: Temperature changes in the mesosphere and stratosphere connected with circulation changes in winter. *J. Atmos. Sci.*, **29**, 756 - 766, doi: 10.1175/1520-0469(1972)029<0756:TCITMA>2.0.CO;2.

Lieberman, R.S., 1999: Eliassen-Palm fluxes of the 2-day wave. *J. Atmos. Sci.*, **56**, 2846 - 2861, doi: 10.1175/1520-0469(1999)056<2846:EPFOTD>2.0.CO;2.

Lieberman, R.S., *et al.*, 2003: The 6.5-day wave in the mesosphere and lower thermosphere: Evidence for baroclinic/barotropic instability. *J. Geophys. Res.*, **108**, 4640, doi: 10.1029/2002JD003349.

Lieberman, R.S., D.M. Riggin, and D.E. Siskind, 2013: Stationary waves in the wintertime mesosphere: Evidence for gravity wave filtering by stratospheric planetary waves. *J. Geophys. Res. Atmos.*, **118**, 3139 - 3149, doi: 10.1002/jgrd.50319.

Lilienthal, F., C. Jacobi, and C. Geibler, 2018: Forcing mechanisms of the terdiurnal tide. *Atmos. Chem. Phys.*, **18**, 15725 - 15742, doi: 10.5194/acp-18-15725-2018.

Limpasuvan, V., C.B. Leovy, and Y.J. Orsolini, 2000: Observed temperature two-day wave and its relatives near the stratopause. *J. Atmos. Sci.*, **57**, 1689 - 1701, doi: 10.1175/1520-0469(2000)057<1689:OTTDWA>2.0.CO;2.

Limpasuvan, V., and D.L. Wu, 2003: Two-day wave observations of UARS Microwave Limb Sounder mesospheric water vapor and temperature. *J. Geophys. Res.*, **108**, 4307, doi: 10.1029/2002JD002903.

Limpasuvan, V., *et al.*, 2012: The roles of planetary and gravity waves during a major stratospheric sudden warming as characterized in WACCM. *J. Atmos. Sol. Terr. Phys.*, **78–79**, 84-98, doi: 10.1016/j.jastp.2011.03.004.

Liu, H.L., and R.G. Roble, 2002: A study of a self-generated stratospheric sudden warming and its mesospheric-lower thermospheric impacts using the coupled TIME-GCM/CCM3. *J. Geophys. Res.*, **107**, doi: 10.1029/2001JD001533.

Liu, H., E. Doornbos, M. Yamamoto, and S. Tulasi Ram, 2011: Strong thermospheric cooling during the 2009 major stratosphere warming. *Geophys. Res. Lett.*, **38**, L12102, doi: 10.1029/2011GL047898.

Livesey, N.J., *et al.*, 2017: Earth Observing System (EOS) Aura Microwave Limb Sounder (MLS) version 4.2X Level 2 data quality and description document. *Rep. JPL D-33509* Rev. **C**, Jet Propul. Lab., Pasadena, Calif.

Long, C.S., *et al.*, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14593 - 14629, doi: 10.5194/acp-17-14593-2017.

Lott, F. and M.J. Miller, 1997: A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Q. J. Roy. Meteorol. Soc.*, **123**, 101 - 127, doi: 10.1002/qj.49712353704.

Madden, R. and P. Julian, 1972: Further evidence of global-scale 5-day pressure waves. *J. Atmos. Sci.*, **29**, 1464 - 1469, doi: 10.1175/15200469(1972)029<1464:FEOGSD>2.0.CO;2.

Manney, G.L., and T.R. Nathan, 1990: Barotropic Stability of Westward-Moving Waves in Realistic Stratospheric Zonal Flows. *J. Atmos. Sci.*, **47**, 775-794, doi: 10.1175/1520-0469(1990)047<0775:BSOWMW>2.0.CO;2.

Manney, G.L., *et al.*, 2008: The evolution of the stratopause during the 2006 major warming: Satellite data and assimilated meteorological analyses. *J. Geophys. Res.*, **113**, D11115, doi: 10.1029/2007JD009097.

Manney, G.L., *et al.*, 2009: Aura Microwave Limb Sounder observations of dynamics and transport during the record-breaking 2009 Arctic stratospheric major warming. *Geophys. Res. Lett.*, **36**, L12815, doi: 10.1029/2009GL038586.

McCormack, J.P., L. Coy, and K.W. Hoppel, 2009: Evolution of the quasi 2-day wave during January 2006. *J. Geophys. Res.*, **114**, D20115, doi: 10.1029/2009JD012239.

McCormack, J.P., L. Coy, and W. Singer, 2014: Intraseasonal and interannual variability of the quasi 2-day wave in the Northern Hemisphere summer mesosphere. *J. Geophys. Res. Atmos.*, **119**, 2928 - 2946, doi:10.1002/2013JD020199.

McFarlane, N.A., 1987: The effect of orographically excited gravity-wave drag on the circulation of the lower stratosphere and troposphere. *J. Atmos. Sci.*, **44**, 1775 - 1800, doi: 10.1175/1520-0469(1987)044<1775:TEOOEG>2.0.CO;2.

McLandress, C., *et al.*, 2006: Large-scale dynamics of the mesosphere and lower thermosphere: An analysis using the extended Canadian Middle Atmosphere Model. *J. Geophys. Res.*, **111**, D17111. doi: 10.1029/2005JD006776.

McLandress, C., *et al.*, 2013: Dynamical control of the mesosphere by orographic and nonorographic gravity wave drag during the extended northern winters of 2006 and 2009. *J. Atmos. Sci.*, **70**, 2152 - 2169, doi: 10.1175/JAS-D-12-0297.1.

McLandress, C., D.A. Plummer, and T.G. Shepherd, 2014: Technical Note: A simple procedure for removing temporal discontinuities in ERA-Interim upper stratospheric temperatures for use in nudged chemistry-climate model simulations. *Atmos. Chem. Phys.*, **14**, 1547 - 1555, doi: 10.5194/acp-14-1547-2014.

Mitchell, D.M., *et al.*, 2015: Signatures of naturally induced variability in the atmosphere using multiple reanalysis datasets. *Q. J. Roy. Met. Soc.*, **141**, 2011 - 2031, doi: 10.1002/qj.2492.

Miyoshi, Y., and T. Hirooka, 2003: Quasi-biennial variation of the 5-day wave in the stratosphere. *J. Geophys. Res.*, **108**, 4620, doi: 10.1029/2002JD003145.

Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA-2. *Geosci. Model Dev.*, **8**, 1339 - 1356, doi: 10.5194/gmd-8-1339-2015.

Monge-Sanz, B.M., M.P. Chipperfield, A.J. Simmons, and S.M. Uppala, 2007: Mean age of air and transport in a CTM: Comparison of different ECMWF analyses. *Geophys. Res. Lett.* **34**, L04801, doi: 10.1029/2006GL028515.

Muller, H.G., and L. Nelson, 1978: A traveling quasi 2-day wave in the meteor region. *J. Atmos. Terr. Phys.*, **40**, 761 - 766, doi: 10.1016/0021-9169(78)90136-8.

Nielsen, K., *et al.*, 2010: Seasonal variation of the quasi 5-day planetary wave: Causes and consequences for polar mesospheric cloud variability in 2007. *J. Geophys. Res.*, **115**, D18111, doi: 10.1029/2009JD012676.

Norton, W.A., and J. Thuburn, 1999: Sensitivity of mesospheric mean flow, planetary waves, and tides to strength of gravity wave drag. *J. Geophys. Res.*, **104**, 30,897 - 30,911, doi: 10.1029/1999JD900961.

Offermann, D., *et al.*, 2003: Zonal asymmetries in middle atmosphere temperatures. *Adv. Space. Res.*, **32**, 1771 - 1780, doi: 10.1016/S0273-1177(03)90475-3.

Onogi, K., *et al.*, 2007: The JRA-25 reanalysis. *J. Meteorol. Soc. Jpn.*, **85**, 369 - 432, doi: 10.2151/jmsj.85.369.

Orbe, C., *et al.*, 2017: Large-Scale Atmospheric Transport in GEOS Replay Simulations. *J. Adv. Mod. Earth Sys.*, **9**, 2545 - 2560, doi: 10.1002/2017MS001053.

Orr, A., *et al.*, 2010: Improved middle atmosphere climate and forecasts in the ECMWF model through a non-orographic gravity wave drag parametrization. *J. Climate*, **23**, 5905 - 5926, doi: 10.1175/2010JCLI3490.1.

Orsolini, Y.J., V. Limpasuvan, and C.B. Leovy, 1997: The tropical stratopause response in the UKMO stratospheric analyses: Evidence for a 2-day wave and inertial circulations. *Q. J. R. Meteorol. Soc.*, **123**, 1707 - 1724, doi: 10.1002/qj.49712354212.

Orsolini, Y.J., *et al.*, 2017: Modelling the descent of nitric oxide during the elevated stratopause event of January 2013. *J. Atmos. Sol. Terr. Phys.*, **155**, 50 - 61, doi: 10.1016/j.jastp.2017.01.006.

Pancheva, D.V., 2006: Quasi-2-day wave and tidal variability observed over Ascension Island during January/February 2003. *J. Atmos. T Sol. err. Phys.*, **68**, 390 - 407, doi: 10.1016/j.jastp.2005.02.028.

Pascoe, C.L., *et al.*, 2005: The quasi biennial oscillation: analysis using ERA-40 data. *J. Geophys. Res.*, **110**, D08105, doi: 10.1029/2004D004941.

Pedatella, N.M., *et al.*, 2014: The neutral dynamics during the 2009 sudden stratosphere warming simulated by different whole atmosphere models. *J. Geophys. Res. Space Physics*, **119**, 1306 - 1324, doi: 10.1002/2013JA019421.

Pedatella, N.M., *et al.*, 2018: How sudden stratospheric warming affects the whole atmosphere. *Eos*, **99**, doi: 10.1029/2018EO092441.

Pfister, L., 1985: Baroclinic instability of easterly jets with applications to the summer mesosphere. *J. Atmos. Sci.*, **42**, 313 - 330, doi: 10.1175/1520-0469(1985)042<0313:BIOEJW>2.0.CO;2.

Plumb, R.A., 1983: Baroclinic instability of the summer mesosphere: A mechanism for the quasi-two-day wave? *J. Atmos. Sci.*, **40**, 262 - 270, doi: 10.1175/1520-0469(1983)040<0262:BIOTSM>2.0.CO;2.

Plumb, A., and J. Eluszkiewicz, 1999: The Brewer–Dobson circulation: Dynamics of the tropical upwelling. *J. Atmos. Sci.*, **56**, 868 - 890, doi: 10.1175/1520-0469(1999)056,0868: TBDCDO.2.0.CO;2.

Polichtchouk, I., *et al.,* 2017: What influences the middle atmosphere circulation in the IFS? *ECMWF Technical Memorandum* No. **809**. doi: 10.21957/mfsnfv15o.

Randall, C.E., *et al.*, 2006: Enhanced NO_x in 2006 linked to strong upper stratospheric Arctic vortex. *Geophys. Res. Lett.*, **33**, L18811, doi: 10.1029/2006GL027160.

Randall, C.E., *et al.*, 2015: Simulation of energetic particle precipitation effects during the 2003–2004 Arctic winter. *J. Geophys. Res. Space Phys.*, **120**, 5035 - 5048, doi: 10.1002/2015JA021196.

Randel, W.J., 1994: Observations of the 2-day wave in NMC stratospheric analyses. *J. Atmos. Sci.*, **51**, 306 - 313, doi: 10.1175/1520-0469(1994)051<0306:OOTDWI>2.0.CO;2.

Remsberg, E.E., *et al.,* 2003: On the verification of the quality of SABER temperature, geopotential height, and wind fields by comparison with Met Office assimilated analyses. *J. Geophys. Res. Atmos.*, **108**, doi: 10.1029/2003JD003720.

Rienecker, M.M., *et al.*, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624 - 3648, doi: 10.1175/JCLI-D-11-00015.1.

Rodgers, C.D., and A.J. Prata, 1981: Evidence for a traveling two-day wave in the middle atmosphere. *J. Geophys. Res.*, **86**, 9661 - 9664, doi: 10.1029/JC086iC10p09661.

Rojas, M., and W. Norton, 2007: Amplification of the 2-day wave from mutual interaction of the global Rossby-gravity and local modes in the summer mesosphere. *J. Geophys. Res.*, **112**, D12114, doi: 10.1029/2006JD008084.

Saha, S., *et al.*, 2010: The NCEP climate forecast system reanalysis. *Bull. Am. Meteo. Soc.*, **91**, 1015 - 1057, doi: 10.1175/2010BAMS3001.1.

Saha, S., *et al.*, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185 - 2208, doi: 10.1175/JCLI-D-12-00823.1.

Sakazaki, T., et al., 2012: Diurnal tides from the troposphere to the lower mesosphere as deduced from TIMED/SABER satellite data and six global reanalysis data sets. *J. Geophys. Res.*, **117**, D13108, doi: 10.1029/2011JD017117.

Sakazaki, T., K. Sato, Y. Kawatani, and S. Watanabe, S., 2015: Three-dimensional structures of tropical nonmigrating tides in a high-vertical-resolution general circulation model. *J. Geophys. Res. Atmos.*, **120**, 1759 - 1775, doi: 10.1002/2014JD022464.

Sakazaki, T., M. Fujiwara, and M. Shiotani, M., 2018: Representation of solar tides in the stratosphere and lower mesosphere in state-of-the-art reanalyses and in satellite observations. *Atmos. Chem. Phys.*, **18**, 1437 - 1456, doi: 10.5194/acp-18-1437-2018.

Sassi, F., R.R. Garcia, B.A. Boville, and H. Liu, 2002: On temperature inversions and the mesospheric surf zone. *J. Geophys. Res.*, **107**, 4380, doi: 10.1029/2001JD001525.

Sato M., J. Hansen, M. McCormick, and J. Pollack, 1993: Stratospheric aerosol optical depths, 1850 - 1990. *J. Geophys. Res.*, **98**, 22987 - 22994, doi: 10.1029/93JD02553.

Sato, K., *et al.*, 2009: On the origins of mesospheric gravity waves. *Geophys. Res. Lett.*, **36**, L19801, doi: 10.1029/2009GL039908.

Sato, K., and S. Hirano, 2019: The Climatology of Brewer-Dobson Circulation and the Contribution of Gravity Waves. *Atmos. Chem. Phys.*, **19**, 4517 - 4539, doi: 10.5194/acp-19-4517-2019.

Semeniuk, K., and T.G. Shepherd, 2001a: The middle-atmosphere Hadley circulation and equatorial inertial adjustment. *J. Atmos. Sci.*, **58**, 3077 - 3096, doi: 10.1175/1520-0469(2001)058<3077:TMAHCA>2.0.CO;2.

Semeniuk, K., and T.G. Shepherd, 2001b: Mechanisms for tropical upwelling in the stratosphere. *J. Atmos. Sci.*, **58**, 3097 - 3115, doi: 10.1175/1520-0469(2001)058<3097:MFTUIT>2.0.CO;2.

Schoeberl, M.R., and D.L. Hartmann, 1991: The dynamics of the stratospheric polar vortex and its relation to springtime ozone depletions. *Science*, **251**, 46 - 52, doi: 10.1126/science.251.4989.46.

Schoeberl, M., A. Douglass, Z. Zhu, and S. Pawson, 2003: A comparison of the lower stratospheric age spectra derived from a general circulation model and two data assimilation systems. *J. Geophys. Res.*, **108**, 4113, doi: 10.1029/2002JD002652.

Shepherd, T.G., K. Semeniuk, and J.N. Koshyk, 1996: Sponge layer feedbacks in middle-atmosphere models. *J. Geophys. Res.*, **101**, 23447 - 23464, doi: 10.1029/96JD01994.

Shepherd, T.G., I. Polichtchouk, R.J. Hogan, and A.J. Simmons, 2018: Report on Stratosphere Task Force. *ECMWF Technical Memorandum* **824**.

Shiotani, M., and I. Hirota, 1985: Planetary wave-mean flow interaction in the stratosphere: a comparison between northern and southern hemispheres. *Q. J. Roy. Meteorol. Soc.*, **111**, 309 - 334, doi: 10.1002/qj.49711146804.

Simmons, A.J., *et al.*, 2014: Estimating low frequency variability and trends in atmospheric temperature using ERA Interim. *Q. J. Roy. Meteorol. Soc.*, **140**, 329 - 353, doi: 10.1002/qj.2317.

Siskind, D.E., L. Coy, and P. Espy, 2005: Observations of stratospheric warmings and mesospheric coolings by the TIMED SABER instrument. *Geophys. Res. Lett.*, **32**, L09804, doi: 10.1029/2005GL022399.

Siskind, D.E., *et al.*, 2007: On recent interannual variability of the Arctic winter mesosphere: Implications for tracer descent. *Geophys. Res. Lett.*, doi: 10.1029/2007GL029293.

Siskind, D.E. , *et al.*, 2010: Case studies of the mesospheric response to recent minor, major, and extended stratospheric warmings. *J. Geophys. Res.*, **115**, D00N03, doi: 10.1029/2010JD014114.

Siskind, D.E., and J.P. McCormack, 2014: Summer mesospheric warmings and the quasi 2-day wave. *Geophys. Res. Lett.*, **41**, 717 - 722, doi: 10.1002/2013GL058875.

Smith, A.K., 1996: Longitudinal variations in mesospheric winds: evidence for gravity wave filtering by planetary waves. *J. Atmos. Sci.*, **53**, 1156 - 1173, doi:10.1175/1520-0469(1996)053<1156:LVIMWE>2.0.CO;2.

Smith, A.K., 1997: Stationary planetary waves in upper mesospheric winds. *J. Atmos. Sci.*, **54**, 2129 - 2145, doi: 10.1175/1520-0469(1997)054<2129:SPWIUM>2.0.CO;2.

Smith, A.K., R.R. Garcia, A.C. Moss, and N.J. Mitchell, 2017: The semiannual oscillation of the tropical zonal wind in the middle atmosphere derived from satellite geopotential height retrievals. *J. Atmos. Sci.*, **74**, 2413 - 2425, doi: 10.1175/JAS-D-17-0067.1.

Smith, T.M., and R.W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854 - 1997). *J. Climate*, **16**, 1495 - 1510, doi: 10.1175/1520-0442-16.10.1495.

Talaat, E.R., J.-H. Yee, and X. Zhu, 2001: Observations of the 6.5-day wave in the mesosphere and lower thermosphere. *J. Geophys. Res.*, **106**, 20715 - 20723, doi: 10.1029/2001JD900227.

Talaat, E.R., J.-H. Yee, and X. Zhu, 2002: The 6.5-day wave in the tropical stratosphere and mesosphere. *J. Geophys. Res.*, **107**, 4133, doi: 10.1029/2001JD000822.

Tomikawa, Y., *et al.*, 2008: Wintertime temperature maximum at the subtropical stratopause in a T213L256 GCM. J. *Geophys. Res.*, **113**, D17117, doi:10.1029/2008JD009786.

Tomikawa, Y., *et al.*, 2012: Growth of planetary waves and the formation of an elevated stratopause after a major stratospheric sudden warming in a T213L256 GCM. *J. Geophys. Res.*, **117**, D16101, doi: 10.1029/2011JD017243.

Tunbridge, V.M., D.J. Sandford, and N.J. Mitchell, 2011: Zonal wave numbers of the summertime 2-day planetary wave observed in the mesosphere by EOS Aura Microwave Limb Sounder. *J. Geophys. Res.*, **116**, D11103, doi: 10.1029/2010JD014567.

Tung, K.K., and J.S. Kinnersley, 2001: Mechanisms by which extratropical wave forcing in the winter stratosphere induces upwelling in the summer hemisphere. *J. Geophys. Res.*, **106**, 22781 - 22791, doi: 10.1029/2001JD900228.

Uppala, S.M., *et al.*, 2005: The ERA-40 reanalysis. *Q. J. Roy. Meteorol. Soc.*, **131**, 2961 - 3012, doi: 10.1256/qj.04.176.

Walterscheid, R.L., G.G. Sivjee, and R.G. Roble, 2000: Mesospheric and lower thermospheric manifestations of a stratospheric warming event over Eureka, Canada (80°N). *Geophys. Res. Lett.*, **27**, 2897 - 2900, doi: 10.1029/2000GL003768.

Wang, Y.M., J. Lean, and N. Sheeley Jr., 2005: Modeling the Sun's magnetic field and irradiance since 1713. *J. Astrophys.*, **625**, 522 - 538, doi: 10.1086/429689.

Watanabe, S., *et al.*, 2008: General aspects of a T213L256 middle atmosphere general circulation model. *J. Geophys. Res.*, **113**, D12110, doi: 10.1029/2008JD010026.

Waters, J.W., *et al.*, 2006: The Earth Observing System microwave limb sounder (EOS MLS) on the Aura satellite. *IEEE Trans. Geosci. Remote Sens.*, **44**, 1075 - 1092, doi: 10.1109/TGRS.2006.873771.

Waugh, D.W., and L.M. Polvani, 2010: Stratospheric polar vortices, in The Stratosphere: Dynamics, Transport, and Chemistry. *Geophys. Monogr. Ser.*, **190,** edited by L. M. Polvani, A. H. Sobel, and D. W. Waugh, pp. 43–57, AGU, Washington, DC, doi:10.1029/2009GM000887.

Wu, D.L., *et al.*, 1993: Observations of the quasi 2-day wave from the High Resolution Doppler Imager on UARS. *Geophys. Res. Lett.*, **20**, 2853 - 2856, doi: 10.1029/93GL03008.

Yue, J., W. Wang, A.D. Richmond, and H.-L. Liu, 2012: Quasi-two-day wave coupling of the mesosphere and lower thermosphere-ionosphere in the TIME-GCM: Two-day oscillations in the ionosphere. *J. Geophys. Res.*, **117**, A07305, doi: 10.1029/2012JA017815.

Zhang, X., *et al.*, 2006: Monthly tidal temperatures 20 - 120 km from TIMED/SABER. *J. Geophys. Res.*, **111**, A10S08, doi: 10.1029/2005JA011504.

Zou, C.-Z., *et al.*, 2014: Recalibration and merging of SSU observations for stratospheric temperature trend studies. *J. Geophys. Res. Atmos.*, **119**, 13,180 - 13,205, doi: 10.1002/2014JD021603.

Major abbreviations and terms

2DFFT	two-dimensional fast Fourier transform
AMIP	Atmospheric Model Intercomparison Project
ATOVS	Advanced TIROS Operational Vertical Sounder
CFSR	Climate Forecast System Reanalysis of the NCEP
CFSv2	Climate Forecast System version 2
DJF	December-January-February
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño Southern Oscillation
EOF	Empirical Orthogonal Function
EOS	Earth Observing System
ERA-40	ECMWF 40-year reanalysis
ERA-Interim	ECMWF interim reanalysis
ERA5	the fifth major global reanalysis produced by ECMWF
ERSST	Extended Reconstructed Sea Surface Temperature
IFS	Integrated Forecast System
JJA	June-July-August
JRA-25	Japanese 25-year Reanalysis
JRA-55	Japanese 55-year Reanalysis
JRA-55AMIP	Japanese 55-year Reanalysis based on AMIP-type simulations
JRA-55C	Japanese 55-year Reanalysis assimilating Conventional observations only
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2
MLS	Microwave Limb Sounder
MLT	Mesosphere Lower Thermosphere
NCEP	National Centers for Environmental Prediction
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
NRLSSI	Naval Research Laboratory model for Solar Spectral Irradiance
PMC	Polar Mesospheric Cloud
PW-1 / PW-2	Planetary Wave number 1/Planetary Wave number 2
QBO	Quasi-Biennial Oscillation
QFDW	Quasi-Five-Day Wave
QTDW	Quasi-Two-Day Wave
SABER	Sounding of the Atmosphere using Broadband Emission Radiometry
SAO	Semi-Annual Oscillation
SH	Southern Hemisphere
S-RIP	SPARC Reanalysis Intercomparison Project
SSU	Stratospheric Sounding Unit
SSW	Sudden Stratospheric Warming
TIMED	Thermosphere • Ionosphere • Mesosphere • Energetics and Dynamics
TIROS	Television and InfraRed Observation Satellite
TOVS	TIROS Operational Vertical Sounder
USLM	Upper Stratosphere Lower Mesosphere

Chapter 12: Synthesis Summary

Chapter authors

Masatomo Fujiwara	Faculty of Environmental Earth Science, Hokkaido University	Japan
Gloria L. Manney	(1) NorthWest Research Associates (2) Department of Physics, New Mexico Institute of Mining and Technology	USA
Lesley J. Gray	(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford (2) NERC National Centre for Atmospheric Science	UK
Jonathon S. Wright	Department of Earth System Science, Tsinghua University	China
James Anstey	Canadian Centre for Climate Modelling and Analysis Environment and Climate Change Canada	Canada
Thomas Birner	(1) Meteorological Institute, Ludwig-Maximilians-University Munich (2) Institute of Atmospheric Physics, German Aerospace Center (DLR Oberpfaffenhofen) *previously at: Department of Atmospheric Science, Colorado State University, USA*	Germany
Sean Davis	Chemical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Edwin P. Gerber	Courant Institute of Mathematical Sciences, New York University	USA
V. Lynn Harvey	Laboratory for Atmospheric and Space Physics, University of Colorado	USA
Michaela I. Hegglin	Department of Meteorology, University of Reading	UK
Cameron R. Homeyer	School of Meteorology, University of Oklahoma	USA
John A. Knox	Department of Geography, University of Georgia	USA
Kirstin Krüger	Department of Geosciences, University of Oslo	Norway
Alyn Lambert	Jet Propulsion Laboratory, California Institute of Technology	USA
Craig S. Long	Climate Prediction Center, National Oceanic and Atmospheric Administration *(retired)*	USA
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan
Beatriz M. Monge-Sanz	Atmospheric, Oceanic and Planetary Physics, University of Oxford	UK
Michelle L. Santee	Jet Propulsion Laboratory, California Institute of Technology	USA
Susann Tegtmeier	Institute of Space and Atmospheric Studies, University of Saskatchewan *previously at: GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany*	Canada
Simon Chabrillat	Royal Belgian Institute for Space Aeronomy	Belgium
Zachary D. Lawrence	(1) Cooperative Institute for Research in Environmental Sciences, Univ. of Colorado (2) NOAA Physical Sciences Laboratory (PSL) (3) NorthWest Research Associates	USA
David G. H. Tan	European Centre for Medium-Range Weather Forecasts *(retired)*	UK
David R. Jackson	Met Office	UK
Saroja Polavarapu	Climate Research Division, Environment and Climate Change Canada	Canada
Gilbert P. Compo	(1) Cooperative Institute for Research in Environmental Sciences, University of Colorado; (2) Physical Sciences Laboratory, National Oceanic and Atmospheric Administration	USA
Rossana Dragani	European Centre for Medium-Range Weather Forecasts	UK
Wesley Ebisuzaki	National Oceanic and Atmospheric Administration	USA
Yayoi Harada	Japan Meteorological Agency	Japan
Chiaki Kobayashi	Japan Meteorological Agency	Japan
Krzysztof Wargan	(1) National Aeronautics and Space Administration (2) Science Systems and Applications, Inc.	USA
Jeffrey S. Whitaker	National Oceanic and Atmospheric Administration	USA

Content

12.1 Key findings and recommendations by chapter 585
12.1.1 Chapter 3: Overview of Temperature and Winds 585
12.1.2 Chapter 4: Overview of Ozone and Water Vapour 587
12.1.3 Chapter 5: Brewer–Dobson Circulation 588
12.1.4 Chapter 6: Extratropical Stratosphere–Troposphere Coupling 591
12.1.5 Chapter 7: Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS)....................592
12.1.6 Chapter 8: Tropical Tropopause Layer (TTL) 594
12.1.7 Chapter 9: Quasi-Biennial Oscillation (QBO) 597
12.1.8 Chapter 10: Polar Processes 599
12.1.9 Chapter 11: Upper Stratosphere and Lower Mesosphere 601
12.2 Overall findings and reanalysis user recommendations 604
12.3 Recommendations for improving reanalyses and their evaluation 604
12.3.1 S-RIP survey results and related product needs from S-RIP studies 605
12.3.2 Data access issues 607
12.3.3 Documentation issues 607
12.4 Prospects for the future 608
Major abbreviations and terms 609

12.1 Key findings and recommendations by chapter

This section lists the key findings and recommendations from each of *Chapters 3 - 11*. The key findings collectively provide a concise overview of the content and results for the corresponding chapter, while the recommendations include both guidelines for reanalysis data users and suggestions for reanalysis data producers. Each subsection also includes a summary figure assessing the reliability of selected reanalyses with respect to key diagnostics, which is reproduced from the summary section of the corresponding chapter. (Please refer to the footnote [1] for the meaning of the evaluation terms, *i.e.*, demonstrated suitable, suitable with limitations, use with caution, demonstrated unsuitable, and unevaluated.) These assessments, while inherently subjective, are intended to provide the reader with an overview of the relative quality of the diagnostic. So, for example, across a given diagnostic the relative performance of the different reanalyses can be compared, and (*e.g.*, for a given reanalysis) the performance across different diagnostics can be compared. Only those diagnostics specifically examined either in this report or in previously published papers are assigned a score in the table; otherwise they are marked unevaluated. Although not all diagnostics we use can be evaluated against observations, we attempt to assign evaluation scores to any key diagnostics that can be readily summarized. For those that cannot be compared to observations, our assessment reflects consistency with other processes and current understanding of the phenomenon in question. Diagnostics that preclude simple classification (*e.g.*, the assessment of polar transport processes yielded results that varied by hemisphere, time of year, altitude, location in the polar vortex, and species) are omitted from these summary figures. Readers interested in further information about any diagnostic should refer to the corresponding chapter and section of the report, for which a key is provided in each summary figure.

It is noted here that, as explained in *Chapter 1* (**Table 1.1**), some ERA5 data have been available since July 2018, ERA5 data from 1979 onward have been available since January 2019, and a preliminary version of ERA5 1950 - 1978 data have been available since November 2020. Because most of the studies in this report were finalized before ERA5 was readily available, full evaluation of ERA5 has not been made. However, Chapter 2 includes information on the ERA5 system, and some chapters show ERA5 results for some diagnostics.

A key for all abbreviations used in this chapter is provided at the end.

12.1.1 Chapter 3: Overview of Temperature and Winds

In this chapter, we have examined reanalysis representations of key diagnostics related to temperature and winds. A summary of the diagnostics evaluated in this chapter is provided in **Figure 12.1**, which also directs the reader to the appropriate chapter section for further information. Below, we briefly summarize the key findings from this chapter and recommendations for both the appropriate use of and potential for improving reanalysis temperature and wind fields.

Key Findings of Chapter 3:

- More recent reanalyses from all centres consistently outperform earlier versions. (*e.g.*, JRA-55 *vs*. JRA-25; MERRA-2 *vs*. MERRA).
- Drifts and jumps in the long-term temperature time series can occur due to changes in available data sources. These irregularities are most pronounced at altitudes above 10 hPa. Greatest caution is advised when determining trends with reanalysis temperature data sets above 10 hPa.
- The more recent reanalyses have fewer discontinuities in their temperature and wind time series owing to improved data assimilation techniques and smoother transitions among different sets of observations.

[1] (As in *Chapter 1*, *Section 1.3*)
Demonstrated suitable: the reanalysis product could be directly validated using observational or physical constraints and was found to be in close agreement with expectations
Suitable with limitations: the reanalysis product could be directly validated using observatial or physical constraints and exhibited limited agreement; or, appropriate constraints were unavailable but reanalysis products were consistent beyond specific limitations as described in the text
Use with caution: the reanalysis system contains all elements necessary to provide a useful representation of this variable or process, but that representation has evident red flags (*e.g.*, disagreement with available observations; meaningful disagreements among reanalyses that cannot be resolved at this point)
Demonstrated unsuitable: the reanalysis product has been flagged as unable to represent processes that are key for this diagnostic as assessed in this report or by previous studies. This category is reserved for situations where the reanalysis is missing something fundamental in its structure (*e.g.*, a model top at 3 hPa means NCEP-NCAR R1 is 'demonstrated unsuitable' for studying processes in the USLM)
Unevaluated: the performance of the reanalysis product with respect to this diagnostic or variable has not been examined in this report or by previous studies

- The transition from the TOVS to ATOVS satellite periods starting around 1998 - 1999 is problematic for all reanalyses. In the stratosphere, the transition from three broad SSU infrared channels to five narrower AMSU/ATMS microwave channels proves to be problematic for data assimilation.

- The more recent reanalyses agree quite well with each other in the lower and middle stratosphere. All reanalyses have greater differences in the upper stratosphere and lower mesosphere. The latter discrepancies result from differences in model top, vertical resolution, data assimilation techniques, and data that are assimilated. *Chapter 2* provides detailed information about each reanalysis system.

- Temperature biases exist between the various reanalyses in the UTLS, especially before 1998. Temperatures in this region do not harmonize until after 2005, when widespread GNSS-RO observations became available.

- The agreement between Singapore radiosonde winds and reanalysis QBO winds at Singapore is better in the second half than the first half of the 1980 - 2014 record, consistent with improved constraints on reanalysis winds due to the gradual increase in the number of radiosonde observations over time. We expect that future reanalyses will have better QBO winds as forecast models become better able to produce a spontaneous QBO in the tropics.

Recommendations from Chapter 3:

- Users of any reanalysis should proceed with greatest caution when intercomparing reanalyses, and particularly when attempting to detect trends and/or changes in climate above the tropopause (see also *Section 12.2*).

- Improving the TOVS period would be highly beneficial to future reanalyses, especially for climate studies. However, the TOVS period may never be as good as the ATOVS period due to the relative sparsity and coarser vertical resolution of the assimilated data.

- Improvements to the variational bias correction schemes for handling the broad SSU weighting functions and improvements to the forecast models (especially the non-orographic gravity wave parameterizations, so that forecast models can generate a realistic QBO on their own) are some of the ways the TOVS time period can be improved upon.

- It may benefit each "satellite-era" reanalysis to begin their reanalysis several years earlier using just conventional data. This most likely will help harmonize the reanalyses' temperature structure below 10 hPa at the start of assimilating satellite data.

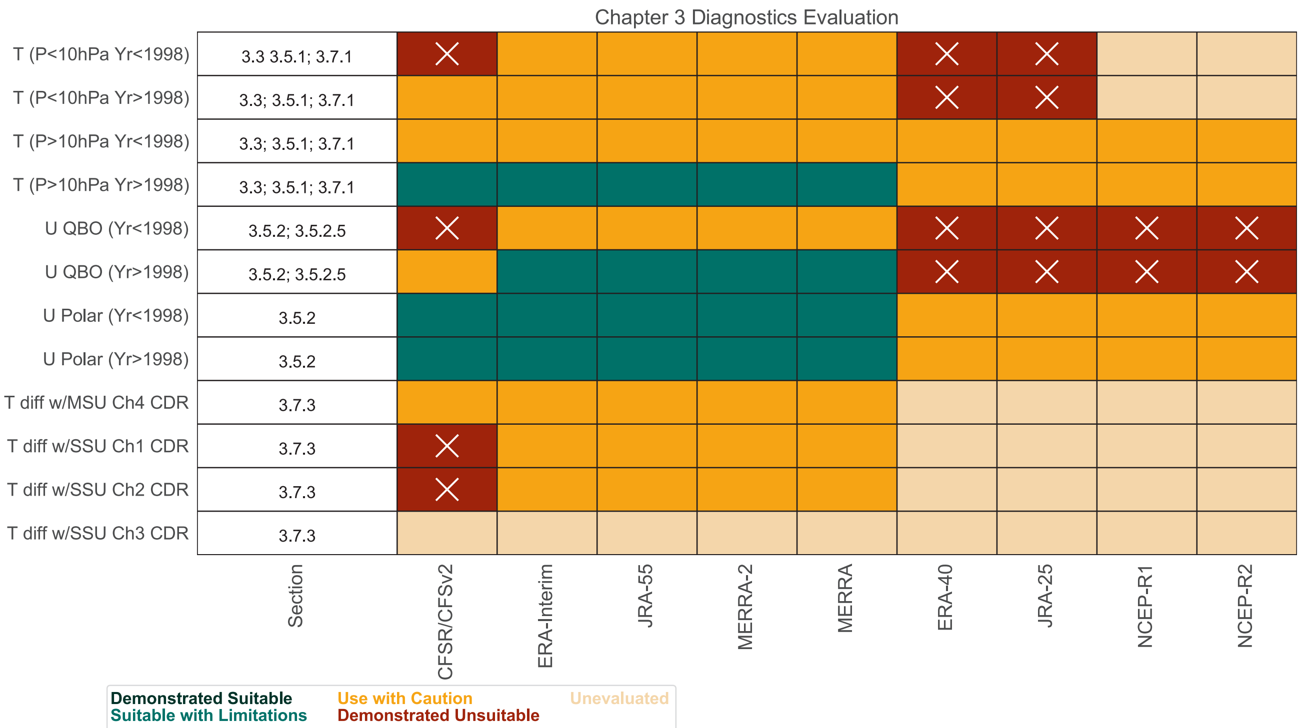


Figure 12.1: *(Same as **Figure 3.26**.) A summary of the diagnostics evaluated in Chapter 3: Overview of Temperature and Winds. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. Note that the score corresponding to "demonstrated suitable" was not assigned to any of the diagnostics listed here, so the darkest green colour does not appear in this table. "T" = temperature; "P" = pressure; "Yr" = year, "U" = zonal wind; "QBO" = Quasi-Biennial Oscillation; "T diff w/" = temperature difference with; "MSU" = Microwave Sounding Unit (a satellite instrument); "Ch" = Channel; "CDR" = climate data record; "SSU" = Stratospheric Sounding Unit (a satellite instrument).*

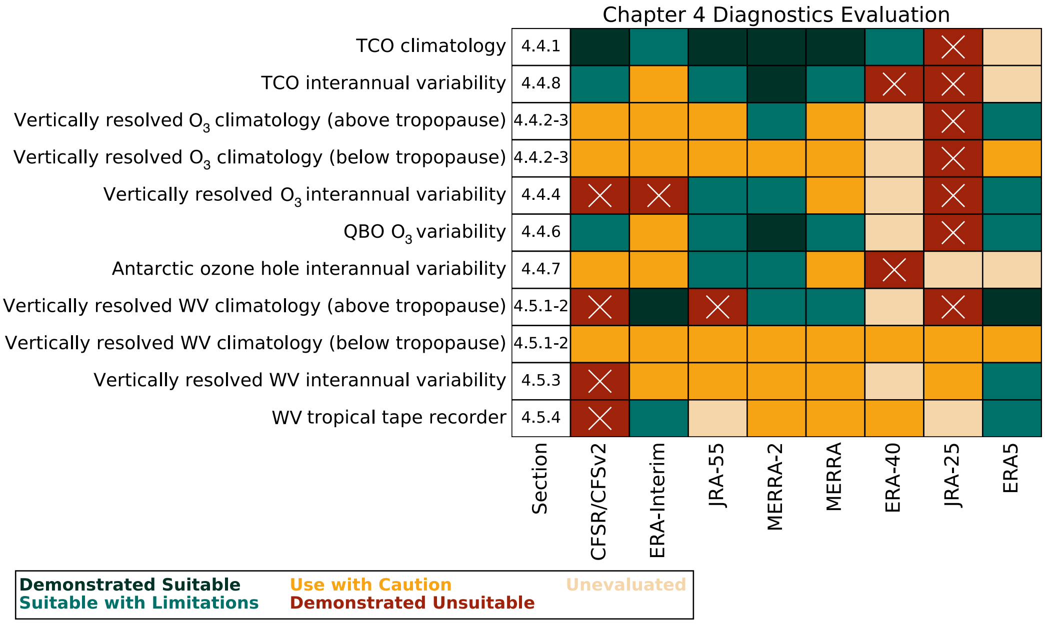


Figure 12.2: *(Same as **Figure 4.21**.) A summary of the diagnostics evaluated in Chapter 4: Overview of Ozone and Water Vapour. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "TCO" = Total Column Ozone; "QBO" = Quasi-Biennial Oscillation; "WV" = water vapour.*

12.1.2 Chapter 4: Overview of Ozone and Water Vapour

In this chapter, we have assessed the reanalysis representations of key diagnostics related to ozone and water vapour. A summary of the diagnostics evaluated in this chapter is provided in **Figure 12.2**, which directs the reader towards the appropriate chapter section for further information. Below, we briefly summarize the key findings from this chapter and recommendations for both use of and improvements to reanalysis ozone and water vapour fields.

Key findings of Chapter 4:

- The treatment of ozone and water vapour varies substantially among reanalyses, both in terms of their representation of these species and assimilated observations.
- The latest generation of reanalyses all assimilate satellite total column ozone observations, with some including vertically-resolved measurements.
- Currently none of the reanalyses directly assimilate WV observations in the stratosphere, although they do assimilate temperature and tropospheric humidity observations that can impact their stratospheric water vapour concentrations.
- Comparisons against assimilated observations of total column ozone (TCO) show that reanalyses generally reproduce TCO well in sunlight regions, within ~10 DU (~3%).
- The lack of TCO observations in polar night, and lack of representation of heterogeneous chemistry in most reanalyses, lead to relatively larger errors in representing TCO in the Antarctic ozone hole.
- From the middle to upper stratosphere, climatological reanalysis ozone profiles are within ±20% of observations.
- Biases are generally larger (~50%) for both water vapour and ozone in the upper troposphere and lower stratosphere.
- Significant discontinuities exist in reanalysis water vapour and ozone fields due to transitions in the observing system.

Recommendations from Chapter 4:

- Users should generally use caution when using reanalysis ozone fields for scientific studies and should check that their results are not reanalysis-dependent.
- Reanalysis stratospheric water vapour fields should generally not be used for scientific data analysis (except perhaps for ERA5). Any examination of these fields must account for their inherent limitations and uncertainties.
- In order to improve reanalysis ozone fields, reanalysis centres should work towards improved chemical parameterisations of ozone as well as assimilation of vertically-resolved ozone measurements (*e.g.*, from limb sounders) and measurements in polar night (*e.g.*, from IR nadir sounders).
- In order to improve reanalysis water vapour fields, future efforts should include the collection and assimilation of observational data with sensitivity to stratospheric water vapour, the reduction of reanalysis temperature biases in the TTL, and improvements in the representation of other processes that affect the stratospheric entry mixing ratio.

12.1.3 Chapter 5: Brewer–Dobson Circulation

This chapter presented both a direct comparison of Brewer-Dobson Circulation (BDC)-related dynamical diagnostics from the reanalysis datasets and transport tracer simulations using reanalysis products to drive different offline chemistry-transport models (CTMs). The direct dynamical diagnostics support intercomparison among the reanalyses, whereas the CTM simulations allow comparison against observation-based mean age-of-air (AoA) and stratospheric water vapour distributions, time series, and trends. A summary assessment of representation of the BDC in major reanalyses is provided in **Figure 12.3**, which directs the reader to the appropriate section of the chapter for further information. In the following, we briefly summarize our key findings and recommendations.

Key findings from dynamical diagnostics in Chapter 5:

- The BDC is generally much more consistent and weaker in more recent products compared to their older versions, although there are still significant differences in basic climatological diagnostics for some fields (*e.g.*, shallow branch wave driving, tropical upwelling structure and seasonality, upwelling strength below 70 hPa).
- Dynamical diagnostics show spurious fluctuations in CFSR; this product should thus not be used for long-term trend or interannual variability analyses.
- Estimates of long-term trends (for 1979 - 2016) in tropical upwelling are inconsistent: MERRA-2 and JRA-55 show positive trends, ERA-Interim shows a negative trend, and ERA5 shows no trend.
- Interannual variability and long-term trends in poleward mass transport through the turnaround latitudes ("tropical outwelling") are inconsistent; this suggests that the shallow branch of the BDC is not well constrained, even in the most recent products.
- Latitudinally and vertically resolved trends in residual circulation transit times (RCTTs) show some coherent signatures of a strengthening of the BDC (decreasing RCTTs, especially for the shallow branch), although the afore-mentioned inconsistencies across products also manifest in this diagnostic (especially for the deep branch).

Key findings from transport tracer simulations in Chapter 5:

- Simulations based on more recent reanalyses produce mean AoA in much better agreement with observations than those based on the previous generation of reanalyses (*e.g.*, ERA-Interim *vs*. ERA-40), indicating that reanalysis representations of the BDC have improved. However, significant discrepancies still remain in AoA and tracer distributions among reanalyses, with the spread of AoA obtained using different reanalyses as large as that obtained by using different CCMs.
- Differences among reanalysis diabatic heating rates[2] are evident and are a major factor affecting offline simulations of stratospheric tracers using diabatic models. Vertical transport within the tropics is too slow in MERRA and MERRA-2, in agreement with smaller diabatic heating rates compared to the other reanalyses. However, this slower tropical transport is evident in both diabatic and kinematic simulations, indicating that the slower BDC in the GEOS-5 system is not solely attributable to the radiation budget. The RCTT diagnostic also shows longer residence times for MERRA and MERRA-2.
- Our offline simulation results show large spread in the values and signs of AoA trends over 1989 - 2010, depending on the reanalysis and on the region of the stratosphere. For the MIPAS period (2002 - 2012) only ERA-Interim is in good agreement with the observed trends, regardless of the offline model used. A positive trend in the mean AoA in the NH is a robust feature in our studies and is in agreement with other observed phenomena. We emphasize that much investigation is still needed on BDC trends and that these trends should be interpreted with caution regardless of source, as natural variability and changes in the observation system make them highly sensitive to the choice of analysis period.
- Large spread in AoA among reanalyses emerges from two main sources: i) differences among the underlying models used to produce the reanalyses, and ii) the relatively weak constraints on stratospheric transport provided by assimilated observations in reanalyses. AoA diagnostics are affected by many other Earth system phenomena, including the stratospheric QBO signal, ENSO variability, and volcanic eruptions, indicating that improvements in the models and the data assimilation systems can both aid in achieving more accurate BDC representations in future reanalyses.

[2] Please note that the diabatic heat budget is not closed in reanalyses. This lack of closure occurs because the data assimilation step can cause changes in temperature that add or remove heat from the system. This analysis increment can be considered as a separate 'diabatic' term in the thermodynamic energy equation, but its application differs amongst reanalyses. Notably, the inclusion of the analysis increment as an additional tendency term in MERRA and MERRA-2 may in turn affect other physical tendency terms produced by the atmospheric model, as the latter are archived during the IAU corrector step rather than the predictor step (see *Chapter 2*, *Section 2.3*). By contrast, tendencies produced by other reanalyses are archived prior to the analysis during the initial forecast/predictor step. The analysis tendency is required to close the budget in either case, but these distinctions should be taken into account when evaluating or interpreting reanalysis diabatic heating products.

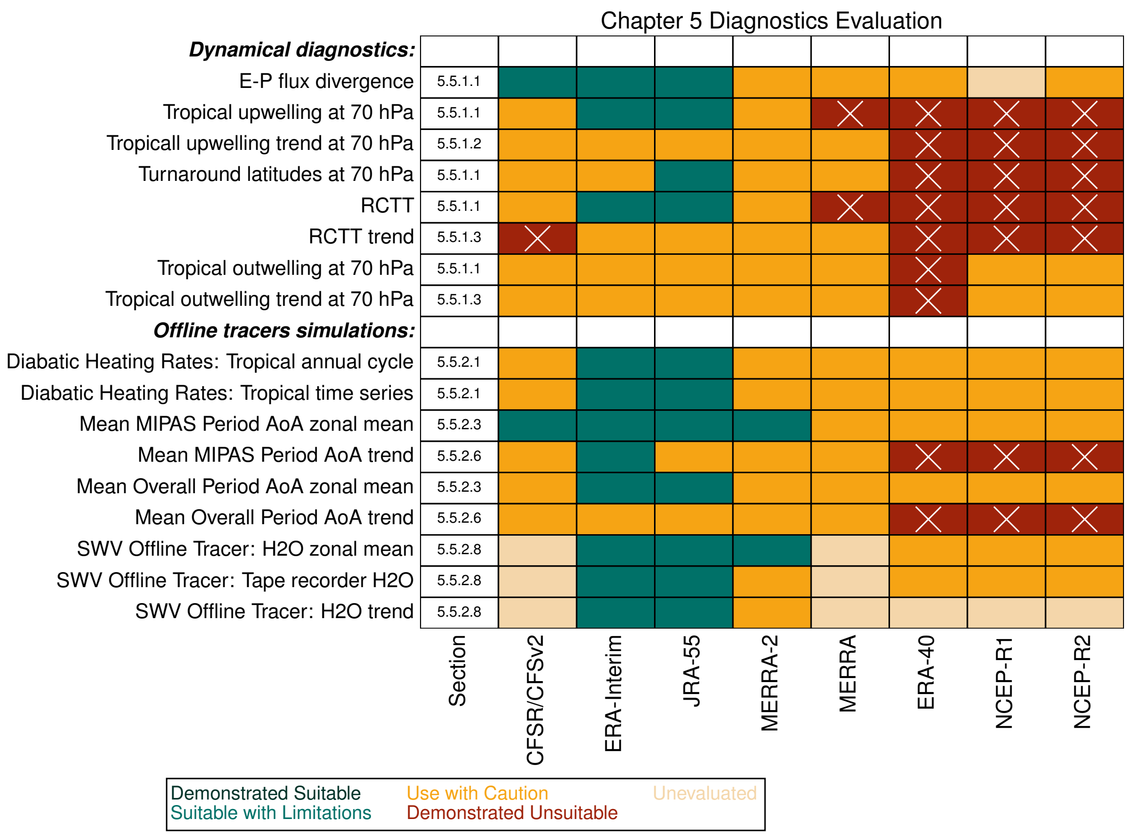


Figure 12.3: *(Same as* ***Figure 5.50****.) A summary of the diagnostics evaluated in Chapter 5: Brewer-Dobson Circulation. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. Note that the score corresponding to "demonstrated suitable" was not assigned to any of the diagnostics listed here, so the darkest green colour does not appear in this table. "E-P flux" = Eliassen-Palm flux; "RCTT" = Residual Circulation Transit Time; "MIPAS" = Michelson Interferometer for Passive Atmospheric Sounding (a satellite instrument); "AoA" = Age of Air; "SWV" = Stratospheric Water Vapour.*

- MERRA-2 shows difficulties in reproducing QBO-related BDC variability before 1995 relative to ERA-Interim and JRA-55. Another feature that is present in MERRA-2 but not in these other two reanalyses is the assimilation of Aura MLS temperatures from 2004 onwards at altitudes above 5hPa. The additional constraints provided by these data can affect stratospheric dynamics, and therefore BDC diagnostics.

Recommendations from Chapter 5:

- MERRA-2 may not be a good option for years before 1995, as it has difficulty reproducing observed QBO variability in stratospheric transport, which also affects its ability to reproduce QBO-related BDC variability.
- Among the more recent reanalyses, CFSR has been found to be problematic for BDC studies, especially with respect to interannual variability and long-term trends. Numerous published studies have also shown that older reanalyses like ERA-40, NCEP-NCAR R1, and NCEP-DOE R2 provide unrealistic representations of the BDC and other stratospheric processes. We therefore discourage the use of these older reanalyses for studies of the stratospheric circulation and associated tracer transport.
- Whenever possible we recommend that users not restrict themselves to only one product when conducting studies of the BDC or related transport. In particular, for the period after 2000, comparisons among MERRA-2, JRA-55, and ERA-Interim can help to distinguish robust from non-robust diagnostics.
- We recommend that users work with reanalysis data on model levels for offline simulations and diagnostics related to the shallow branch of the BDC.
- For future reanalyses, we recommend that reanalysis producers: i) provide variable uncertainty information; ii) provide variables at higher vertical resolution, especially within the UTLS region; iii) provide pressure level data at pressures less than 1 hPa (important for RCTT calculations); iv) archive data at higher frequencies; v) archive additional relevant variables (*e.g.*, heating rates) by default.
- The recently released ERA5 includes most of these features, although the resolution around the UTLS is still coarser than desired.

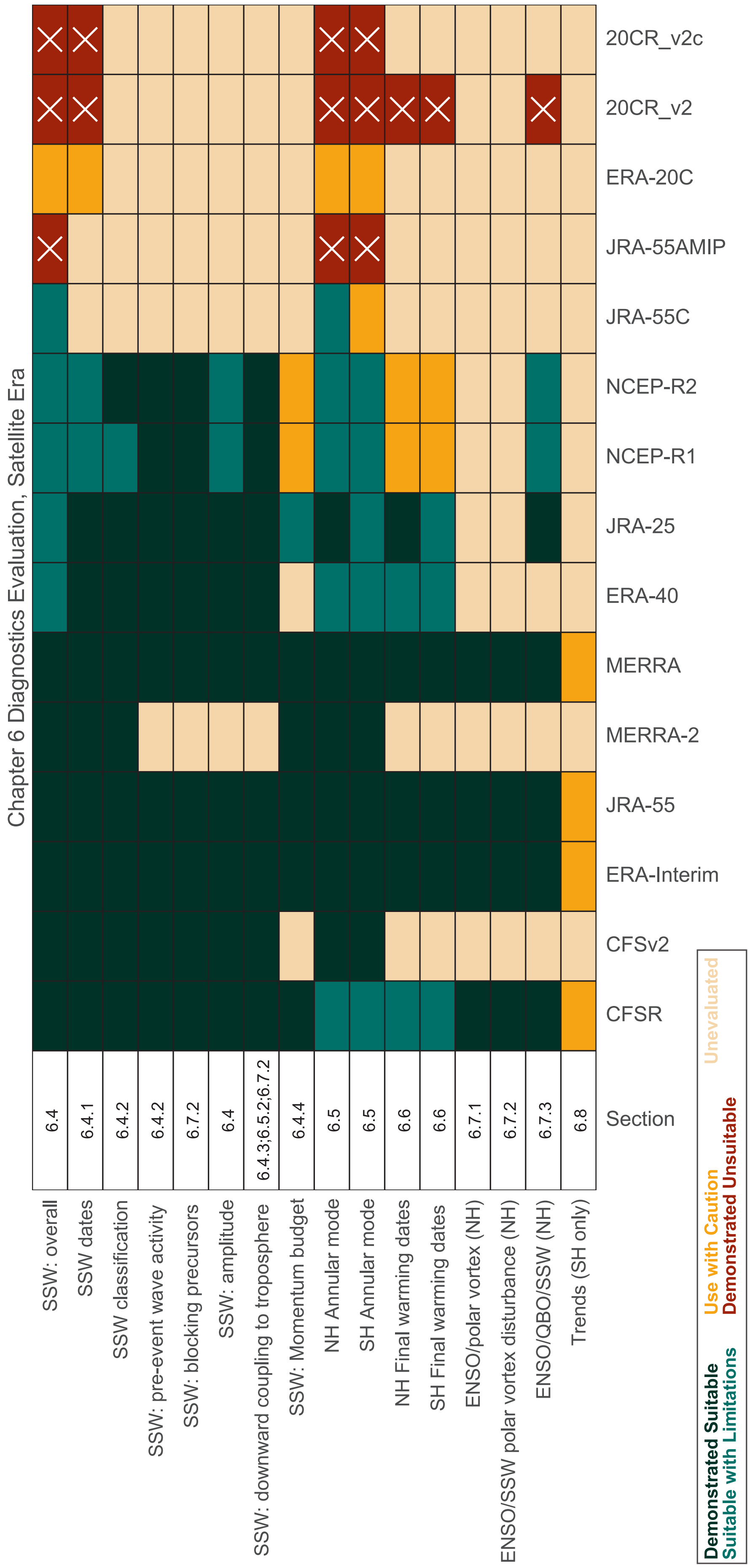


Figure 12.4: *(Same as* ***Figure 6.23****.) Evaluation of reanalyses during the satellite era (1979 onward) based on diagnostics computed for Chapter 6: Extratropical Stratosphere-Troposphere Coupling. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "SSW" = Sudden Stratospheric Warmings; "NH" = Northern Hemisphere; "SH" = Southern Hemisphere; "ENSO" = El Niño–Southern Oscillation; "QBO" = Quasi-Biennial Oscillation.*

- Further model development will be required to improve the representation of the BDC in future reanalyses. Aspects that require particular attention are: i) gravity wave drag parameterisations; ii) representations of radiative gases and aerosols in the stratosphere; iii) cloud and convection parameterisations, especially in tropical latitudes; iv) assimilation of stratospheric winds; v) model vertical resolution in the UTLS; and vi) extension of the vertical range to incorporate mesospheric processes.

- Sustained long-term observation platforms are required to monitor changes in the strength and structure of the BDC, to keep evaluating how well current and future reanalyses represent major stratospheric circulation patterns. Therefore, we strongly recommend the creation and sustained support of such observation platforms, and that they operate long enough to cover time scales relevant to the evolution and trends of the BDC.

12.1.4 Chapter 6: Extratropical Stratosphere–Troposphere Coupling

Atmospheric reanalyses are vital for evaluating stratosphere-troposphere coupling due to the lack of direct observations of the large-scale atmospheric circulation. In this chapter, we examined the representation of coupling between the troposphere and stratospheric polar vortices across the reanalyses. We assessed the reanalyses in terms of their internal consistency and in terms of their consistency with one another. Summary assessments of key stratosphere-troposphere coupling diagnostics are provided in **Figure 12.4** for the satellite era (1979 and later) and **Figure 12.5** for the pre-satellite era (1958 - 1978). Both figures direct the reader to the appropriate section of the chapter for further information. In the following, we briefly summarize key findings and recommendations based on our evaluation.

Key findings of Chapter 6:

- In the satellite era (1979 - onward), the representation of large scale stratosphere-troposphere circulation is very consistent across all full-input reanalyses. On synoptic scales, the more recent reanalyses (ERA-Interim, JRA-55, MERRA, and MERRA-2, and to a slightly lesser extent, CFSR/CFSv2) become more clearly superior.

- Our ability to assess and understand stratosphere-troposphere coupling is primarily limited by sampling uncertainty, that is, by the comparatively large natural variability of the circulation relative to the length of the satellite record. As an example, various efforts have sought to characterize the break-down of the polar vortex during a Sudden Stratospheric Warmings (SSW) as a split or displacement event. Methodological differences among the classifications proposed in the literature, however, result in a partial agreement (for two-thirds of SSW events). In contrast, applying the same definition to different reanalyses yields nearly identical results.

- Although measures of stratosphere-troposphere coupling determined from earlier reanalyses are generally not statistically distinct from results obtained with a more recent reanalysis, the more recent products show demonstrable improvement, particularly with respect to internal consistency (*e.g.*, the momentum budget) and at higher levels (10 hPa and above).

- Reanalysis datasets broadly agree on trends in the austral polar vortex related to ozone depletion since 1979. In contrast, there are no discernible trends in Northern Hemisphere polar vortex variability.

- Pre-satellite era reanalyses (1958 - 1978) appear to be of good quality in the Northern Hemisphere, and therefore can be used to reduce sampling uncertainty in measures of stratosphere-troposphere coupling by approximately 20 %. We emphasize that this represents a more significant reduction in uncertainty than achieved by shifting from an earlier generation reanalysis to a more recent reanalysis.

- Pre-satellite era reanalyses of the Southern Hemisphere are generally of poor quality, and can only be used to reduce sampling uncertainty with great caution.

- A conventional-input reanalysis of the Northern Hemisphere (JRA-55C) matches full-input reanalyses well up to 10 hPa, supporting the validity of pre-satellite reanalysis products in this hemisphere. JRA-55C's representation of the Southern Hemisphere is not as accurate, suggesting that satellite measurements are more critical in this hemisphere due to the reduced density of conventional observations.

- Surface-input reanalyses have also been evaluated. ERA-20C captures not only the correct statistical climatology of the Northern Hemisphere stratospheric polar vortex, but also much of its actual variability (correctly representing the timing of about half of observed SSWs). This suggests it may be suitable for exploring low-frequency variability of the stratosphere-troposphere coupled system. The representation of the stratospheric vortex in NOAA 20CR v2/v2c, however, is demonstrably poor.

Recommendations from Chapter 6:

- We recommend the use of more recent reanalysis products. As a matter of best practice, we urge all users to avoid the use of earlier reanalyses unless the project requires the use of an older product, and special care is taken to justify that the older product is otherwise consistent with more recent reanalyses. In particular, we note for users that modern reanalyses can be obtained, in addition to their native high-resolution grids, at a coarser resolution that is comparable to that of earlier reanalyses and thus more manageable in size, but which still captures the best representation of the large-scale circulation.

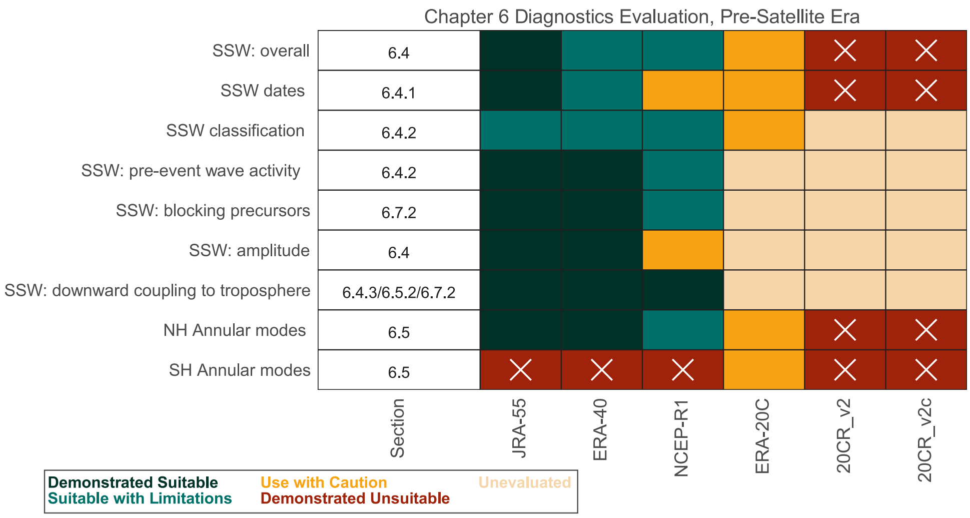


Figure 12.5: *(Same as **Figure 6.24**.) Evaluation of reanalyses during the pre-satellite era (1958 - 1978) based on diagnostics computed for Chapter 6: Extratropical Stratosphere-Troposphere Coupling. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "SSW" = Sudden Stratospheric Warmings; "NH" = Northern Hemisphere; "SH" = Southern Hemisphere.*

- The consistency of trends associated with the Antarctic ozone hole (for the period 1979 forward) suggest that reanalyses may be reliably capturing the influence of stratospheric ozone loss. One must exercise great caution in the interpretation of trends in the reanalyses, however, as they can be spuriously caused by changes in the observations assimilated over time, an issue that could systematically affect all products. Additional support from direct observations and/or understanding of the mechanism(s) help build confidence in trends found in the reanalyses.

- When an extended record is needed to reduce sampling uncertainty, we recommend the use of pre-satellite era reanalyses (1958 - 1978) in the Northern Hemisphere, but caution against their use in the Southern Hemisphere.

- Due to significant biases in the mean state and variability of the polar vortex in the NOAA 20CR surface-input reanalysis, we do not recommend it for the purpose of investigating stratosphere-troposphere coupling.

- ERA-20C may be suitable, with caution, for exploring the low-frequency variability of the stratosphere-troposphere coupled system.

- As our ability to quantify the large scale coupling between the stratosphere and troposphere is primarily limited by sampling uncertainty, we recommend that future reanalysis products extend their analysis prior to the satellite era.

12.1.5 Chapter 7: Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS)

In this chapter, we have evaluated diagnostics that are critical to understanding ExUTLS dynamical and transport processes, including the extratropical tropopause; upper tropospheric (UT) jet streams; mixing and transport diagnostics; and ozone distributions and evolution. Because representing these processes requires high resolution, we focus on recent full-input reanalyses, including MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2, with the conventional input JRA-55C also included for a few diagnostics. **Figure 12.6** summarizes the results for the main diagnostics evaluated in this chapter, and directs the reader to the appropriate section of the chapter for further information. Because most of the diagnostics evaluated in *Chapter 7* cannot be verified using direct observations, there are very few cases where we can rate the reanalyses as "demonstrated suitable".

We summarize our key findings and recommendations below.

Key Findings of Chapter 7:

- The reanalyses evaluated here agree well on the placement of the lapse-rate tropopause, both with each other and with data from high-resolution radiosonde observations. CFSR/CFSv2 shows the smallest errors with respect to radiosonde-based lapse-rate tropopause data.

- Long-term trends in tropopause characteristics are in broad agreement both among the reanalyses and with observations, except for CFSR/CFSv2.

- The representation of multiple lapse-rate tropopause altitudes, which indicate lateral stratosphere-troposphere exchange (STE) events between the tropical UT and extratropical LS, is highly dependent on the vertical grid resolution of the reanalysis. CFSR/CFSv2 has the highest frequency of multiple tropopauses, as well as the highest ExUTLS resolution among the reanalyses evaluated here.

- Using pressure and model-level versions of CFSR/CFSv2, we show that the coarser vertical resolution of the pressure-level fields makes them unsuitable for identifying tropopause locations, especially in multiple-tropopause situations.

- JRA-55C is unsuitable for identifying multiple tropopauses because of its inability to qualitatively reproduce the distributions in SH high latitudes.

- Despite a general under-representation of multiple tropopause frequency compared to observations, most modern reanalyses reproduce the pattern and sign of observed long-term trends.

- The reanalyses show good overall agreement in representing the climatologies of UT jets and the sub-vortex jet in the lowermost stratosphere.

- Robust trends in UT jets (latitude, altitude, and windspeed) are limited to particular longitude regions and seasons. Disagreement among the reanalyses is most common for the SH jets; in particular, MERRA-2 and/or CFSR/CFSv2 sometimes differ from the other reanalyses even in the sign of the SH jet latitude trend.

- Kinematic STE is in broad agreement among the reanalyses, with some important differences in the magnitudes and long-term changes of troposphere-to-stratosphere transport and stratosphere-to-troposphere transport. Transport estimates are sensitive to the choice of vertical coordinate (*i.e.*, diabatic vs. kinematic) and the period analyzed.

- Mixing diagnostics including effective diffusivity and PV gradients as a function of equivalent latitude (EqL) show generally good agreement in both climatological seasonal cycles and interannual variability.

- Mass flux across the 380 K isentropic surface agrees well among MERRA-2, ERA-Interim, and JRA-55, but CFSR/CFSv2 shows inconsistencies in the seasonal cycle.

- Climatological ozone distributions and seasonal cycles show good qualitative agreement. Given large differences in the ozone products assimilated and the methods of assimilating them, this points to good representations of the dynamics in the UTLS, where ozone changes are primarily driven by dynamical and transport processes.

- Reanalysis ozone fields mapped in EqL generally reproduce at least qualitatively the interannual variability in MLS-observed ozone, but ERA-Interim shows several step function changes that are related to changes in the versions of MLS ozone assimilated. For example, large biases in ERA-Interim UTLS ozone arise in mid-2009 through 2012 owing to the use of an early version of MLS near real time data.

Recommendations from Chapter 7:

- Only the recent high-resolution reanalyses (MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 are such reanalyses evaluated herein) are suitable for ExUTLS dynamical and transport studies. Dynamical diagnostics derived from these reanalyses indicate that they are all suitable for use in such studies with some limitations. Earlier reanalyses (*e.g.*, ERA-40, NCEP-NCAR R1, and NCEP-DOE R2) are not suitable for detailed UTLS studies and are not evaluated here.

- A few diagnostics (*e.g.*, effective diffusivity in CFSR/CFSv2; ozone in ERA-Interim) show substantial discontinuities when assessed over many years, and thus should be used with greatest caution and awareness.

- Because many diagnostics in this chapter cannot be directly compared with observational data, it is important that ExUTLS studies use multiple reanalyses and assess agreement among them whenever possible.

- For diagnostics that cannot be directly compared with data, and in light of similar changes in input data, agreement among the reanalyses should be regarded as a necessary but by no means sufficient condition for robustness of trends.

- As is the case for diagnostics described in other chapters (*e.g.*, *Chapter 10*), differences between the PV fields arising from differing products provided by the reanalysis centres add to uncertainties in the evaluations. It would be helpful in the future for all reanalysis centres to provide PV on the model grids.

- The results from reanalyses assimilating MLS ozone (which has relatively high vertical resolution compared to other ozone profilers currently used) show promise for future improvements. More attention to consistently assimilating high-resolution ozone observations in future reanalyses would be extremely beneficial to understanding the processes controlling ozone in this region, where it is of great importance to the radiative balance.

- Future work is needed to better elucidate the role of various elements of model design in producing observed differences in tropopause location and characteristics (*e.g.*, through idealized simulations with the core models of each reanalysis).

- In the future, the accuracy of tropopause identifications in reanalyses should improve as the vertical grid spacing decreases. These diagnostics should be evaluated in forthcoming reanalyses (most immediately in ERA5) and the impacts of these improvements on estimates of STE and their long-term changes should be explored.

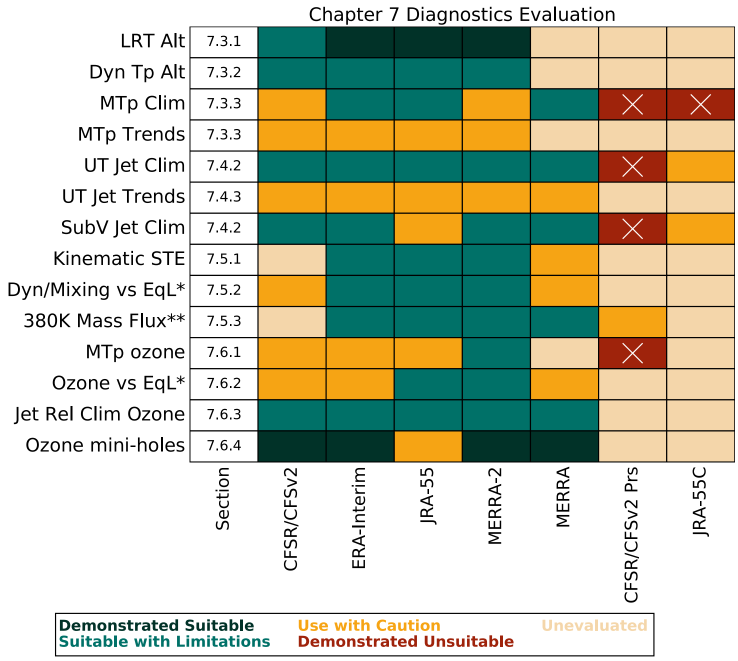


Figure 12.6: *(Same as **Figure 7.36**.) A summary of the diagnostics evaluated in Chapter 7: Extratropical Upper Troposphere and Lower Stratosphere (ExUTLS). The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. Because analyses as a function of equivalent latitude (EqL; marked by *) depend critically on PV (which is used to compute EqL), reanalyses where we have concerns about the PV fields are rated "use with caution" even in the absence of obvious "red flags". "CFSR/CFSv2 Prs" indicates CFSR/CFSv2 was used as interpolated to standard pressure levels, and the 380 K mass flux analysis (marked by **) was done using pressure level data for all reanalyses. All other diagnostics were calculated using model level data for all reanalyses. "LRT" = lapse-rate tropopause; "Alt" = altitude; "Dyn" = dynamical; "Tp" = tropopause; "MTp" = multiple tropopause; "UT" = upper troposphere/tropospheric; "Clim" = climatology; "SubV" = subvortex; "STE" = stratosphere-troposphere exchange; "Jet Rel" = in coordinates relative to the subtropical jet core location.*

- The accuracy of transport estimates from reanalyses is largely unknown, since global estimates of transport from observing systems are not available and the outcomes are sensitive to the input fields and methods used. Comparison of transport calculations using reanalysis wind fields and trace gas observations is one path to examine the accuracy of transport in reanalyses.

- If possible, errors in transport calculations should be increasingly gleaned from comparison of trajectory calculations driven by the reanalysis winds to long-duration balloon observations. However, such observations are infrequent and sometimes assimilated into the reanalysis, which limits their utility for validation studies.

- Given known errors in trajectory and other transport calculations that arise from coarse temporal resolution of input wind fields, more frequent 3D wind field outputs are desired from future reanalyses. Such wind fields, which are already available for ERA5, will allow for improved understanding of transport and STE.

- Increased horizontal and vertical grid resolution will also be beneficial for reducing errors in transport calculations and enabling analysis of processes at smaller scales.

12.1.6 Chapter 8: Tropical Tropopause Layer (TTL)

In this chapter, we have investigated the extent to which reanalysis data sets reproduce key characteristics of the TTL, including the cold point and lapse rate tropopause, the vertical structure and distribution of clouds within the TTL, basic dynamical processes and circulation patterns, transport statistics and residence times derived from trajectory simulations, equatorial wave activity, and long-term changes in the width of the tropical belt. We have also evaluated how key differences in reanalysis performance within the TTL impact upon regional and seasonal aspects of the South Asian Summer Monsoon (SASM) anticyclone. Summary assessments of reanalysis products in the TTL are provided in **Figure 12.7** for the global tropics and in **Figure 12.8** for the SASM. Key findings and recommendations from this chapter are outlined below.

Key findings of Chapter 8:

- Advances in reanalysis and observational systems over recent years have led to a clear improvement in TTL reanalysis products over time. In particular, the reanalyses ERA-Interim, ERA5, MERRA-2, CFSR, and JRA-55 show very good agreement after 2002 in terms of the vertical TTL temperature profile, meridional tropopause structure, and interannual variability.

Long-term temperature trends from reanalyses and adjusted radiosonde data indicate significant cooling in the upper TTL (above the cold point).

- While climatological TTL temperatures from reanalyses agree very well with observations with relatively small low biases, the cold point and lapse rate tropopause show warm biases, most likely related to the fact that the discrete values corresponding to reanalysis model levels are unable to reproduce the observed minimum temperature as recorded in a near-continuous profile.

- Cloud fields in the tropical UTLS vary greatly in both magnitude and vertical distribution across reanalyses. Differences in cloud fraction and cloud water content impact the radiation budget both at the top-of-atmosphere and within the UTLS, and the effects of differences in cloud and convection parameterizations can be identified in vertical profiles of temperature and humidity in the tropical troposphere.

- There are large differences among reanalysis diabatic heating rate products[3] within the TTL, which are known to influence transport statistics and rates of ascent in trajectory simulations of cross-tropopause transport in this region. Differences among reanalysis diabatic heating rates in the tropical UTLS are not limited to any one component: longwave, shortwave, and non-radiative components all show substantial discrepancies.

- Lagrangian transport studies demonstrate large differences in reanalysis temperatures at the dehydration point and in TTL residence times. However, the data sets agree on the spatial distribution of dehydration locations and produce roughly similar distributions, seasonal cycles, and interannual variations of TTL residence time.

- Equatorial wave activity and corresponding temperature anomaly patterns at 100 hPa are similar among the reanalyses, including the characteristic horseshoe-shaped structures that resemble the stationary wave response to tropical heating. However, the strength of the wave activities, their spectral magnitudes, and the intensity of temperature response differ among the reanalyses, with the latter differences depending on the aspects of the dynamical model and/or assimilation system.

- Metrics of the width of the TTL based on the zonally-resolved subtropical jet and tropopause break show robust changes in only a few regions and seasons and poor agreement of the resulting zonal-mean annual-mean values. The diagnostics based on the zonal-mean subtropical jet and tropopause break, on the other hand, suggest stronger trends in the width of the TTL than their zonally-resolved counterparts. Overall, the two subtropical jet diagnostics are more consistent than the two tropopause break diagnostics, possibly related to smoother variations in the zonal wind field relative to the tropopause break.

- Modern reanalyses agree well regarding the climatological position and evolution of area extent and moments of the SASM anticyclone, although there are notable differences in the distribution of SASM anticyclone centre locations. All of the reanalyses indicate slightly higher CPT temperatures and lower CPT heights in the SASM anticyclone compared to GNSS-RO satellite observations.

- Distributions of ozone volume mixing ratios within the SASM anticyclone are qualitatively consistent among reanalyses and broadly consistent with observations. However, none of the evaluated reanalyses are able to reproduce the low ozone mixing ratios within the SASM anticyclone.

- Cloud properties, convection, radiative heating, and omega fields for the SASM UTLS differ significantly among reanalyses on a regional scale as these properties are only weakly constrained by assimilated observations. These differences impact derived transport processes in the UTLS, and residence times based on diabatic Lagrangian transport calculations reveal large differences.

Recommendations from Chapter 8:

- In the TTL, temperature on native model levels should be used rather than the standard pressure-surface data sets. Various diagnostics such as the cold point and lapse rate tropopause and the analysis of equatorial waves are demonstrably improved when model-level data are used. For a more realistic representation of the tropical tropopause levels, data sets that combine low temperature biases with high vertical resolution should be used.

- Long-term drifts in high cloud fraction, OLR, and LWCRE are present in almost all reanalyses, and often disagree in terms of sign, timing, or magnitude. These products should generally not be used for trend or time series analysis without independent verification. Among the reanalyses, ERA5 shows greater stability in time and stronger correlations with observed variability for these cloud and radiation metrics and may therefore offer a more reliable characterization of long-term variations in related metrics relative to earlier reanalyses.

[3] See the footnote on diabatic heating rates in reanalyses in *Section 12.1.3.*

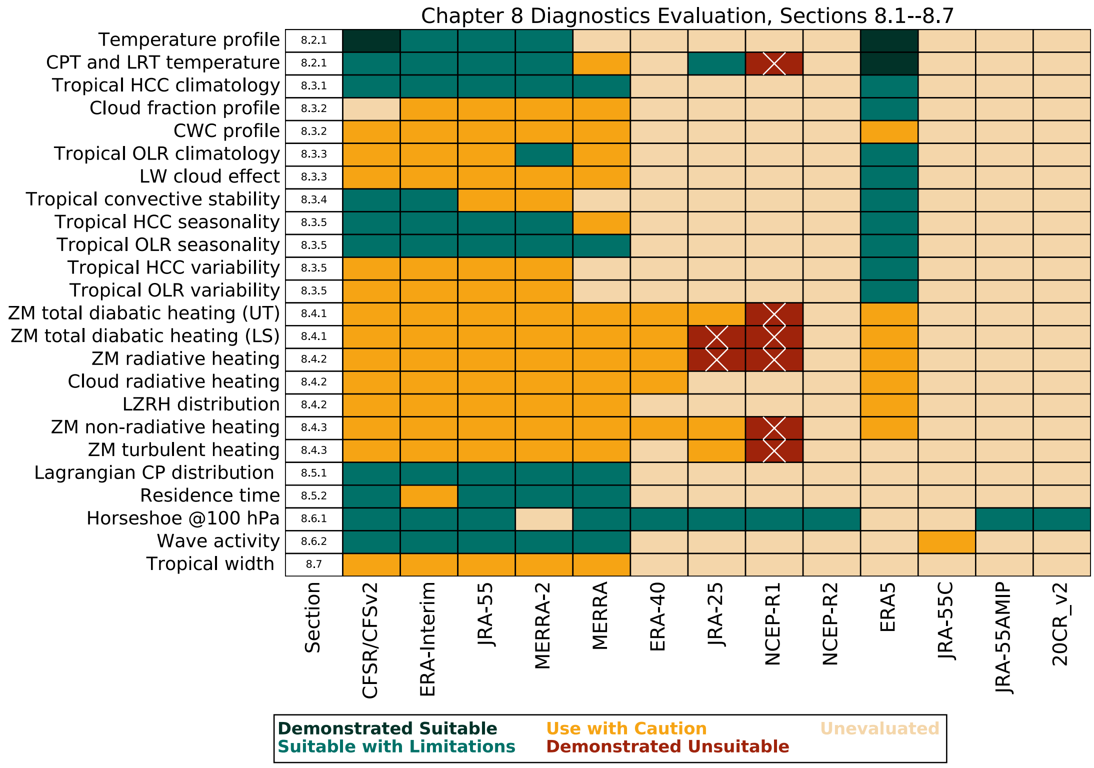


Figure 12.7: *(Same as* ***Figure 8.72****, top.) A summary of the diagnostics evaluated in Sections 8.2-8.7 of Chapter 8: Tropical Tropopause Layer (TTL). The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "CPT" = cold point tropopause; "LRT" = lapse rate tropopause; "HCC" = high cloud cover fraction; "CWC" = cloud water content; "OLR" = outgoing longwave radiation; "LW" = longwave; "ZM" = zonal mean; "UT"=Upper Troposphere; "LS"=Lower Stratosphere; "LZRH" = level of zero net radiative heating; "CP" = cold point.*

- Given large differences in reanalysis diabatic heating products and related metrics within the tropical UTLS, researchers using these fields to drive or nudge model simulations of this region should use multiple reanalyses whenever possible.

- When applying metrics of tropical width based on the subtropical jet or tropopause break, it is recommended to use multiple reanalyses and to be aware of the caveat that the zonal-mean diagnostics suggest stronger trends than their zonally-resolved counterparts.

- For analyses involving the SASM anticyclone it is recommended to use more recent reanalyses. In particular, researchers are encouraged to avoid NCEP-NCAR R1 and NCEP-DOE R2 data sets and the geopotential height field of the MERRA-2-ANA pressure-level data when possible.

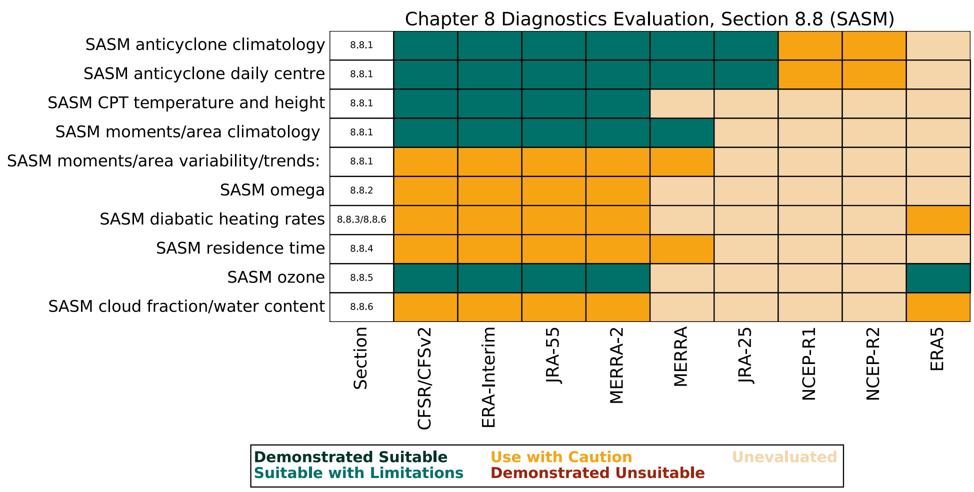


Figure 12.8: *(Same as* ***Figure 8.72****, bottom.) A summary of the diagnostics evaluated in Section 8.8 of Chapter 8: South Asian Summer Monsoon (SASM). The "Section" column at left indicates where in Section 8.8 each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "CPT" = cold point tropopause.*

- Transport simulations for the SASM domain that use diabatic heating rates to represent vertical motion should use multiple reanalyses if possible and carefully consider the representation of convective sources to the TTL. MERRA-2 diabatic heating rates should only be used at 370 K potential temperature level and above.

- Ozone in the UTLS above the SASM should be carefully validated against observations, and cloud and radiative heating should be used with caution for all reanalyses.

12.1.7 Chapter 9: Quasi-Biennial Oscillation (QBO)

In this chapter, we have investigated the representation of the quasi-biennial oscillation (QBO) and tropical stratospheric variability in atmospheric reanalyses. An assessment of key QBO-related diagnostics is provided in **Figure 12.9**, which includes cross-references for finding further information in the chapter text. Here we provide a concise summary along with recommendations on which reanalyses are appropriate to use for various diagnostics of the QBO and tropical stratospheric variability.

Key findings of Chapter 9:

- Reanalyses broadly agree with FUB winds on the evolution of the zonal-wind QBO apart from the older NCEP reanalyses (NCEP-NCAR R1 and NCEP-DOE R2), although even these adequately reproduce the phase of the QBO. The main error in NCEP-NCAR R1 and NCEP-DOE R2 is that the QBO wind amplitude is substantially underestimated (by a factor of 2 - 4, depending on the altitude considered).

- Inter-reanalysis spread in QBO winds has decreased in recent years, consistent with increasing observations to constrain the reanalyses. However, differences between JRA-55 and JRA-55C show no long-term trend, indicating that the increased satellite data assimilated into JRA-55 over the 1973 - 2012 period does not substantially affect the QBO winds. This suggests that satellite observations are less important than conventional observations for constraining the QBO.

- Most inter-reanalysis spread in QBO winds occurs during QBO phase transitions, especially the QBO-W (westerly) onset which is often delayed by ~ 1 - 2 months compared with FUB winds. These onsets are also delayed when compared with the MERRA-2 reanalysis, which uses a forecast model that spontaneously generates a QBO. Hence, we attribute the delays to lack of sufficiently strong westerly momentum deposition in the tropical stratosphere, that can only be provided by wave drag.

- There is substantial inter-reanalysis spread in strength and spatial structure of zonal winds in the tropical upper troposphere and tropopause region (both zonal-mean and zonally-varying components). This has implications for modelling tropical wave propagation (*i.e.*, how the background winds filter upward propagation of waves that force the QBO and SAO, including parameterized gravity waves). Small changes in wave filtering at lower altitudes can have substantial effects on wave forcing at higher altitudes.

- There is uncertainty in how much zonal asymmetry is present in the QBO, especially at 70 hPa, given that assimilation of winds in the tropics is dominated by the Singapore radiosondes. Inter-reanalysis spread is greatest over the oceans where there is a lack of radiosonde observations. Inter-reanalysis spread has reduced in recent years but spatial patterns remain unchanged, especially at 70 hPa where the flow is less zonally symmetric. QBO-related vertical velocity anomalies have comparable magnitude to the background vertical velocity, though the magnitudes of both vary among the reanalyses.

- Reanalysis QBO temperature anomaly evolutions compare well with sonde and GNSS-RO observations (all reanalyses considered here assimilate radiosondes, and the four recent ‘full-input’ reanalyses (ERA-Interim, CFSR, JRA-55, MERRA-2) assimilate GNSS-RO data, albeit over slightly different periods). Peak-to-peak QBO zonal-mean temperature variations are ~ 2 K at 70 hPa and ~ 1 K near the tropical tropopause (100 hPa), corresponding to 25 - 30 % and 15 - 20 % the size of the annual cycle, respectively. Zonal asymmetries are also evident, with QBO amplitude in the Indonesian region roughly 30 % larger than the zonal-mean amplitude. Comparison with GNSS-RO, which are spatially homogeneous, suggests this is a real feature rather than an artefact of the strong influence of the Singapore observations. This may have implications for QBO influences on convection and precipitation.

- There is good agreement on the relative contributions of the various tropical waves to forcing the QBO. The greatest inter-reanalysis spread is in the Kelvin wave contribution during the descending QBO-W phase. There is significant natural variability (*i.e.*, from one QBO phase to the next) in the various contributions. The vertical advection term differs widely among reanalyses, including in its sign, consistent with large inter-reanalysis differences in vertical velocity.

- Although assimilation of satellite observations does not have a major impact on the QBO wind evolution (as noted above) it nevertheless has an indirect impact via improved representation of different components of the waves that force the QBO, which may in turn contribute to improvements in details such as the spread in the timing of QBO phase changes referred to above. There is clear evidence that representations of tropical waves changed after introduction of the AMSU satellite observations in ~ 1998. Assuming that the observations are more accurate in the latter period, we recommend that the more recent data be used for studies of wave diagnostics.

- There are clear differences in wave characteristics when derived on model versus pressure surfaces. They are qualitatively similar, but for quantitative results model levels are better. Comparison of wave characteristics with satellite observations (HIRDLS, SABER, COSMIC, and AIRS) shows consistency between the reanalyses and high correlations in the tropical lower stratosphere with all observations except AIRS. Correlations with HIRDLS and SABER are notable because these observations are not assimilated by any of the reanalyses and thus provide independent validation. Reanalysis momentum fluxes in the lower tropical stratosphere correlate well with HIRDLS but less well with SABER.

- There is good inter-reanalysis agreement on teleconnections between the QBO influence and NH winter polar vortex (Holton-Tan effect), with clear impacts in early winter (November - January). A late winter reversal of this response (February - March) seen in the 1979 - 2016 analysis is not robust in the longer 1958 - 2016 period, highlighting the importance of using as long a data record as possible.

- There is no evidence for an early- or mid-winter QBO influence on SH vortex strength but good reanalysis agreement that the final SH warming occurs later in QBO-W than QBO-E when the phase is defined using 20 hPa QBO winds.

- In boreal winter there is a QBO impact on the strength of the tropical upper tropospheric winds of ~4 - 5 m s^{-1}, accompanied by an impact on the winter hemisphere subtropical jet near 30° latitude. There is good agreement of this signal for 1980 - 2016 in the four recent full-input reanalyses, but some details are not robust when the longer period 1958 - 2016 is examined.

- A QBO modulation of mean sea level pressure (MSLP) is found in NH winter over the extended 1958 - 2016 period in the JRA-55 reanalysis. The pattern, which in January resembles the North Atlantic Oscillation (NAO) pattern, is almost identical to that found in a recent study that combined ERA-40 / ERA-Interim to achieve a similarly long data record, suggesting that choosing either method for lengthening the data period is adequate for MSLP analysis.

- Analysis of the JRA-55 and ERA-Interim reanalyses over the satellite era demonstrate a QBO modulation of tropical precipitation, and both compare well with independent GPCC satellite observations. The response is mostly robust to inclusion of the pre-satellite years of JRA-55.

Recommendations from Chapter 9:

- Most reanalyses are suitable for determining the QBO phase but comparing several reanalyses is recommended for estimating the timing of phase transitions. MERRA-2 agrees best with the FUB at 30 hPa and is likely to provide the most accurate transition times at this level, but is a poor choice for 10 hPa QBO phase due to unusual features earlier in its record.

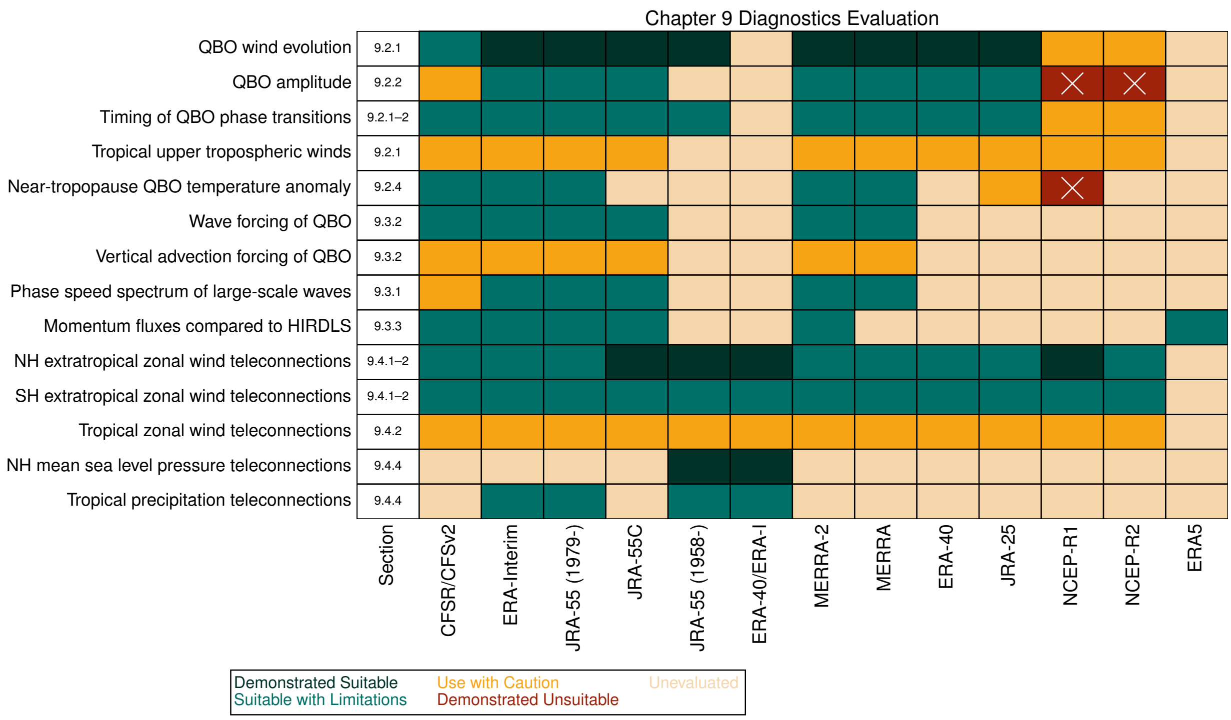


Figure 12.9: *(Same as **Figure 9.65**.) A summary of the diagnostics evaluated in Chapter 9: Quasi-Biennial Oscillation (QBO). The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "HIRDLS" = High Resolution Dynamics Limb Sounder (a satellite instrument); "NH" = Northern Hemisphere; "SH" = Southern Hemisphere.*

- The most recent reanalyses are recommended for comparison of QBO characteristics (amplitude, period, *etc.*) with climate models. JRA-55 provides the longest record and thus the most statistically robust estimates. MERRA-2 may also be a good choice because its representation of the QBO does not rely on the data assimilation to correct a severe model bias (*i.e.*, the lack of a QBO); however, at least based on the diagnostics presented here, this may not be important for most applications (and the aforementioned caveat about the 10-hPa level winds should also be noted). CFSR is less suitable than JRA-55, MERRA-2, or ERA-Interim because it underestimates the QBO amplitude compared to the other reanalyses.

- Conventional-input reanalyses are adequate for studies of the QBO as long as tropical radiosonde data are assimilated; JRA-55C appears to be as suitable for examining the QBO as JRA-55 (although its record is slightly shorter).

- Although not examined in the report, surface-input reanalyses (*e.g.,* ERA-20C) are not recommended for QBO studies. If a QBO exists in such a reanalysis it will be entirely produced by the forecast model; even if it is realistic, the lack of assimilated tropical stratospheric wind observations means that the QBO phase timing will almost certainly be incorrect.

- For studies of tropical temperature and meridional wind spectra any of the modern reanalyses are equally suitable since they show relatively small differences.

- For estimates of QBO wave forcing (*e.g.,* Eliassen-Palm flux divergence) care is required since there is substantial inter-reanalysis spread. Without suitable observations for validation it is not clear which reanalysis, if any, is most accurate, so comparison of several reanalyses is recommended. Given the very large natural (seasonal, inter-annual) variability of the QBO forcing terms, analysis of a long data period is recommended where appropriate.

- For QBO studies that involve the vertical advection term, comparison of as many of the modern reanalyses as possible is recommended because of large inter-reanalysis spread in vertical velocity in the lower tropical stratosphere. Model-level diagnostics are recommended since wave quantities can be damped by vertical interpolation. The post-1998 period is more reliable for evaluating wave spectra and QBO wave forcing.

- For investigation of QBO-vortex teleconnections we recommend using the longest available data records to maximise the signal-to-noise ratio (see also *Section 12.1.4*). However, while using pre-satellite era data to extend the data period is recommended for analysis of features at levels below ≈ 10 hPa, caution is required at the higher levels (*e.g.*, evaluating results from the pre- and post-satellite eras separately). For QBO studies of the SH, pre-satellite era data should be used with caution.

- For studies of the QBO impact on tropical / subtropical tropospheric circulation and surface precipitation the maximum available data period is recommended (*e.g.*, JRA-55 for 1958 - 2016 or concatenating the ERA-40 and ERA-Interim datasets). Care is required to distinguish the QBO signal from the ENSO signal.

- We recommend that reanalysis centres include 15 hPa and 40 hPa levels as standard output levels. The QBO amplitude peaks at 15 hPa in the FUB data, so model-reanalysis comparisons require this level for accurate validation of the models. The 40 hPa level, which is also in the FUB data, is highly correlated with the NH polar vortex response, and was the level at which the unusual easterly layer (the "QBO disruption") first emerged during 2015/16 NH winter.

12.1.8 Chapter 10: Polar Processes

In this chapter, we examined diagnostics of relevance to polar chemical processing and dynamics based on recent full-input reanalyses, including MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2. The selected diagnostics primarily target winter conditions. Observational datasets, reanalysis-driven CTM simulations, and operational analyses were also examined for some metrics. A summary evaluation of selected diagnostics examined in *Chapter 10* is provided in **Figure 12.10** as a quick reference to help users identify which reanalyses may be most suitable for a given issue related to stratospheric polar chemical processing. The key findings of this work, along with recommendations that follow from them, are summarized below.

Key findings of Chapter 10:

- In both polar regions, differences between temperatures from recent full-input reanalyses display an annual cycle. Using ERA-Interim as a reference, time series (2008 - 2013) of the differences in lower stratospheric daily polar-cap temperatures between the other reanalyses and the reference showed mainly positive deviations in summer but mainly negative deviations in winter, with the largest differences reaching ~ 1 K in the Antarctic and ~ 0.5 K in the Arctic. Thus, intercomparisons of the same reanalyses could find temperature discrepancies of opposite sign, depending on the season being examined.

- Polar winter temperatures from recent full-input reanalyses are in much better agreement in the lower and middle stratosphere than were those from older reanalysis systems.

- In the Southern Hemisphere especially, a dramatic convergence toward better agreement between the reanalyses is seen after 1999.

Average absolute differences from the reanalysis ensemble mean (REM) in wintertime daily minimum temperatures poleward of 40°S have been reduced from over 3 K prior to 1999 to generally less than 0.5 K in the most recent decade, while average differences in the area with temperatures below PSC thresholds have been reduced from over 1.5% of a hemisphere to less than about 0.5%. Other polar temperature and vortex diagnostics suggest a more complex picture, showing similar improvements for some reanalyses but persistent differences for others. The convergence toward better agreement is less apparent in the Northern Hemisphere.

- For many polar temperature and vortex diagnostics, reanalyses generally agree better in the Antarctic, where winters tend to have similar duration and potential for polar chemical processing every year, and thus the sensitivity to differences in meteorological conditions among reanalyses is low. In contrast, the generally warmer and more disturbed vortex and large interannual variability of Arctic winters lead to conditions that are frequently marginal, and thus the sensitivity to reanalysis differences is high.

- Comparisons of polar-cap averaged diabatic heating rates[4] in the lower stratosphere show that MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 give consistent results for the climatology and day-to-day evolution at pressures greater than about 20 hPa and should generally be suitable for polar processing studies.

- Comparisons of ERA-Interim, MERRA, and MERRA-2 with long-duration balloon observations in the Antarctic show that they reproduce the temperature and horizontal wind fluctuations of the balloons at about the 30% level; thus a significant portion of the atmospheric gravity wave spectrum is not captured by the reanalyses.

- An evaluation of trajectory calculations from a Lagrangian transport model using long-duration balloon observations in the Antarctic found typical error growth rates of 60 - 170 km day^{-1} over 15-day trajectories for a subset of full-input reanalyses.

- Winter-long simulations from a chemistry transport model driven by different full-input reanalyses generally produce very similar results through most of the season for most species. However, substantial disparities between model runs are seen where composition gradients are largest. In particular, comparisons with satellite long-lived tracer measurements indicate that the model underestimates the strength of confined diabatic descent inside the winter polar vortex to varying degrees depending on the specific reanalysis used to force the model. As a consequence, considerable spread between the different simulations becomes evident by late winter.

- Estimates of chemical ozone loss based on satellite observations are relatively insensitive to the choice of reanalysis used to interpolate the measurements to isentropic surfaces and identify the vortex boundary. In contrast, chemical loss estimates based on simulated ozone fields from a chemistry transport model can differ substantially; a case study showed that forcing the model with different reanalyses yielded differences in the estimates of chemical ozone loss in the Antarctic vortex core as large as ~ 25 DU (20% - 30%).

Recommendations from Chapter 10:

- Any of the recent full-input reanalyses (MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2) can be suitable for studies of lower stratospheric polar processing. However, substantial differences between the various reanalyses are found in some instances; therefore, the choice of which reanalysis to use in a given study may depend on the specific science questions being addressed.

- Temperature biases in older meteorological reanalyses often rendered them unsuitable for accurately modeling interannual variability in PSC formation and consequent denitrification, chlorine activation, and chemical ozone loss; in particular, ERA-40, NCEP-NCAR R1, and NCEP-DOE R2 are obsolete and should no longer be used for studies of polar stratospheric chemical processing and dynamics.

- Because of the limitations of earlier reanalyses, it was not uncommon for modeling studies to try to match observed chlorine activation and/or ozone loss by imposing arbitrary systematic adjustments of 1 - 2 K or more on reanalysis temperatures. Increased confidence in the accuracy of current polar reanalysis temperatures provides tighter constraints on model parameterizations of microphysics/chemistry used to represent polar chemical processing. As a consequence, strong justification should be provided in modeling studies seeking to ascribe deficiencies in modeled chlorine activation and/or ozone loss to reanalysis temperature biases.

- Despite the overall good agreement between the polar temperatures from current full-input reanalyses, whenever feasible it is best to employ multiple reanalyses, even for studies involving recent winters for which differences between reanalyses are likely to be small; using more than one reanalysis allows estimation of uncertainties and the potential impact of those uncertainties on the results, especially for quantities that cannot be directly compared with observations.

[4] See the footnote on diabatic heating rates in reanalyses in *Section 12.1.3*.

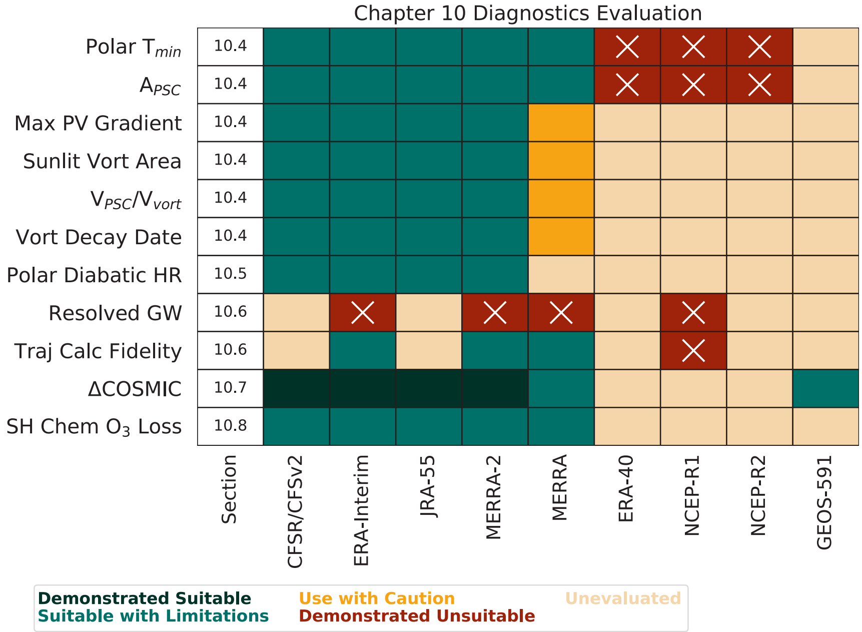


***Figure 12.10:** (Same as **Figure 10.26**.) A summary of the diagnostics evaluated in Chapter 10: Polar Processes. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. "Polar T_{min}" = minimum temperatures poleward of 40°; "A_{PSC}" = area of temperatures below PSC existence thresholds; "Max PV Gradient" = daily maximum gradients in potential vorticity, a measure of vortex strength; "Sunlit Vort Area" = area of the polar vortex in sunlight; "V_{PSC}/V_{vort}" = winter-mean volume of air with temperature below the nitric acid trihydrate PSC threshold, expressed as a fraction of the volume of air in the vortex; "Vort Decay Date" = the last day before which the vortex area is above 1 % of a hemisphere continuously for 30 days; "Polar Diabatic HR" = Diabatic heating rates in the polar vortex region; "Resolved GW" = resolved atmospheric gravity wave spectrum; "Traj Calc Fidelity" = fidelity of reanalysis-driven trajectory calculations from a Lagrangian transport model; "ΔCOSMIC" = differences between reanalysis and COSMIC GNSS-RO temperatures; "SH Chem O_3 Loss" = estimates of chemical loss in the Antarctic ozone hole from a chemistry transport model forced by reanalyses.*

- Reanalysis temperatures are generally unsuitable for assessment of trends in temperature-based diagnostics. Major changes in assimilated data inputs are often made at approximately the same time in all reanalyses, hindering determination of the impact of such changes through reanalysis intercomparisons. Caution is especially advised for the estimation of trends in diagnostics that aggregate low temperatures over months and/or vertical levels in the Northern Hemisphere, such as the winter-mean fraction of the vortex volume with air cold enough for PSCs to exist; such diagnostics are particularly sensitive to the specific PSC threshold chosen, which is subject to non-negligible interannual variability.

12.1.9 Chapter 11: Upper Stratosphere and Lower Mesosphere

In this chapter, we examined differences among reanalyses in the upper stratosphere and lower mesosphere among full-input reanalyses that provide data in this part of the atmosphere (MERRA, MERRA-2, ERA-40, ERA-Interim, JRA-25, JRA-55, and CFSR/CFSv2). A summary assessment of the diagnostics examined in this chapter is provided in **Figure 12.11**. Researchers interested in exploring a particular phenomenon within the USLM should consult the appropriate section of the chapter before proceeding, as indicated in the first column of this figure. Key findings and recommendations from this chapter are outlined below.

Key findings of Chapter 11:

- Differences among the reanalyses 1) decrease with time due to improvements in assimilated observational data, 2) increase with altitude due to differences in model top, sponge layers, and gravity wave drag treatments, and 3) increase nearer the Equator where sparse observations leave key dynamical phenomena largely unconstrained.
- Although no single reanalysis system is clearly better in representing all aspects of the USLM, higher-top systems such as MERRA and MERRA-2 are essential for capturing mesospheric circulation features such as the SAO and the QTDW.
- Differences in the satellite data assimilated into reanalyses as a function of time introduce discontinuities in both basic state variables and higher order diagnostics. This precludes trend studies based on a single reanalysis system.
- Differences in temperature among the reanalyses increase with height into the mesosphere at all latitudes. Likewise the inter-reanalysis differences in zonal wind increase with height especially in the equatorial region.

- Seasonal mean temperature differences defined with respect to MERRA are larger in older reanalyses (ERA-40 and JRA-25) and smaller in newer reanalyses (MERRA-2, ERA-Interim, and JRA-55).
- Westerly and easterly jets in the winter and summer stratosphere, respectively, are well reproduced in all of the evaluated reanalyses.
- The descending branch of the residual circulation in the winter stratosphere is strongest in MERRA, consistent with results prepared for *Chapter 5* (not shown; *Thomas Birner, personal communication*, 2021).
- Anomalous vertical temperature gradients around 3 hPa in JRA-25 lead to anomalous flow in the winter stratosphere. These features are not observed in the other reanalyses.
- Noisy meridional and vertical winds in ERA-40 cause larger dispersion of air parcels, which leads to "younger" age of air values and a weaker subtropical barrier in the stratosphere.
- Throughout the year, MERRA-2 has weaker cross-equatorial flow, a weaker middle-atmosphere Hadley circulation, and a westerly bias in the tropical USLM compared to ERA-Interim, JRA-55, and MERRA.
- Signatures of long-term variability due to the ENSO, the QBO, the 11-year solar cycle, and volcanic eruptions are evident in JRA-55, MERRA-2, and ERA-Interim; however, there are substantial differences among these reanalyses in the USLM, especially at equatorial latitudes.
- The mean SAO amplitude is reasonable in ERA-Interim, JRA-55, MERRA, and MERRA-2; comparison between JRA-55 and JRA-55C highlights the crucial role of assimilating satellite temperatures for accurately representing the SAO.
- The spatial patterns and magnitudes of inertial instability frequency are in good agreement among MERRA, MERRA-2, ERA-Interim, and JRA-55.

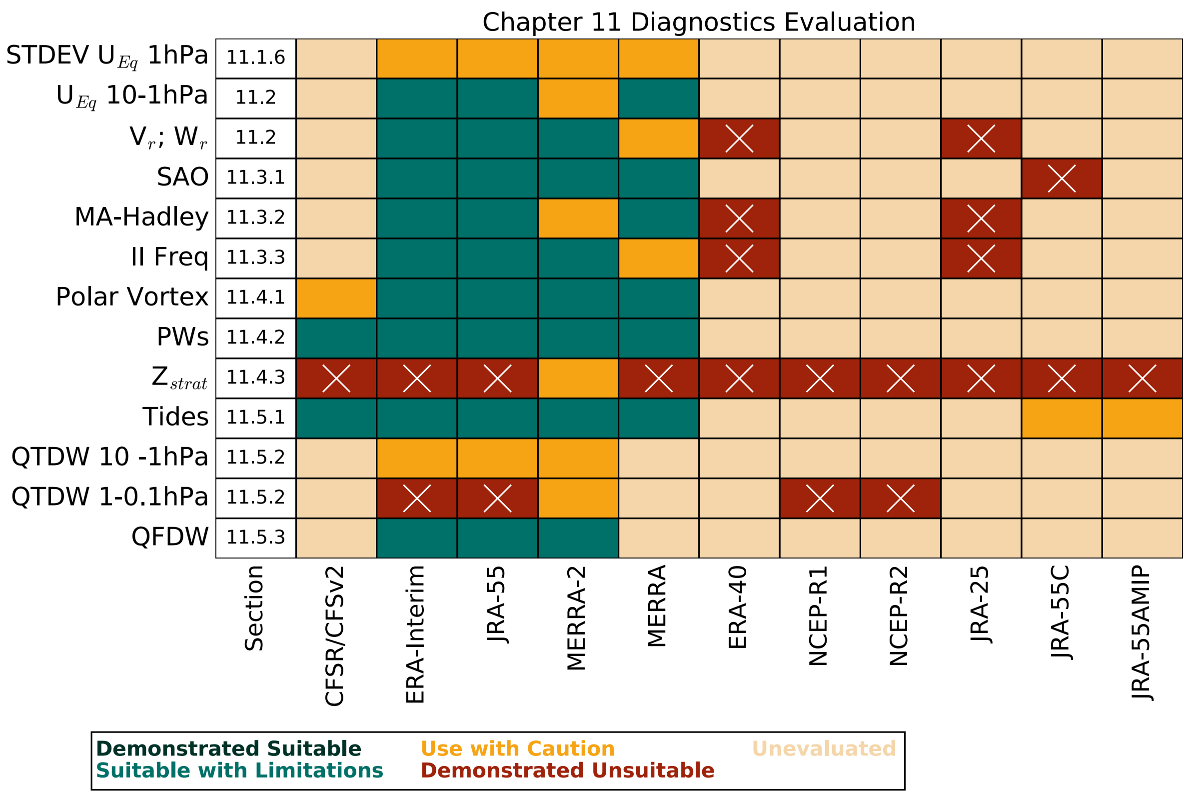


Figure 12.11: *(Same as **Figure 11.52**.) A summary of the diagnostics evaluated in Chapter 11: Upper Stratosphere and Lower Mesosphere. The "Section" column at left indicates where in the chapter each diagnostic is described. See the beginning of this chapter for the meaning of the evaluation terms. Note that the score corresponding to "Demonstrated Suitable" was not assigned to any of the diagnostics listed here, so the darkest green colour does not appear in this table. The full names of the abbreviated diagnostics can be found in the Chapter 11 sections and subsections. Briefly, "STDEV" = the standard deviation; "U_{Eq}" = the zonal wind at the Equator; "V_r" and "W_r" = the residual circulation meridional and vertical velocities, respectively; "SAO" = the Semi-Annual Oscillation; "MA-Hadley" = the middle-atmosphere Hadley circulation; "II Freq" = the occurrence frequency of inertial instability; "PWs" = planetary waves; "Z_{strat}" = the height of the stratopause with emphasis on elevated stratopause events; "QTDW" = the quasi-2-day wave; "QFDW" = the quasi-5-day wave.*

- MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 all capture multi-year winter mean polar vortex characteristics in both hemispheres, although CFSR wintertime vortex frequencies in the 50 ° to 70 ° latitude bands are 10 - 20 % lower in both hemispheres than those based on the other four reanalyses.

- MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 sufficiently capture the multi-year mean seasonal evolution of the polar vortex at the stratopause during 2005 - 2015 (interannual variability is not assessed).

- Quasi-stationary PW-1 amplitudes show remarkable agreement among the reanalyses and with MLS observations in the extratropics during winter; larger differences are seen at lower latitudes and during summer.

- Elevated Stratopause (ES) events represent strong, transient departures from climatological conditions in the Arctic USLM. These events are generally unconstrained by observations (with the exception of MERRA-2 which assimilates temperatures from Aura MLS after 2004). Their representation in all reanalyses depends strongly on the nature of the sponge layer in the forecast model used to produce the reanalysis, and thus cannot be regarded as trustworthy.

- While reanalyses reproduce the global patterns of the diurnal and semi-diurnal migrating tides, their amplitudes are underestimated by 20 - 50 % compared to SABER observations.

- The representation of the quasi-2-day wave is qualitatively similar in MERRA-2, ERA-Interim, and JRA-55, but with 50 % differences in amplitude.

- There is excellent agreement in the representation of the quasi-5-day wave among MERRA-2, ERA-Interim, and JRA-55, suggesting that the origin and propagation of this wave involve stratospheric processes that are well represented in these systems.

Recommendations from Chapter 11:

- Scientific studies using reanalyses in the USLM should make every effort to also include comparisons with independent observations. This imperative will require sustained engagement from reanalysis data users, new observational campaigns and operational measurement platforms for evaluation of reanalysis data, and a renewed commitment to replace the aging satellites currently relied upon for temperature and constituent observations of the middle atmosphere.

- Large discontinuities that occur due to differences in the data assimilation process preclude trend studies based on any single reanalysis system.

- There are large temperature and wind differences among the reanalyses in the tropical USLM. Using two or more reanalyses datasets to study phenomena (*e.g.*, the SAO, the diurnal tide) in this region of the atmosphere is recommended to increase confidence.

- There are large uncertainties in MERRA-2 zonal winds in the tropics; MERRA, ERA-Interim, and JRA-55 are in better agreement with each other up to 1 hPa than they are with MERRA-2.

- There are large uncertainties in "older" reanalysis datasets in the USLM; the meridional circulation in the stratosphere and mesosphere is more realistic in MERRA-2, ERA-Interim, and JRA-55 than in MERRA, ERA-40, and JRA-25.

- Both Eulerian-mean and residual-mean meridional flows in ERA-40 are noisier than those in the other reanalyses, thus, science studies based on ERA-40 residual circulation velocities would likely generate noisier results.

- JRA-55C is not suitable for studies of the SAO.

- Low polar vortex frequency of occurrence biases in CFSR/CFSv2, due to high polar temperatures and a weak polar night jet, render this reanalysis dataset less suitable for polar vortex studies compared to MERRA, MERRA-2, ERA-Interim, or JRA-55.

- MERRA, MERRA-2, ERA-Interim, JRA-55, and CFSR/CFSv2 are all suitable for studying quasi-stationary PW-1 patterns in the winter extratropics, but care should be exercised for studies focusing on the subtropics or the summer.

- Reanalyses should not be relied upon for studying ES events. Even for MERRA-2, the underlying forecast model does not capture the evolution of ES events correctly and so derived quantities (other than temperatures that are directly assimilated) should be treated with caution.

- Older reanalyses such as ERA-40 or JRA-25 are not suitable for tidal studies.

- Tidal results should not be extrapolated from one time to another as the representation of tides is sensitive to the satellite data assimilation.

- There are large uncertainties in using reanalysis data to study 5-day and 2-day wave normal modes; different reanalyses may yield different results.

12.2 Overall findings and reanalysis user recommendations

Several common findings and recommendations emerge from the detailed and extensive reanalysis comparisons described in *Chapters 3–11* and summarized above:

- All studies find substantial improvements in the most recent generation of reanalyses, even in cases where the older reanalyses are adequate for some diagnostics. We thus recommend that studies using full-input reanalyses be done with CFSR/CFSv2, ERA-Interim (and/or ERA5), JRA-55, and/or MERRA-2 rather than reanalyses from previous generations.

- In particular, NCEP-NCAR R1 and NCEP-DOE R2 are inadequate for many diagnostics. These reanalyses are deprecated based not only on the findings of this activity, but also on a wealth of comparisons spanning more than a decade. With the availability of modern reanalyses providing coverage of the pre-satellite (*e.g.*, JRA-55 and ERA5) and the ability to obtain those reanalyses on coarser grids, there should be no reason to continue using these older reanalyses.

- A number of studies find deficiencies in CFSR/CFSv2 relative to its peers (ERA-Interim, JRA-55, MERRA-2), and in some cases the changes between CFSR and CFSv2 are sufficiently large that it may not be appropriate to use these two datasets as if they were continuous.

- Several studies have shown that valuable information on the pre-satellite era can be obtained from conventional-input reanalyses that do not assimilate satellite data. For such studies, it is essential to first assess how the diagnostics in question compare with the full-input reanalysis from the same system during the satellite era.

- The vast majority of studies described herein found scientific benefit in using multiple reanalyses and comparing the results. This type of approach is especially important for diagnostics that cannot be directly compared with observations. In cases where the reanalyses agree well, results based on multiple reanalyses still provide valuable uncertainty estimates.

- All reanalyses show some level of discontinuities related to major changes in data inputs, a key example being changes associated with (and improvements in reanalysis agreement after) the switch from TOVS to ATOVS around 1998/1999. As different assimilation systems handle these changes in different ways, the impacts also differ across reanalyses.

- While several studies reported herein show valuable information obtained from studying trends in reanalysis data, great caution should be used in conducting such studies, not least because of the impacts of observing system changes as mentioned in the previous point. Trend studies should always compare multiple reanalyses, and consistency among results for multiple reanalyses should be viewed as a necessary but not sufficient condition for robustness.

- Many of the studies described herein benefitted from using the highest vertical resolution available and, for some (especially studies of conditions at and around the tropopause), this high vertical resolution proved critical. We thus recommend using reanalysis products on model levels in all analyses for which sharp vertical gradients or fine-scale vertical features may be important.

- Several quantities (notably diabatic heating rates[5], ozone and water vapour, and products related to clouds and convection) are handled and reported very differently across different reanalyses. Careful consideration of how the individual products are produced is necessary when using and comparing them.

- Several chapters have emphasized the importance of continuing data records (especially satellite trace gas data), for which ongoing records are in jeopardy due to aging instruments and an uncertain commitment to future missions. These data are essential benchmarks for evaluating reanalysis data both directly (validation of reanalysis products) and indirectly (evaluation of reanalysis-driven or nudged CTM and CCM simulations). In addition, several currently available homogenized satellite datasets have been shown to provide important improvements in reanalysis products when assimilated, so continuing (and improving upon) records such as these should be a priority.

12.3 Recommendations for improving reanalyses and their evaluation

One important aspect of the S-RIP activity was the involvement of reanalysis centres, as well as the continuing dialog between representatives of these centres and the reanalysis data users who conducted studies for S-RIP. A number of recommendations for future work have emerged from these interactions, including recommendations related to future reanalysis development, improvements in the output products, data formats, or grids, and the need for further observations both for assimilation into reanalysis systems and for evaluation of reanalysis products.

[5] See the footnote on diabatic heating rates in reanalyses in *Section 12.1.3*.

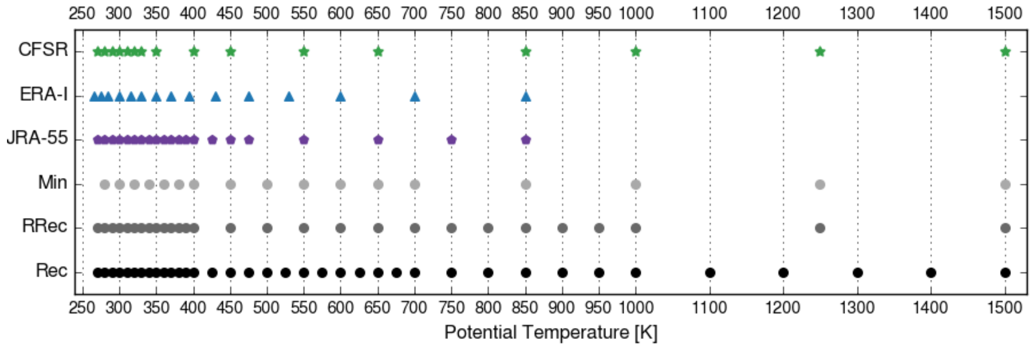


Figure 12.12: *Current and proposed isentropic levels from the surface through the midstratosphere. The following three sets were proposed for the survey (* indicates levels that are above the top of some of the most recent models/analyses).*
Min: *280, 300, 320, 340, 360, 380, 400, 450, 500, 550, 600, 650, 700, 850, 1000, 1250, 1500, 1750, 2000, 2500, 3000, 3500*, 4000* K.* ***RRec:*** *270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 450, 500, 550, 600, 650, 700, 750, 800 ,850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2500, 3000, 3500*, 4000* K.* ***Rec:*** *270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2250, 2500, 3000, 3500*, 4000* K. Based on the results of this survey, we recommend RRec (the middle resolution).*

12.3.1 S-RIP survey results and related product needs from S-RIP studies

To help clarify the output product format needed by reanalysis users, we conducted several surveys related to the adequacy of currently available products and output grids. The results of these surveys are summarized briefly below. The number of respondents was 28. Overall, 63 % of respondents expect to need reanalysis data at or near the resolution of the native model grids, while 74 % of respondents need data either on model levels or on isobaric or isentropic grids that are finer or more extensive than currently available. Detailed surveys on user needs for data on isentropic and isobaric levels resulted in the following:

Isentropic Levels:

- Approximately 70 % of respondents need data on isentropic surfaces.
- Of those, 82 % say the currently available levels are inadequate.
- Of the three sets of levels we proposed (**Figures 12.12** and **12.13**), 28 % say they need the finest resolution, 44 % the medium resolution, and 28 % the coarsest resolution.
- **Products most needed on isentropic surfaces:**
  - › Pressure / Temperature: 100 %
  - › Potential Vorticity: 94 %
  - › Zonal and Meridional Winds: 89 %
  - › Ozone mixing ratio: 66 %
  - › Specific Humidity: 50 %
  - › Montgomery Streamfunction: 44 %

The grey dots in **Figures 12.12** and **12.13** show the three sets of common isentropic levels we proposed for the survey. Based on the survey results, among Min, RRec, and Rec (see **Figures 12.12** and **12.13** for their definitions), we recommend the following set RRec:

RRec: 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2500, 3000, 3500*, 4000* K

where * indicates levels that are above the top of some of the most recent models/analyses.

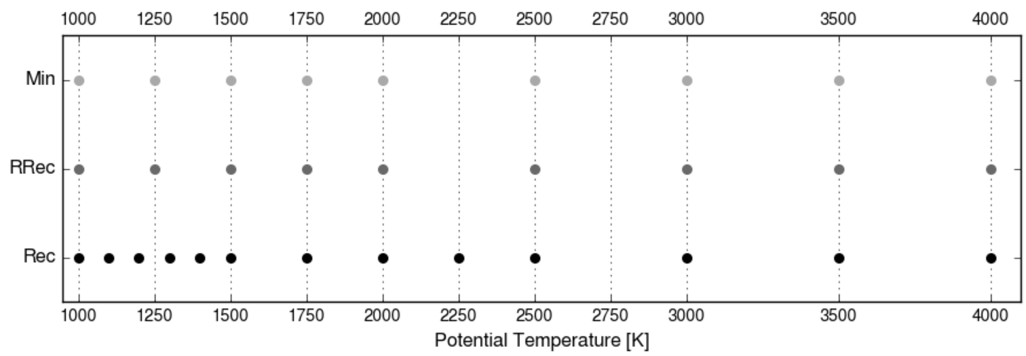


Figure 12.13: *As for* ***Figure 12.12****, but for proposed isentropic levels in the USLM. Based on the results of our survey, we recommend RRec (the middle resolution).*

Pressure Levels:

- Approximately 81 % of respondents need data on pressure surfaces.
- Of those, 84 % say the currently available levels are inadequate.
- 95 % say the proposed additional levels (see below) would be useful to them.

The standard ERA-Interim output diagnostic levels are:

1000, 975, 950, 925, 900, 875, 850, 825, 800, 775, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 225, 200, 175, 150, 125, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, 1 hPa

Other recent full-input reanalyses use very similar pressure-level grids for data distribution.

We propose additional levels at 85, 60, 40, and 15 hPa (to improve resolution in the vicinity of the tropical tropopause and the QBO) and 0.7, 0.3, 0.1, 0.03, and 0.01 hPa (to improve coverage of the USLM).

In summary, our recommendation on common pressure levels for future reanalysis products is as follows:

1000, 975, 950, 925, 900, 875, 850, 825, 800, 775, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 225, 200, 175, 150, 125, 100, 85, 70, 60, 50, 40, 30, 20, 15, 10, 7, 5, 3, 2, 1, 0.7, 0.3, 0.1, 0.03, 0.01 hPa

Figure 12.14 illustrates the vertical grid spacing for these requested pressure levels, as well as vertical grid spacings for the current standard pressure levels and the model levels used in producing ERA-Interim and ERA5. Other levels suggested by survey respondents but not represented in our recommendation are 80 hPa for UTLS studies and 0.5 hPa and 0.2 hPa for USLM studies.

In addition, both the surveys conducted and many of the S-RIP studies reported herein suggest community needs for:

- Diabatic heating rates from all physics on model grids, with all reanalyses reporting a consistent minimum product. Currently some reanalyses report only LW and SW heating rates on model levels, whereas others also report heating rates from all physics. There are also other differences in how the reanalysis provide diabatic heating rates that make them difficult to use and compare. Diabatic heating rates are critical for transport studies, especially in the upper troposphere and stratosphere. For consistency and comparability, it would be helpful for future reanalyses to provide (on model levels) temperature tendencies from (1) all physics, (2) all-sky radiation, and (3) clear-sky radiation, with the latter two provided separately for the LW and SW components. Additional terms (*e.g.*, convection, large-scale condensation, turbulence, assimilation, *etc.*) are also valuable for evaluating individual reanalyses and conducting scientific studies, and we suggest that these terms be provided as computational resources and model formulation permit.
- Integration with satellite simulators where possible. The inclusion of the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) in MERRA-2 provided valuable context not only for observational validation, but also for understanding differences between MERRA-2 and other reanalysis cloud products. As these simulators and their use in climate model evaluation expands, their application to reanalysis products becomes increasingly relevant. The provision of model-resolution reanalysis outputs at high frequency is a welcome step toward facilitating offline application of satellite simulators. However, computational resources permitting, full integration within the reanalysis model would go a long way toward enabling wider and more effective use of these tools.

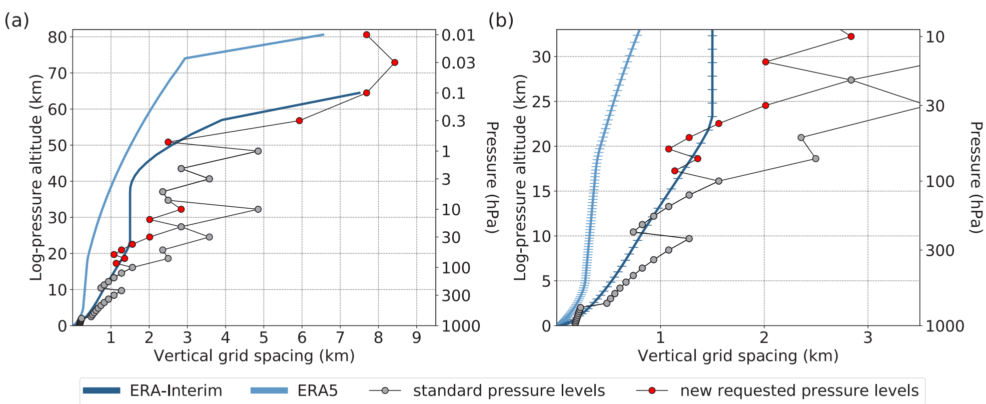


__Figure 12.14:__ Vertical profiles of vertical grid spacing for the requested pressure levels (see text), as well as vertical grid spacings for current standard pressure levels (from ERA-Interim; see text) and model levels (from ERA-Interim and ERA5). Panel (a) shows vertical grid spacing from the surface to 0.01 hPa (illustrating the proposed extension of the vertical grid), while panel (b) provides a zoomed view from the surface to 10 hPa (illustrating the requested finer resolution around the tropical tropopause and the lower part of the QBO).

- Information on uncertainty estimates for reanalysis products (especially basic fields such as temperatures and winds). Despite recent advances in this regard (*e.g.*, the ensemble of data assimilations produced as part of ERA5), such estimates remain problematic to produce for complex data assimilation systems. It may therefore be a useful goal for continuing S-RIP efforts to produce such estimates based on intercomparisons.

- Availability of a common data format for all reanalyses. This would be in line with current practice in the climate modelling community (*e.g.*, CMIP), for which a common data format has proved invaluable for intercomparison studies. Adoption of community standards for variable and file metadata in tools for preprocessing reanalysis data prior to download (see below) would boost the effectiveness of these tools and support the evaluation and intercomparison of future reanalysis products.

12.3.2 Data access issues

Ease of access to reanalysis datasets on multiple grids has improved greatly over the years of the S-RIP activity. We note and commend recent efforts to improve accessibility to both model-level and native-grid products for studies where resolution is critical and reduced-resolution products for cases where resolution is not critical and disk space or bandwidth are limiting factors. Nevertheless, the ever-increasing data volume for new reanalyses remains the largest current and future challenge to data access, as illustrated by difficulties in obtaining, storing, and processing ERA5 data at high resolution and on model levels. It will be essential to devise solutions for these challenges, not least in light of the numerous cases documented in this report for which the high resolution that engenders such large file sizes proved important both for fair evaluation of the reanalyses and for clarifying understanding of the diagnostics under evaluation.

We envision solutions for this issue taking multiple forms, from simple improvements in procedures and infrastructure to extensive investment in distributed processing and server-side applications. Developments in the latter direction have been extremely valuable for the S-RIP activity, as some reanalysis centres have begun dedicating computational resources for users to conduct simple preprocessing steps (*e.g.*, regridding, subsampling, and temporal averaging) prior to downloading data. These tools reduce the computational overhead for reanalysis data users, thus speeding up analysis and allowing access to a more complete set of reanalysis products for cross-validation and hypothesis testing. Particularly useful features include options for remapping data onto user-selected grids, subsetting regions or variables, and daily averaging. We express here our appreciation for the resources and hard work that reanalysis centres and their employees have put into making these tools available, as well as our wholehearted support for further investment in this direction.

Many participants of S-RIP have also benefitted from a designated group workspace on the JASMIN "super-data-cluster" in the United Kingdom. Funded by the National Environment Research Council (NERC) and the UK Space Agency, this platform provided data storage and analysis tools that facilitated some of the more computationally intensive tasks undertaken by S-RIP participants. Resources permitting, further investment in the server-side tools provided by reanalysis centres might adopt some of the capabilities of this type of group workspace, such as temporary storage of intermediate products and/or more flexible pre-processing tools. Such developments would be invaluable for making new reanalyses accessible to a wider community of data users.

Helpful steps for improving data access can also be taken without requiring large investments of funds or computational resources. For example, standard sets of metadata to support widely used scripting tools (*e.g.*, parameter tables or grid description files for use with the Climate Data Operators developed at the Max-Planck-Institut für Meteorologie) could be used to construct 'recipes' for users to convert data from a more concise, centre-preferred format (*e.g.*, GRIB) to a more verbose and user-friendly format (*e.g.*, CF-compliant NetCDF4 with standard naming conventions). Such recipes could be organized by and provided together with pre-defined data collections (*e.g.*, upper-air analysis, forecast diagnostics, *etc.*), as was previously done for some reanalyses using GrADS control files. Regular testing of output data against common data processing tools and manipulations (*e.g.*, remapping, area selection, merging or averaging in time) would also be helpful. Often small adjustments to the grid description or other aspects of the file metadata are all that is needed to ensure compatibility with a wide range of software tools for climate data analysis. Although these steps cannot address barriers associated with computational overhead, they can substantially reduce the 'learning curve' for users interested in adopting and applying a new reanalysis dataset. Community resources like S-RIP (presuming it continues in some form) and reanalyses.org can also play valuable roles in creating, updating, and distributing these types of tools.

12.3.3 Documentation issues

It is critical for information on the models and assimilation systems to be kept current and accessible. In the past, documentation for reanalyses has often been sparse, out-of-date, difficult to find, or all of the above. For some centres (notably ECMWF), this situation has improved in recent years. It is important to have information available both on the ideas and assumptions behind the original model schemes (generally accessible in some form now, though not always easy to find), and on how those schemes have evolved since their original publication, in some cases 20 - 30 years ago (generally not available now).

We hope that one legacy of S-RIP will be to provide a model for immediately, consistently, and systematically documenting each new reanalysis, and for bringing and keeping documentation on existing reanalyses up to date. The detailed information presented in *Chapter 2: Description of the Reanalysis Systems* could serve as a template in this regard.

12.4 Prospects for the future

S-RIP was originally planned to continue until 2018 (*i.e.*, 5 years starting from the Planning Meeting in 2013). However, a fundamental goal of S-RIP is to provide well-organized feedback to the reanalysis centres, thus forming a "virtuous circle" of assessment, improvements in reanalyses, further assessment, and further improvements in reanalyses. To this end, calculations of diagnostics suited to numerous types of studies have been and are being developed for current reanalyses. These diagnostics can then be easily extended and applied to the assessment of future reanalyses. Since most reanalysis centres have ongoing programmes to deliver new and improved reanalyses, it may be valuable to continue S-RIP beyond this initial period of 8 years. The SPARC SSG meeting in 2022 will therefore provide a critical opportunity for that body to review the value of S-RIP, with input from the reanalysis centres and atmospheric science and climate researchers, and discuss how the continuing goals of systematic evaluation of reanalyses can be supported into the future.

Regardless of the future development of S-RIP, it is important for this project to leave a lasting legacy through publication of its report that helps to sustain international interest in the assessment of reanalyses. A primary goal of the project is to establish tighter links between reanalysis providers and SPARC-related researchers. It is thus hoped that outcomes from the S-RIP assessment will facilitate and even drive future reanalysis developments in a systematic, standardised way, in place of the ad hoc approaches that have been used previously. A further legacy will be the creation of public archives (at BADC/CEDA and NOAA; see *Chapter 1*, *Section 1.5*) of processed reanalysis data with standard formats and resolutions, which will help to enable both further intercomparisons and scientific analyses without repetition of expensive pre-processing steps. This ensemble of derived data sets is freely available to researchers worldwide, and is intended to be a useful tool for reanalyses assessment beyond the lifetime of the project.

Acknowledgements

We acknowledge the scientific guidance and sponsorship of the World Climate Research Programme, coordinated in the framework of SPARC. We are grateful to Theodore Shepherd and Greg Bodeker (past SPARC co-chairs) for strong encouragement and valuable advice during the proposal phase of the project during 2011 - 2012, and Joan Alexander (serving SPARC co-chair during 2013 - 2015), Neil Harris (serving SPARC co-chair since 2014), Judith Perlwitz (serving SPARC co-chair during 2016 - 2020), and Seok-Woo Son (serving SPARC co-chair since 2019) for continued support and encouragement. We are also very grateful to the review handling editors of this report, Judith Perlwitz, Seok-Woo Son, Vincent-Henri Peuch, Karen Rosenlof, and Gufran Beig, and to the reviewers. The Information Initiative Center of Hokkaido University, Japan, has hosted the S-RIP web server since 2014. We thank the reanalysis centres for providing their support and data products. The British Atmospheric Data Centre (BADC) of the UK Centre for Environmental Data Analysis (CEDA) has provided a virtual machine for data processing, a group workspace for data storage, and public data archive sites. We thank Yulia Zyulyaeva for contributions to S-RIP as an original member of the working group, as a chapter colead through October 2014, and as the designer of the S-RIP logo. We thank Diane Pendlebury for contributions to S-RIP as a chapter co-lead through August 2015. We thank Quentin Errera, the lead of the SPARC Data Assimilation working group, for co-organizing the 2014, 2015, 2016, and 2017 S-RIP workshops. We also thank John McCormack for co-organizing the 2017 workshop. Travel support for some participants of the 2013 planning meeting, the 2014, 2015, 2016, and 2017 workshops, and the 2018 chapter-lead meeting was provided by SPARC. We thank Peter Haynes and Gabriele Stiller for their work as guest editors for the special issue "The SPARC Reanalysis Intercomparison Project (S-RIP)" in the journals Atmospheric Chemistry and Physics (ACP) and Earth System Science Data (ESSD); we offer special thanks to William Lahoz (1960 - 2019) for his role as the other guest editor of this special issue, which he performed excellently even as he battled his final illness. We thank the hard works at the SPARC Office, in particular Mareike Heckl, Brigitte Ziegele, and Sabrina Zechlau for their typesetting of the report. Work at the Jet Propulsion Laboratory, California Institute of Technology, was carried out under a contract with the National Aeronautics and Space Administration.

Major abbreviations and terms

20CR	20th Century Reanalysis of NOAA and CIRES
3D	three dimensional
AIRS	Atmospheric Infrared Sounder
Alt	altitude
AMSU	Advanced Microwave Sounding Unit
AoA	Age of Air
A_{PSC}	area of temperatures below PSC existence thresholds
ATMS	Advanced Technology Microwave Sounder
ATOVS	Advanced TIROS Operational Vertical Sounder
BADC	British Atmospheric Data Centre
BDC	Brewer-Dobson Circulation
Calc	Calculation
CCM	Chemistry-Climate Model
CDR	Climate Data Record
CEDA	Centre for Environmental Data Analysis
CF	Climate and Forecast
CFMIP	Cloud Feedback Model Intercomparison Project
CFSR	Climate Forecast System Reanalysis of NCEP
CFSv2	Climate Forecast System, version 2
Ch	Channel (e.g., Ch1: Channel 1)
Chem	Chemical
CIRES	Cooperative Institute for Research in Environmental Sciences (NOAA and University of Colorado Boulder)
Clim	Climatology
CMIP	Coupled Model Intercomparison Project
COSMIC	Constellation Observing System for Meteorology Ionosphere and Climate
COSP	CFMIP Observation Simulator Package
CP	Cold Point
CPT	Cold-Point Tropopause
CTM	Chemistry-Transport Model
CWC	Cloud Water Content
DOE	Department of Energy
DU	Dobson unit
Dyn	Dynamical
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño–Southern Oscillation
E-P flux	Eliassen-Palm flux
EqL	Equivalent Latitude
ERA-20C	ECMWF 20th century reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA5	the fifth major global reanalysis produced by ECMWF
ERA-Interim (or ERA-I)	ECMWF interim reanalysis
ES	Elevated Stratopause
ExUTLS	Extratropical Upper Troposphere and Lower Stratosphere
FLXHR	Level 2B Fluxes and Heating Rates of CloudSat data product

FUB	Freie Universität Berlin
GEOS-5	Goddard Earth Observing System Model of the NASA, Version 5
GNSS-RO	Global Navigation Satellite System Radio Occultation
GPCC	Global Precipitation Climatology Centre
GrADS	Grid Analysis and Display System
GRIB	GRIdded Binary or General Regularly-distributed Information in Binary form
HCC	High Cloud Cover
HIRDLS	High Resolution Dynamics Limb Sounder
IAU	Incremental Analysis Update
ITCZ	Intertropical Convergence Zone
JASMIN	a data intensive supercomputer for environmental science at United Kingdom
Jet Rel	in coordinates relative to the subtropical jet core location
JRA-25	Japanese 25-year Reanalysis
JRA-55	Japanese 55-year Reanalysis
JRA-55AMIP	Japanese 55-year Reanalysis based on AMIP-type simulations
JRA-55C	Japanese 55-year Reanalysis assimilating Conventional observations only
LRT	Lapse-Rate Tropopause
LS	Lower Stratosphere
LW	Long-Wave
LWCRE	Long-Wave Cloud Radiative Effect
LZRH	Level of Zero net Radiative Heating
MA-Hadley	the middle-atmosphere Hadley circulation
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2
MERRA-2-ANA	MERRA-2 "analysis" data products that result directly from the Gridpoint Statistical Interpolation (GSI) analyses
MERRA-2-ASM	MERRA-2 "assimilation" data products that are the result of applying the Incremental Analysis Update (IAU)
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
MSLP	Mean Sea Level Pressure
MSU	Microwave Sounding Unit
MTp	Multiple Tropopause
NAO	North Atlantic Oscillation
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NERC	National Environment Research Council of the United Kingdom
NetCDF4	Network Common Data Form, Version 4
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
OLR	Outgoing Longwave Radiation
P	pressure
PSC	Polar Stratospheric Cloud
PV	Potential Vorticity
PW-1	Planetary Wave-1 (wavenumber one of planetary waves)
PWs	Planetary Waves
QBO	Quasi-Biennial Oscillation
QBO-E	QBO Easterly phase
QBO-W	QBO Westerly phase
QFDW	Quasi 5-Day Wave
QTDW	Quasi 2-Day Wave
R1 (or NCEP-R1)	NCEP-NCAR Reanalysis 1

R2 (or NCEP-R2)	NCEP-DOE Reanalysis 2
RCTT	Residual Circulation Transit Time
Rel	Relative
REM	Reanalysis Ensemble Mean
RO	Radio Occultation
RRec	the final recommended set of isentropic levels
SABER	Sounding of the Atmosphere using Broadband Emission Radiometry
SAO	Semi-Annual Oscillation
SASM	South Asian Summer Monsoon
SH	Southern Hemisphere
SPARC	Stratosphere-troposphere Processes And their Role in Climate
S-RIP	SPARC Reanalysis Intercomparison Project
SSG	Scientific Steering Group
SSU	Stratospheric Sounding Unit
SSW	Sudden Stratospheric Warmings
STDEV	Standard Deviation
STE	Stratosphere-Troposphere Exchange
SubV	subvortex
SW	Short-Wave
SWV	Stratospheric Water Vapour
TCO	Total Column Ozone
TIROS	Television Infrared Observation Satellite
T_{min}	minimum temperatures
TOVS	TIROS Operational Vertical Sounder
T	temperature
Tp	Tropopause
Traj	Trajectory
TTL	Tropical Tropopause Layer
U	zonal wind
U_{Eq}	the zonal wind at the Equator
USLM	Upper Stratosphere and Lower Mesosphere
UT	upper troposphere / upper tropospheric
UTLS	Upper Troposphere and Lower Stratosphere
V_{PSC}	winter-mean volume of air with temperature below the nitric acid trihydrate PSC threshold
V_{vort}	volume of air in the vortex
V_r	the residual circulation meridional velocity
W_r	the residual circulation vertical velocity
WV	water vapour
Yr	year
ZM	Zonal Mean
Z_{strat}	the height of the stratopause with emphasis on elevated stratopause events

SPARC Office
c/o Deutsches Zentrum für Luft-
und Raumfahrt e.V. (DLR)
Institut für Physik der Atmosphäre
Münchener Str. 20
D-82234 Oberpfaffenhofen
Germany
office@sparc-climate.org