WCRP Climate Model Intercomparisons: Where we are and how we got there

Peter J. Gleckler
Program for Climate Model Diagnosis and Intercomparison (PCMDI)

- A brief background/history of WCRP Model Intercomparison Projects (“MIPs”)
- The Coupled Model Intercomparison Project (CMIP5) and related MIPs
- The Earth System Grid Federation (ESGF)
- Climate model performance metrics

Brief history: WCRP climate model intercomparisons

1990-1995 **NO data standards!**

- WGNE Atmospheric Model Intercomparison Project (**AMIP**) megabytes
  - 30 Atmospheric GCMs perform a common experiment (prescribed SST & sea-ice)
  - Standard model output and "diagnostic subprojects"

1995-2000 **Limited data standards**

- **AMIP II** – tighter experimental protocol, more extensive diagnostics
- The Coupled Model Intercomparison Project (**CMIP**)

2000 - 2003

- The Coupled Model Intercomparison Phase II (**CMIP2**) gigabytes

A brief history of climate model intercomparison (ii)

Clearly defined data standards

2003 – Present

- The Coupled Model Intercomparison Phase III - CMIP3 terabytes

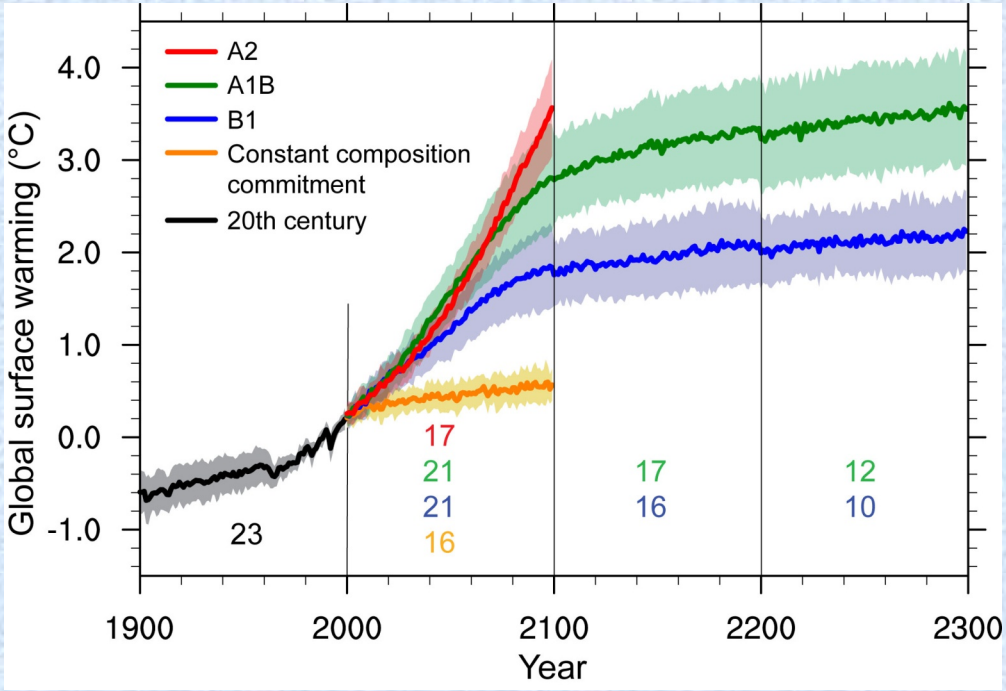
2009 – Present

- **The Coupled Model Intercomparison Phase V - CMIP5** **petabytes**

- Other, more scientifically focused MIPs are formally coordinated with CMIP5 (e.g., CFMIP and PMIP)

CMIP is overseen by the WGCM

CMIP and other model intercomparions facilitate much of the model-based research in IPCC assessments



[AR4 From Summary for Policy Makers]

- 4 of 7 figures in the AR4 “Summary for Policy Makers” are based on CMIP3
- Conclusions based on CMIP3 multi-model ensemble are more robust than on “anecdotal” conclusions from individual modeling studies
- Research from CMIP5 and related MIPs can be expected to be central to the AR5 and future climate change assessments

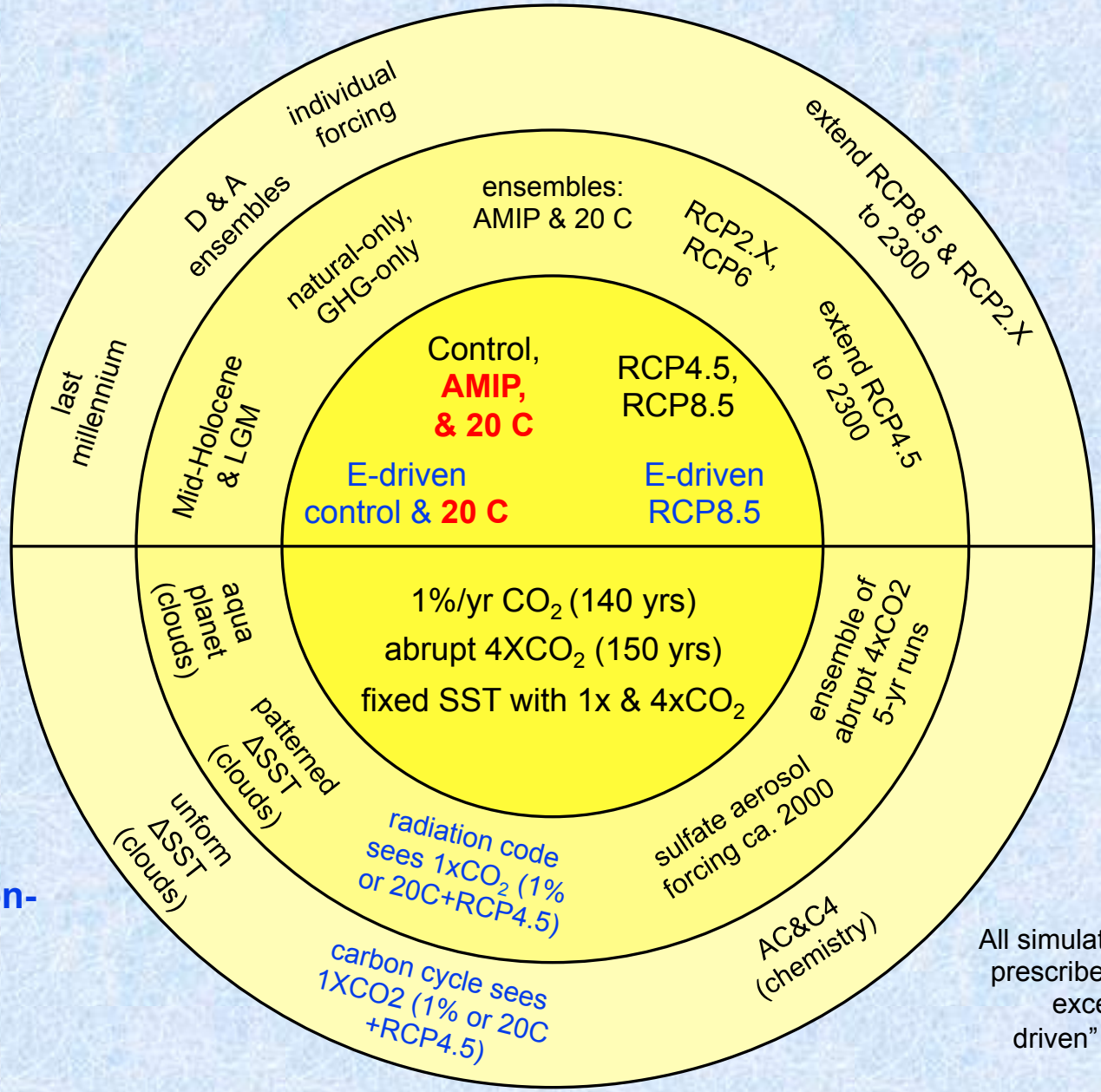
Advancement in coordinated model evaluation via MIPs

- Growth in data volume reflects a large increase in:
  - the number of users (~1000s),
  - complexity and assortment of experiments,
  - dramatic increase in fields/diagnostics being saved,
  - the amount/diversity of research being performed/published

- CMIP5 timelines
  - 2006-2009: Experimental design
  - 2010-2013: Modeling groups perform simulations
  - 2011- 2017+: Research

- Some early perspective on CMIP5:
  - 2006-2013: ~600 journal publications based on CMIP3
  - 2011-2013: ~250 journal publications based on CMIP5

CMIP5 Long-term Experiments

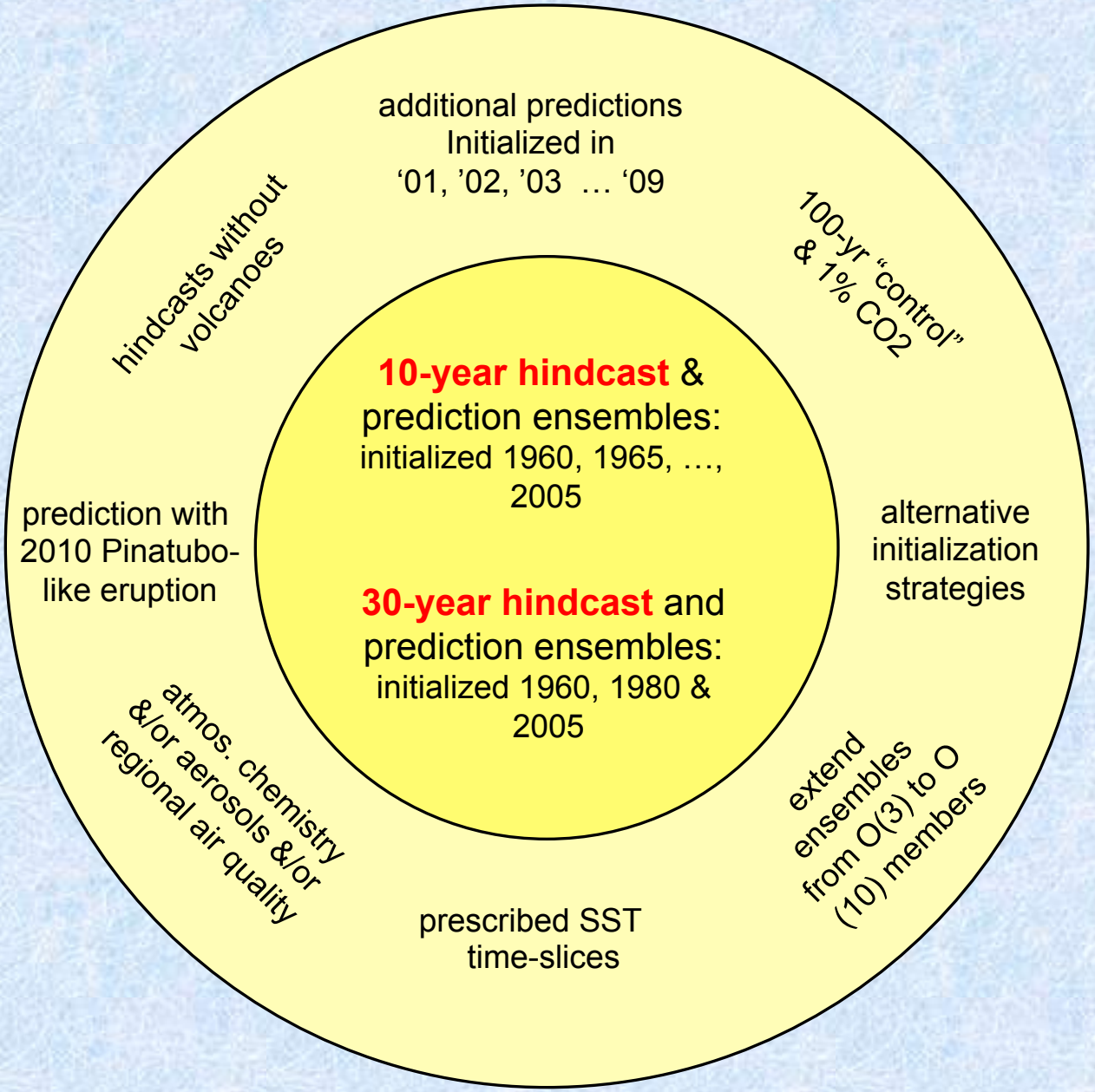


Best for Comparing with obs

Coupled carbon-cycle climate models only

All simulations are forced by prescribed concentrations except those "E-driven" (i.e., emission-driven).

CMIP5 Decadal Predictability/Prediction Experiments



Best for Comparing with obs

CMIP5 output fields cover all parts of the system and include “high frequency” samples.

- Domains (number of monthly variables*):
  - Atmosphere (60)
  - Aerosols (77)
  - Ocean (69)
  - Ocean biogechemistry (74)
  - Land surface & carbon cycle (58)
  - Sea ice (38)
  - Land ice (14)
  - CFMIP output (~100)
- Temporal sampling (number of variables*)
  - Climatology (22)
  - Annual (57)
  - Monthly (390)
  - Daily (53)
  - 6-hourly (6)
  - 3-hourly (23)

*Not all variables saved for all experiments/time-periods

http://cmip-pcmdi.llnl.gov/cmip5/output_req.html

Key to the success of CMIP3-5

- **Sustained support for critical infrastructure**
  - **Community-developed metadata conventions**
    - "Climate-Forecast" metadata convention (CF)
      http://cf-pcmdi.llnl.gov/
  - **Software to ensure data complies to conventions**
    - Climate Model Output Writer (CMOR)
      http://www2-pcmdi.llnl.gov/cmor
  - **Advancing state-of-the-art data delivery methods**
    - Earth System Grid Federation (ESGF)
      http:esgf.org/
- **A highly collaborative spirit within WCRP**

navigation

- Home
- Documents
- Conformance
- Discussion
- Governance
- Working Groups

CF Metadata

NetCDF Climate and Forecast (CF) Metadata Convention

The conventions for climate and forecast (CF) metadata are designed to promote the processing and sharing of files created with the NetCDF API. The CF conventions are increasingly gaining acceptance and have been adopted by a number of projects and groups as a primary standard. The conventions define metadata that provide a definitive description of what the data in each variable represents, and the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction, regridding, and display capabilities.

The CF conventions generalize and extend the COARDS conventions.

Here are the slides for a talk that provides an overview of CF. An expository version of this talk is in this article.

Discussion about CF Metadata takes place in two formats:

1. CF Metadata Trac, and
2. cf-metadata mailing list.

For further explanation of each of these, take a look at the Discussion page.

Quick Links

- CF Conventions Document
- CF Standard Name Table
- CF Metadata Trac
- cf-metadata Mailing List Archives
- CF Conformance Requirements & Recommendations
- CF Compliance Checker

Use of CF-conventions in CMIP5 (example)

```
netcdf rlut_Amon_MPI-ESM-LR_historical_r2i1p1_185001-200512 {
dimensions:
        time = UNLIMITED ; // (1872 currently)
        lat = 96 ;
        lon = 192 ;
        bnds = 2 ;
variables:
        double time(time) ;
                time:bounds = "time_bnds" ;
                time:units = "days since 1850-1-1 00:00:00" ;
                time:calendar = "proleptic_gregorian" ;
                time:axis = "T" ;
                time:long_name = "time" ;
                time:standard_name = "time" ;
        double time_bnds(time, bnds) ;
        double lat(lat) ;
                lat:bounds = "lat_bnds" ;
                lat:units = "degrees_north" ;
                lat:axis = "Y" ;
                lat:long_name = "latitude" ;
                lat:standard_name = "latitude" ;
        double lat_bnds(lat, bnds) ;
```

Crucial coordinate information

Use of CF-conventions in CMIP5 (ii)

Variable description

float rlut(time, lat, lon) ;

rlut:standard_name = "toa_outgoing_longwave_flux" ;

rlut:long_name = "TOA Outgoing Longwave Radiation" ;

rlut:comment = "at the top of the atmosphere (to be compared with satellite measurements)" ;

rlut:units = "W m-2" ;

rlut:cell_methods = "time: mean" ;

rlut:cell_measures = "area: areacella" ;

rlut:history = "2011-06-08T06:43:34Z altered by CMOR: replaced missing value flag (-1e+20) with standard missing value (1e+20)." ;

rlut:missing_value = 1.e+20f ;

rlut:_FillValue = 1.e+20f ;

rlut:associated_files = "baseURL: http://cmip-pcmdi.llnl.gov/CMIP5/ dataLocation gridspecFile: gridspec_atmos_fx_MPI-ESM-LR_historical_r0i0p0.nc areacella: areacella_fx_MPI-ESM-LR_historical_r0i0p0.nc" ;

Use of CF-conventions in CMIP5 (iii)

// global attributes:

:institution = "Max Planck Institute for Meteorology" ;

:institute_id = "MPI-M" ;

:experiment_id = "historical" ;

:source = "MPI-ESM-LR 2011; URL: http://svn.zmaw.de/svn/cosmos/branches/releases/mpi-esm-cmip5/src/mod; atmosphere: ECHAM6 (REV: 4603), T63L47; land: JSBACH (REV: 4603); ocean: MPIOM (REV: 4603), GR15L40; sea ice: 4603; marine bgc: HAMOCC (REV: 4603);" ;

:model_id = "MPI-ESM-LR" ;

:forcing = "GHG Oz SD SI VI LU" ;

:parent_experiment_id = "piControl" ;

:parent_experiment_rip = "r1i1p1" ;

:branch_time = 47481. ;

:contact = "cmip5-mpi-esm@dkrz.de" ;

:history = "Model raw output postprocessing with modelling environment (IMDI) at DKRZ: URL: http://svn-mad.zmaw.de/svn/mad/Model/IMDI/trunk, REV: 3209 2011-06-08T06:43:31Z CMOR rewrote data to comply with CF standards and CMIP5 requirements." ;

Not relevant for observations

navigation

- Home
- PCMDI Home Page
- News
- CDAT
- CMOR
- IPCC AR4 Model Data Portal
- About Us
- Newsletter

news

- CDAT Newsletter, June 2007 2007-06-26
- CDAT 4.1.2 Released 2006-06-07
- CDAT 4.0 Released 2005-11-21
- PCMDI Software Portal Released 2005-09-28
- CDAT 4.0 Beta Released 2005-09-28

More news...

Climate Model Output Rewriter (CMOR)

The "Climate Model Output Rewriter" (CMOR, pronounced "Seymour") comprises a set of C-based functions, with bindings to both Python and FORTRAN 90, that can be used to produce CF-compliant netCDF files that fulfill the requirements of many of the climate community's standard model experiments. These experiments are collectively referred to as MIP's and include, for example, AMIP, CMIP, CFMIP, PMIP, APE, and IPCC scenario runs. The output resulting from CMOR is "self-describing" and facilitates analysis of results across models.

Much of the metadata written to the output files is defined in MIP-specific tables, typically made available from each MIP's web site. CMOR relies on these tables to provide much of the metadata that is needed in the MIP context, thereby reducing the programming effort required of the individual MIP contributors.

CMOR2

- Download
- Documentation
- Tables
- Mailing list
- RELEASE NOTES

For questions concerning CMOR2, contact the cmor list (cmor@lists.llnl.gov).

Old CMOR (CMOR1)

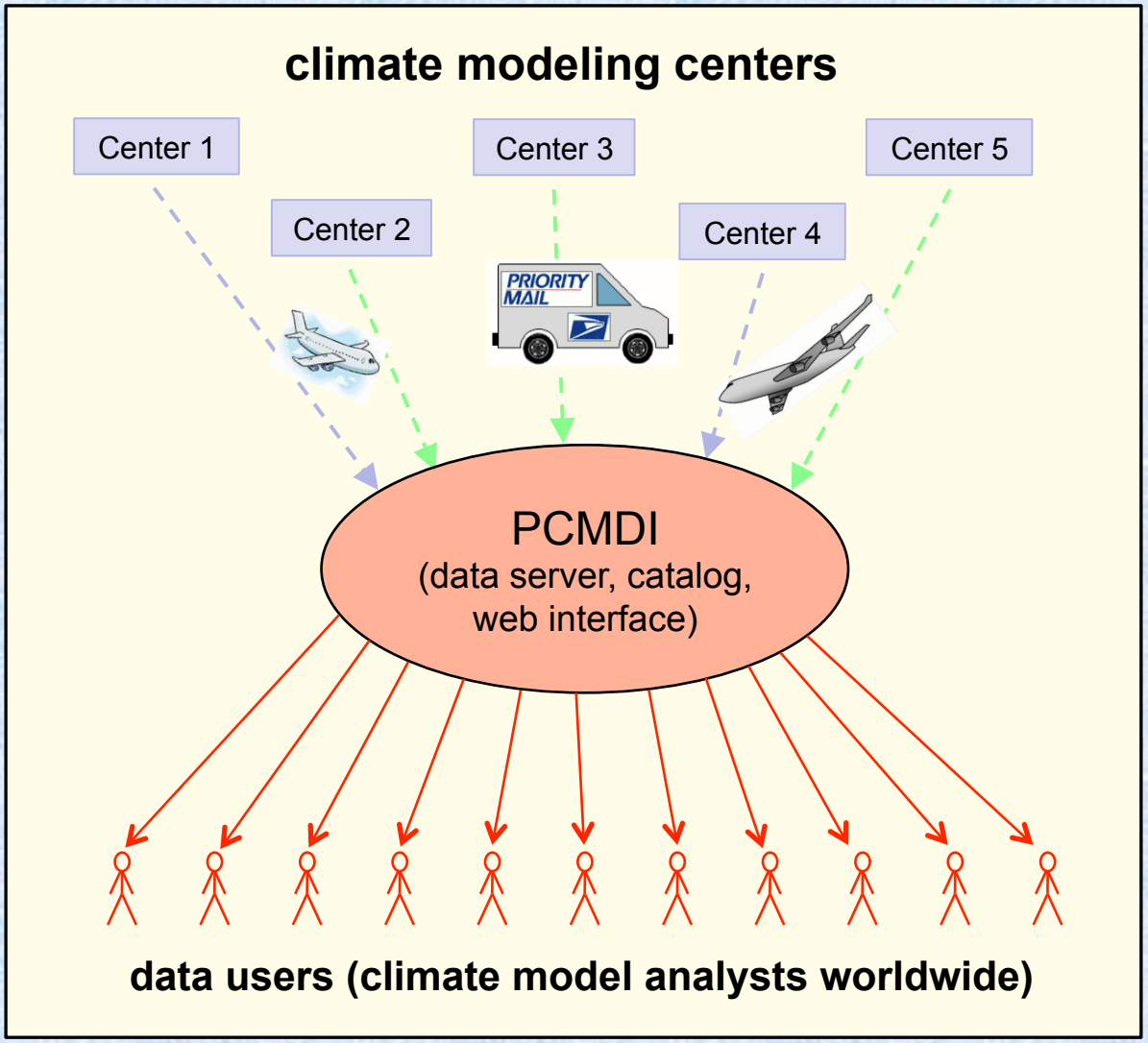
- Download
- Documentation

For questions concerning CMOR, contact Karl Taylor (taylor13@llnl.gov).

Why all this effort dedicated to organizing data?

- Documents a very large amount of information that is important climate modeling research
- A design target for software developers and climate scientists
- Community-based conventions and data delivery (next subject) greatly facilitates climate change research
- These standardizations facilitate reproducibility in metric and methods, crucial for ensuring the integrity of the climate analysis, but also will maintain them into the future, so that quality can be monitored

CMIP3 data handling: ESG* central archive at PCMDI



- Data shipped to PCMDI on hard disks
- Delayed availability
- Hindered corrections

- Search service via web gateway
- Download from single location (ftp, http, OPeNDAP)
- Fragile dependence on a single server.

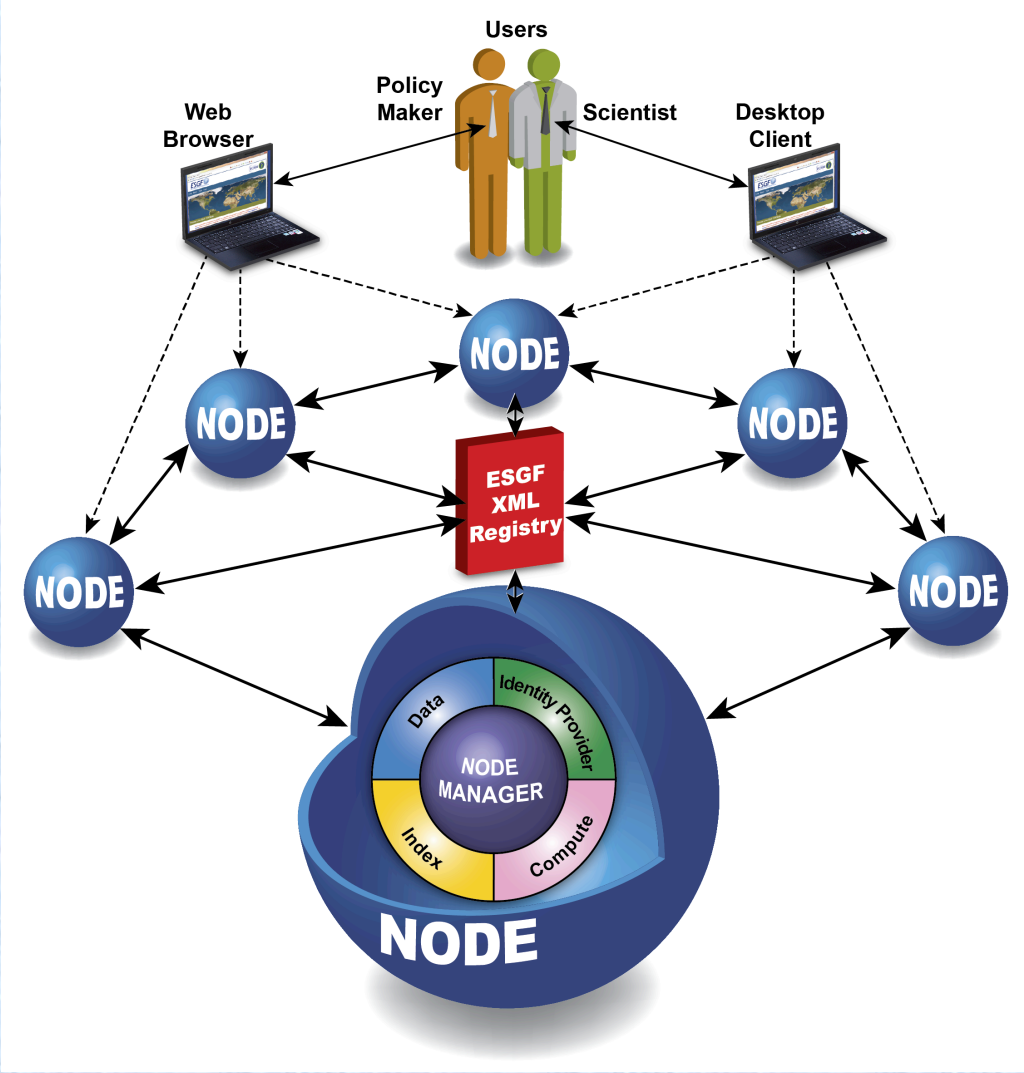
*ESG = Earth System Grid

A brief history of the Earth System Grid (ESG): ESG-I, ESG-II, ESG-CET, ESGF

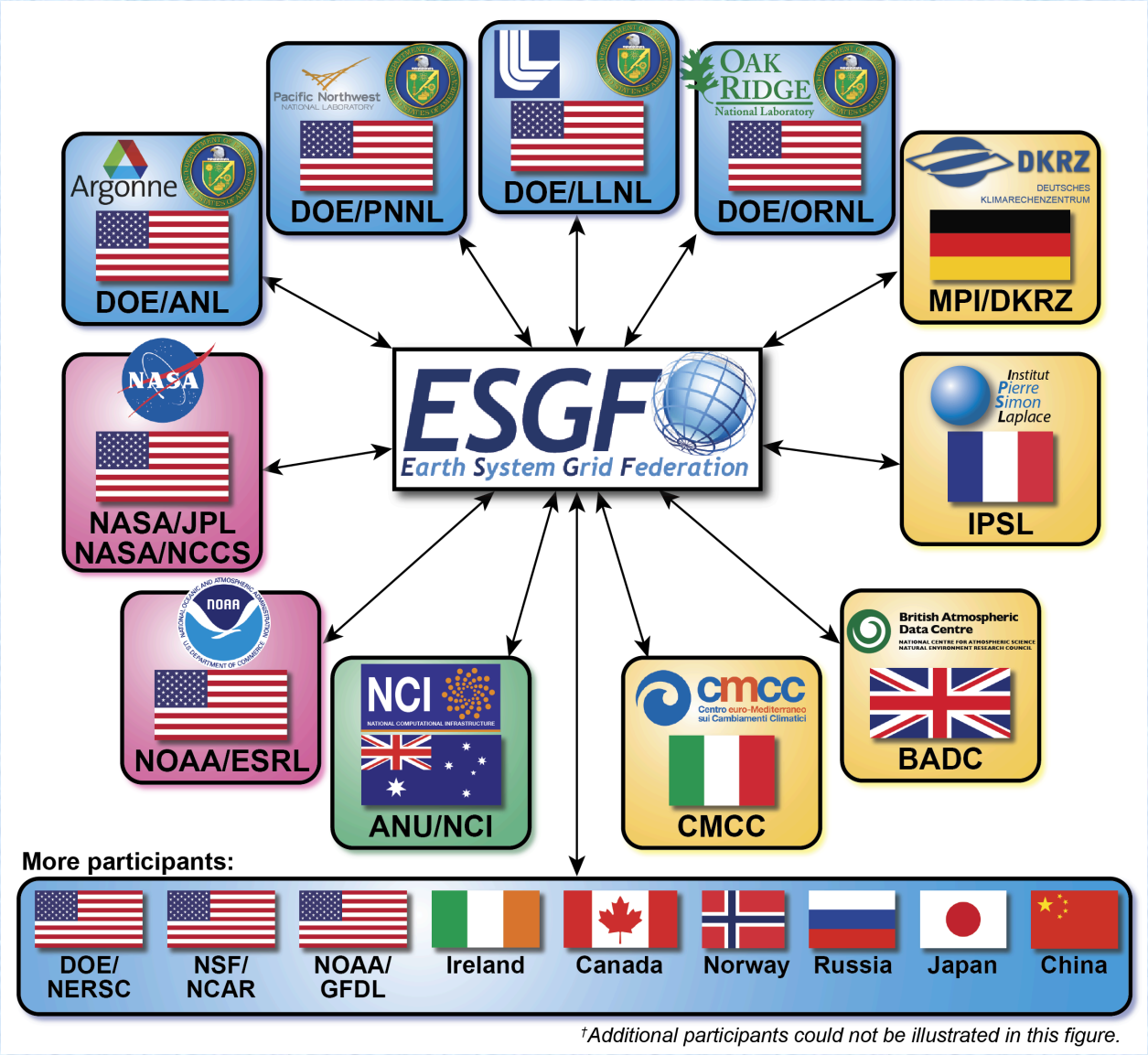
- **ESG-I** funded under DOE's Next Generation Internet (NGI) to address the emerging challenge of climate data 1999 – 2001 (ANL, LANL, LBNL, LLNL, NCAR, USC/ISI)
  - Data movement and replication; Prototype climate "data browser"; Hottest Infrastructure" Award at SC'2000.
- **ESG-II** funded under DOE's Scientific Discovery through Advanced Computing (SciDAC), turning climate data sets into community resources 2001-2006 (ORNL addition)
  - Web-based portal, metadata, access to archival storage, security, operational services, 2004 first operational portal CCSM (NCAR), IPCC CMIP3/AR4 (LLNL); 200 TB of data, 4,000 users, 130 TB served.
- **ESG-CET** funded under DOE's Offices of ASCR and BER to provide climate researchers worldwide with access to: data, information, models, analysis tools, and computational resources required to make sense of enormous climate simulation and observational data sets 2006 – 2011 (PMEL addition)
  - 2010 Awarded by American Meteorological Society (AMS) for leadership which led to a new era in climate system analysis and understanding.
  - CMIP3, CMIP5, CCSM, POP, NARCCAP, C-LAMP, AIRS, MLS, Cloudsat, etc.
  - 25,000 users, 500-800 users active per month, over 1 PB served
- **ESGF P2P** under the DOE's Office of BER, it is an open consortium of institutions, laboratories and centers around the world that are dedicated to supporting research of climate change, and its environmental and societal impact. (Additional U.S. funding from NASA, NOAA, NSF.) The federation includes: multiple universities and institution partners in the U.S., Europe, Asia, and Australia.

Courtesy D. N. Williams, AIMS

Federation connectedness means the user does not have to know where the data resides and critical data is replicated



ESGF is more than CMIP: federated and integrated data from multiple sources



†Additional participants could not be illustrated in this figure.

ESGF
Earth System Grid Federation

PCMDI

Home | Search | Tools | Login | Help

Welcome to this ESGF P2P Node

Quick Search

Keyword:

Search

Advanced Search (Category, Geospatial, Temporal, and more)...

Peer Nodes

- ANL Node
- BADC Node
- BNU Node
- CMCC Node
- DKRZ Node
- DKRZ CMIP5 Node
- NOAA-GFDL Node
- IPSL Node
- NASA-GSFC Node
- NASA-JPL Node
- NCI Node
- NERSC Node
- ORNL Node
- PCMDI Node

About esgf-pcmdi-9

The PCMDI mission is to develop improved methods and tools for the diagnosis and intercomparison of general circulation models (GCMs) that simulate the global climate. The need for innovative analysis of GCM climate simulations is apparent, as increasingly more complex models are developed, while the disagreements among these simulations and relative to climate obser vations remain significant and poorly understood. The nature and causes of these disagreements must be accounted for in a systematic fashion in order to confidently use GCMs for simulation of put ative global climate change.

Resources

Quick Links

- Create Account
- MyProxyLogon
- Expert Search (XML)
- Wget Script Generator
- ESGF aggregated RSS feed
- Contact ESGF

Instructions

- ESGF Full User Guide
- Search Help
- Search Controlled Vocabulary
- Wget Scripts FAQ
- Wget Scripting
- Tutorial: Download Strategies
- Using Globus Online
- Subscribing to RSS Notification

(supported browsers)

ESGF
Earth System Grid Federation

Home | Search | Tools | Login | Help

Welcome to this ESGF P2P Node

Quick Search

Keyword:

Search

Advanced Search (Category, Geospatial, Temporal, and more)...

Peer Nodes

- ANL Node
- BADC Node
- BNU Node
- CMCC Node
- DKRZ Node
- DKRZ CMIP5 Node
- NOAA-GFDL Node
- IPSL Node
- NASA-GSFC Node
- NASA-JPL Node
- NCI Node
- NERSC Node
- ORNL Node
- PCMDI Node

Guest User
ESGF P2P Version 1.4.2-brooklyn_heights-release-master [fe-2.3.2]

ESGF
Earth System Grid Federation

PCMDI

Home | Search | Tools | Login | Help

Current Selections
No search criteria selected

Search

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
To download data: add datasets to your Data Cart, then click on *Expand or wget*.

Temporal Search
Geospatial Search
Clear search constraints and datacart
Search Help
Search Controlled Vocabulary

Search All Sites ☐ **Show All Replicas** ☐ **Show All Versions**

Search Categories

PROJECTS

- C20C (1)
- CMIP5 (58044)
- COUND (4)
- CSSEF (461)
- CSSEF-ADHOC (3)
- EUCLIPSE (80)
- GeoMIP (581)
- LUCID (255)
- PMIP3 (158)
- TAMIP (1344)
- TEST (15)
- ana4MIPs (6)
- c-lamp (28)
- ccsm4 (2)
- cloud-cryo (20)
- euclipse (12)
- geomip (8)
- obs4MIPs (25)
- ultrahighres (2)

Institute
Model
SubModel
Instrument
Experiment Family
Experiment
SubExperiment
Time Frequency
Product
Realm
Variable
Variable Long Name
CMIP Table

Display 10 datasets per page

Add All Displayed to Datacart | Remove All Displayed from Datacart

Results | Data Cart

ESGF Governance is now defined

- ESGF Review Board (ERB) consists* of institutions and individuals whose goal is to advance ESGF technology by providing development direction and oversight. While ESGF's open-source nature facilitates progression via its many contributors and collaborations, the ERB provides synergy and cohesion among various development efforts, thus ensuring changes benefit the community as a whole.

- The ERB serves the following functions:

    - Maintains a roadmap of ESGF, including long-term plans
    - Makes decisions on high-impact code changes to ESGF

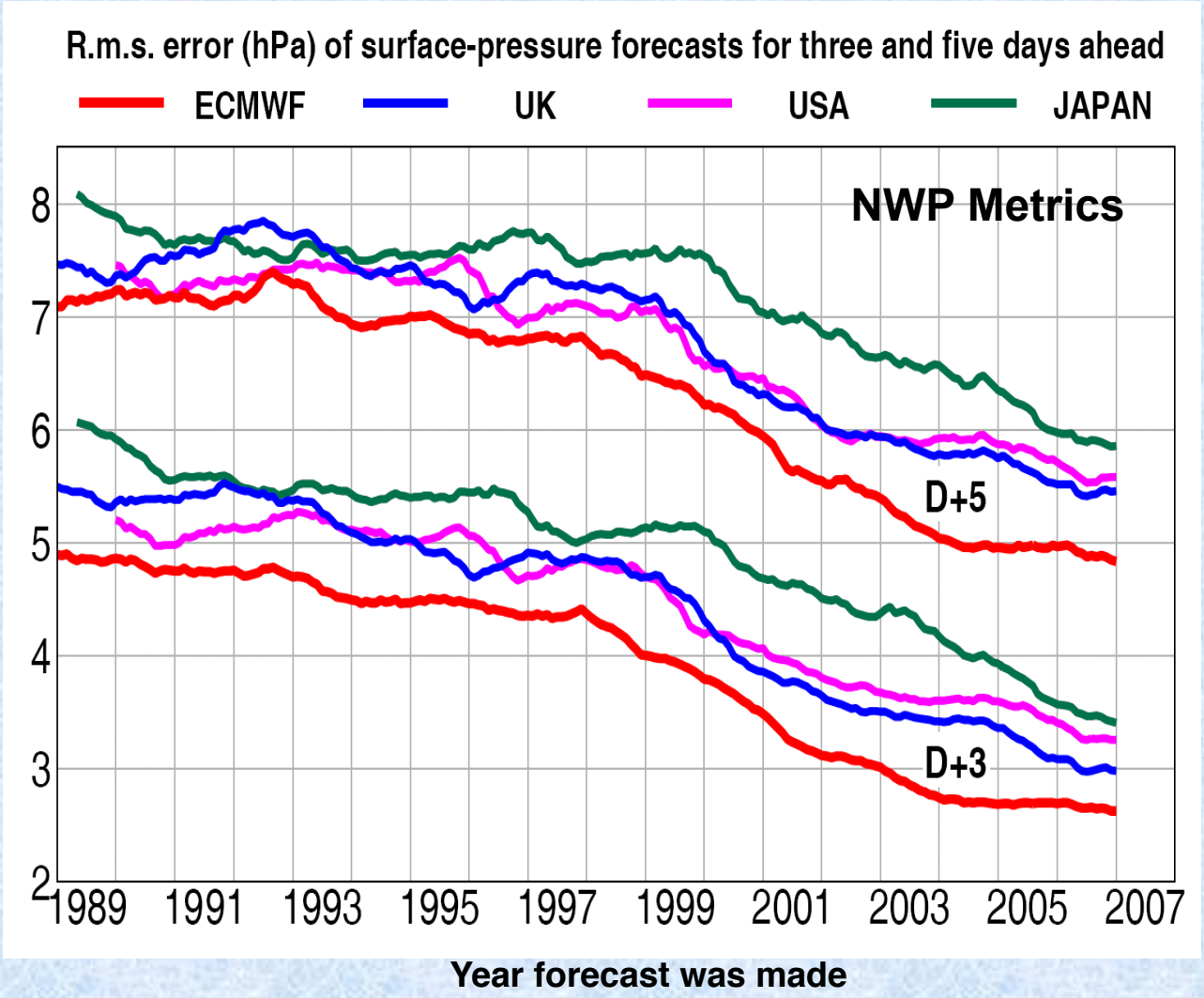
The ERB is composed of:
(1) steering committee (includes representation from funding agencies)
(2) executive committee responsible for meeting sponsors/stakeholders/community and setting and prioritizing work
(3) technical committee, responsible for the development of the system architecture, management of the development lifecycle and scheduling releases.

ESGF governance document available via http://esgf.org/ or by contacting D. Williams

*Committees are now being populated

Courtesy D. N. Williams, AIMS

WCRP efforts to establish routine performance metrics for climate models



The climate modeling community does not yet have routine performance metrics

Courtesy
M.Miller/A.Simmons

What is usually meant by climate model "metrics"?

- "Metrics", as used here, are succinct and objective measures of the quality of a model simulation – usually a scalar quantity
- Quantify errors, usually *not* designed to diagnose reasons for model errors
- Skill in simulating things we have observed: "performance metrics"
- Model reliability for application (e.g., "projection reliability metrics")
  - How accurate are model projections of climate change?
  - Extremely valuable… and… extremely difficult

Questions motivating routine benchmarks for climate models

- Of direct concern to the WGNE and WGCM:
  - Are models improving?
  - Do some models consistently agree with observations better than others?
  - What do models simulate robustly, and what not?

- Related research drivers:
  - How does skill in simulating observed climate relate to projection credibility?
  - Can we justify weighting model projections based on metrics of skill?

First steps… focus on annual cycle (which is in widespread use)

Standard annual cycle:

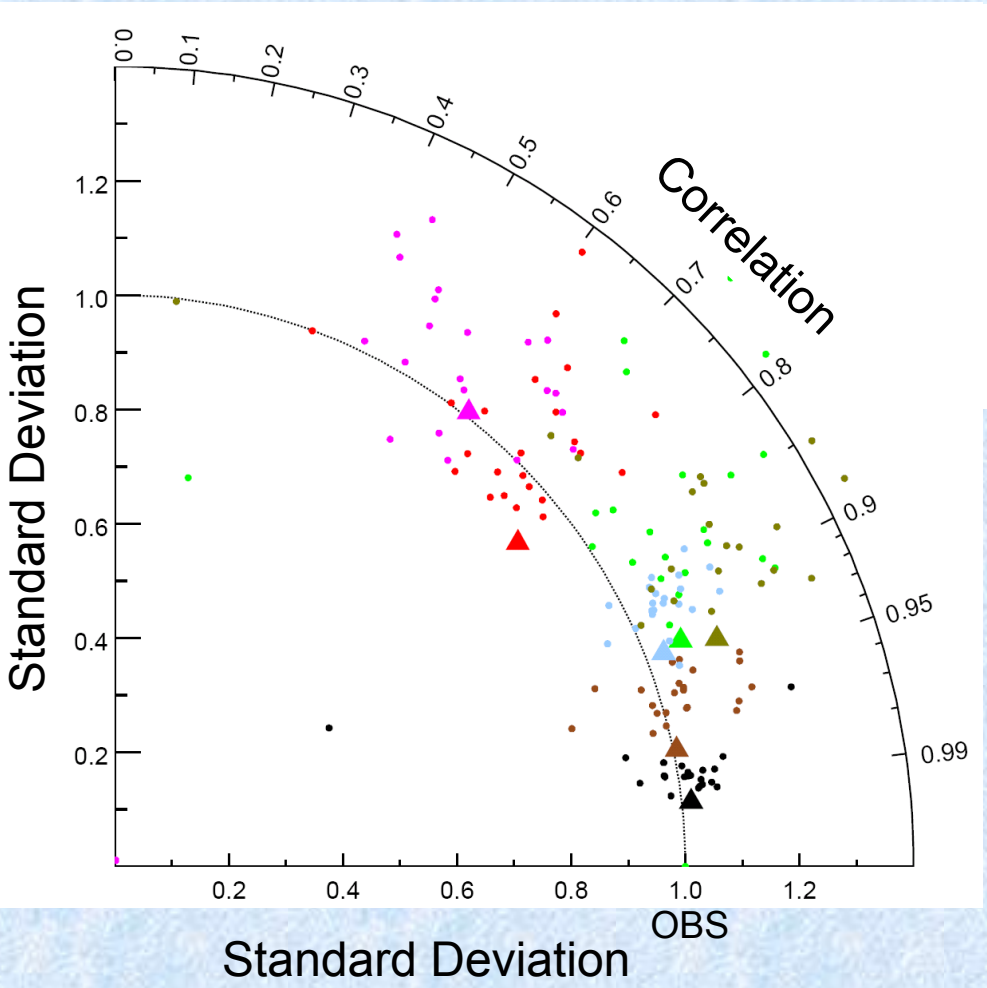
- 15-20 large- to global- scale statistical or “broad-brush” metrics
- Domains: Global, tropical, NH/SH extra-tropics
- 20 year climatologies: Annual mean, 4 seasons
- Routine metrics: bias, centered RMS, MAE, correlation, standard deviation
- Field examples: OLR, T850, q, SST, SSH, sea-ice extent
- Observations: multiple for most cases

Exploring an extended set of metrics, coordinating with other working groups (in progress):

- ENSO (CLIVAR Pacific Panel)
- Monsoons (CLIVAR AAMP)
- MJO (YOTC Task force)
- Carbon cycle in emission-driven ESMs (ILAMB)

Some simulated fields agree much more closely with observations better than others

Taylor diagram for CMIP3 annual cycle global climatology (1980-1999)



Sea Level Pressure: ERA40 reference

Total precipitation rate: CMAP reference

Total Cloud Cover: ISCCP reference

LW radiation TOA (OLR): CERES reference

Reflected TOA Shortwave: ERBE reference

Air Temperature (850 hPa): ERA40 reference

Zonal Wind (850 hPa): ERA40 reference

- Variable dependent skill
- Multi-model mean “superiority”

The essential role of observations for climate model performance metrics

- **The quest for traceability:**
  - Knowing the data came from the appropriate source (ideally the data experts)
  - Accurate information concerning the data product version
  - Documentation on the data product that is relevant for model analysts

- **Quantifying observational uncertainty remains a key challenge:**
  - For some fields, model errors remain >> than observational uncertainty, but not so in many cases
  - Although inadequate, the common path is to characterize obs uncertainty by using multiple products
  - Increasingly, model analysts expect useful quantification of uncertainties

Possible advancements for a community-based effort to establish routine benchmarks for climate models

- The WGNE/WGCM metrics panel is developing an analysis package to be shared with all leading modeling groups. This will include:
  - simple analysis routines
  - **observational data**
  - database of metrics results from all available climate models
- This will enable modeling groups to compare the results from other models within their model development process.