Modeling Infrastructure, Data Standards, and Protocols

Pan-WCRP Modeling Meeting
Exeter, UK

V. Balaji[1]   Karl Taylor[2]

[1]NOAA/GFDL and Princeton University

[2]PCMDI

11 October 2017

Outline

Fast forward to today...

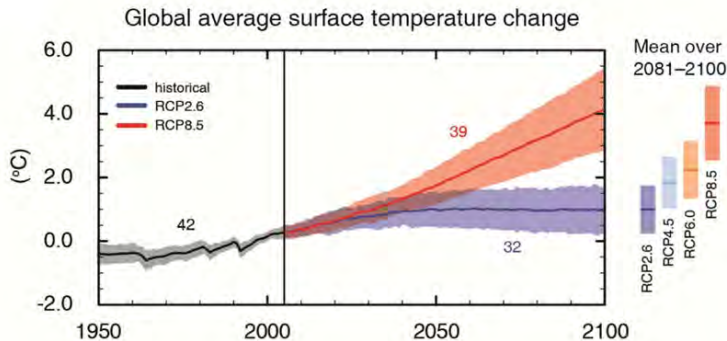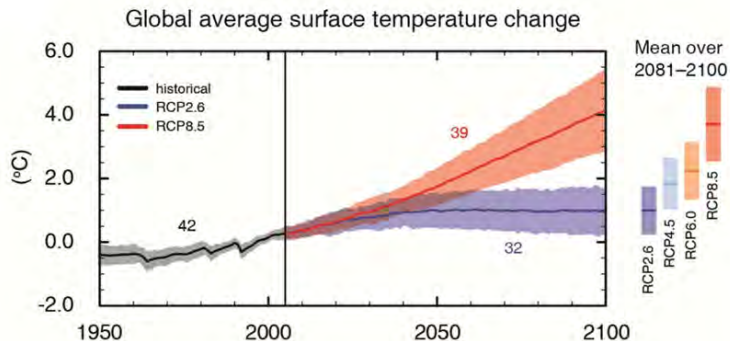Global average surface temperature change

Figure SPM.7 from the IPCC AR5 Report. 20th century warming cannot be explained without greenhouse gas forcings.

Multi-model ensembles for climate projection

Global average surface temperature change

- Critically depends on software, metadata, and data standards: the Earth System Grid Federation (`http://esgf.org`): a 3 PB federated archive.
- Key technical issues like replication, versioning, subsetting, QC, citation.

CMIP6 design: DECK and MIPs
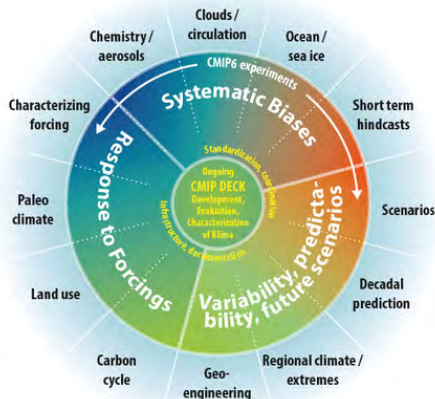
DECK experiments form the core; many specialized MIPs for smaller communities, some 24 of which have been endorsed by CMIP panel. Figure courtesy Meehl et al (*Eos* 2014).

CMIP6 Scientific Design

Coupled Model Intercomparison Project Phase 6 (CMIP6): Design and Organization

Veronika Eyring, Jerry Meehl, Bjorn Stevens, Ron Stouffer, Karl Taylor (CMIP Panel)

Sandrine Bony and Cath Senior (WGCM Co-chairs)

V. Balaji (WGCM Infrastructure Panel co-chair with K. Taylor)

16 January 2015 *(updates to CMIP6 Data Request Timeline on Slide 9)*
Please see the CMIP Panel website for additional information and updates:
http://www.wcrp-climate.org/index.php/wgcm-cmip/about-cmip

Contact for questions: CMIP Panel Chair Veronika Eyring (email: Veronika.Eyring@dlr.de)
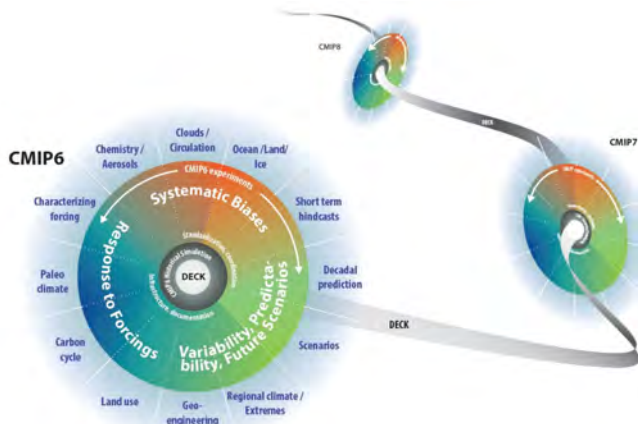
The final CMIP6 Design, possibly with small modifications to the here presented figures and wording, will be published in a GMD Special Issue together with a description of the CMIP6-Endorsed MIPs and the forcing datasets. This Special Issue will open 30 April 2015.
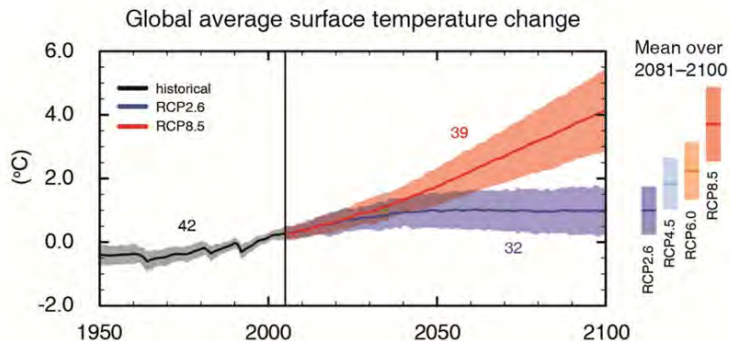
http://goo.gl/FMYRKe

CMIP evolution

DECK is designed to evolve slowly or not at all.

- IPCC Assessment Reports are snapshots of the "state of the science", but not directly linked to CMIP.

Multi-model ensembles for climate projection

Global average surface temperature change

- Critically depends on software, metadata, and data standards: the Earth System Grid Federation (`http://esgf.org`): a 3 PB federated archive.
- Key technical issues like replication, versioning, subsetting, QC, citation.

The global data infrastructure underpinning MIPs

- MIPs, and in general any science involving cross-model comparisons, critically depend on the global data infrastructure – the "vast machine" (Edwards 2010) – making this sort of data-sharing possible.
- Infrastructure should not be a research project.
- Infrastructure should be treated as such by the national and international research agencies, but it is instead funded piecemeal, as a soft-money afterthought. This places the system at risk (NRC 2012: "A National Strategy for Advancing Climate Modeling", ISENES-2 Infrastructure Strategy document, 2012.)

Role of WGCM and its infrastructure panel

- Provide scientific guidance and requirements for the GDI; exert greater influence over its design and features.
- Provide standards governance allowing for orderly evolution of standards.
- Provide design templates (e.g CMOR extensions) for groups designing MIPs and work to ensure their conformance to standards.
- Work with academies and publishers to require adequate data citation and recognition for data providers.
- Intercede with national agencies to provision data infrastructure with adequate and stable long-term funding.

WIP: The WGCM Infrastructure Panel formed 2014

- Chaired by V. Balaji (Princeton/GFDL) and K. Taylor (PCMDI).
- Strategy to develop a series of "position papers" on global data infrastructure and its interaction with the scientific design of experiments. These will be presented to WGCM annual meeting.
    - protocol document for the "endorsed MIPs" delivered. Working with CMIP panel and MIP sponsors on CMIP6 data request.
    - data access policies: would open access simplify the technical design of the infrastructure?
    - data citations. Developing and promoting a path to data citations using DOIs and the emerging data journals, such as ESSD, Nature Scientific Data.
    - projected data volumes for CMIP6, strategies for managing the growth path
- Close involvement of the WIP and CMIP panel (e.g. joint papers)
- Interest from other WCRP working groups! (WGSIP, WGNE)
- Covers not only ESGF requirements but also other tools: ESDOC, CMOR, CF Conventions, ..

WIP Membership

- V. Balaji (co-chair): GFDL
- Karl Taylor (co-chair): PCMDI
- Luca Cinquini: NASA JPL
- Cecelia DeLuca: NOAA
- Sebastien Denvil: IPSL
- Mark Elkington: MOHC
- Francesca Guglielmo, LSCE
- Eric Guilyardi: IPSL
- Martin Juckes: BADC
- Slava Kharin: CCCma
- Michael Lautenschlager: DKRZ
- Bryan Lawrence : NCAS, BADC
- Dean Williams: PCMDI

a blend of computer and climate scientists representing data centers and modeling groups: rotating membership with overlapping 2-year cycles

ESGF, a global network of compute and data nodes

Figure courtesy IPSL.

WIP Position Papers

https://earthsystemcog.org/projects/wip/

- CDNOT Terms of Reference: a charter for the CMIP6 Data Node Operations Team.
- CMIP6 Global Attributes, DRS, Filenames, Directory Structure, and CVs: conventions and controlled vocabularies for consistent naming of files and variables.
- CMIP6 Persistent Identifiers Implementation Plan: a system of identifying and citing datasets used in studies, at a fine grain.
- CMIP6 Replication and Versioning: a system for ensuring reliable and verifiable replication; tracking of dataset versions, retractions and errata.

WIP Position Papers

https://earthsystemcog.org/projects/wip/

- CMIP6 Quality Assurance: systems for ensuring data compliance with rules and conventions listed above.
- CMIP6 Data Citation and Long Term Archival: a system for generating Document Object Identifies (DOIs) to ensure long-term data curation.
- CMIP6 Licensing and Access Control: terms of use and licenses to use data.
- CMIP6 ESGF Publication Requirements: linking WIP specifications to the ESGF software stack, conventions that software developers can build against.
- Errata System for CMIP6: a system for tracking and discovery of reported errata in CMIP6.

Conclusions

- the CMIP multi-model ensemble is an organizing principle and fundamental tool of climate science
- Addresses WCRP Grand Challenges through WGCM-led experimental design
- Design allows both for continuity across CMIP generations and new experiments in response to evolving science
- provides input to IPCC and other Assessment Reports
- WGCM Infrastructure Panel translates CMIP experimental design into requirements for the global data infrastructure
- Close involvement of WIP with ESGF-XC (overlapping membership)
- GMD Paper in CMIP6 Special Issue "Requirements for a global data infrastructure in support of CMIP6" close to submission.