Towards efficient and systematic model benchmarking in CMIP6 and beyond

Peter J. Gleckler
Program for Climate Model Diagnosis and Intercomparison (PCMDI)
Lawrence Livermore National Laboratory, Livermore, CA, USA

Veronika Eyring
Deutsches Zentrum für Luft- und Raumfahrt (DLR)
Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

2016 International Land Model Benchmarking (ILAMB) Workshop
May 16–18, 2016
Washington DC

Talk outline

- Background and motivation
- Earth System Model Evaluation Tool (ESMValTool)
- The PCMDI Metrics Package (PMP)
- Expected advancements to modeling research and development
- Synergies between developing benchmarking capabilities

Further enhancing the scientific value of CMIP

The traditional mode of documenting CMIP research (peer-reviewed publications) will continue as the primary deliverable of CMIP. However, there is a pressing need to more efficiently produce and document established model evaluation results to:

- Advance the science more efficiently (less re-inventing the wheel)
- Facilitate national assessments, the IPCC process, etc.
- More directly contribute to model development (via quick feedback)

Community-based building blocks are a viable mechanism to accomplish this, thanks to the design target provided by the CMIP data conventions

Status and Outlook

CMIP
The World Climate Research Programme's Coupled Model Intercomparison Project

CMIP6 Status

- CMIP6 Experimental Design finalized
- Forcing datasets for DECK and CMIP6 historical simulations finalized by end of April
- CMIP6 Simulation Period (2016-2020)
- Infrastructure in preparation (including data request) by WGCM Infrastructure Panel (WIP)

CMIP6 Participating Model Groups: > 30 using a wide variety of different model versions

21 CMIP6-Endorsed MIPs that build on the DECK and CMIP historical simulations to address a large range of specific questions with WCRP Grand Challenges as scientific backdrop.

CMIP6 Climate Projections part of a CMIP6-Endorsed MIP (ScenarioMIP)

- New scenarios span the same range as the RCPs, but fill critical gaps for intermediate forcing levels and questions for example on short-lived species and land-use.
- Forcings for future scenarios available by end of 2016, climate model projections expected to be available within the 2018-2020 time frame.

A Central Goal of CMIP6 is Routine Evaluation of the Models with Observations

- Efforts to develop community tools and to couple them to the ESGF are underway

Geosci. Model Dev. Special Issue on CMIP6

- Overview of the CMIP6 Experiment Design and Organization (Eyring et al., GMD, 2016)
- Experimental design from all CMIP6-Endorsed MIPs
- Description of the CMIP6 forcing data and infrastructure

=> We expect CMIP6 to continue CMIP's tradition of major scientific advances

http://www.wcrp-climate.org/index.php/wgcm-cmip/about-cmip

Earth System Model Evaluation Tool (ESMValTool)

A community diagnostic and performance metrics tool for routine evaluation of Earth System Models in CMIP

http://www.esmvaltool.org/

- **Standard namelists** reproduce sets of diagnostics & performance metrics that have demonstrated their importance in ESM evaluation in the peer-reviewed literature
- **Priority of the effort** so far has been to target specific scientific themes, incl.:
  - Selected Essential Climate Variables (ECVs)
  - A range of known systematic biases common to ESMs (e.g. coupled tropical climate variability, monsoons, Southern Ocean, continental dry biases)

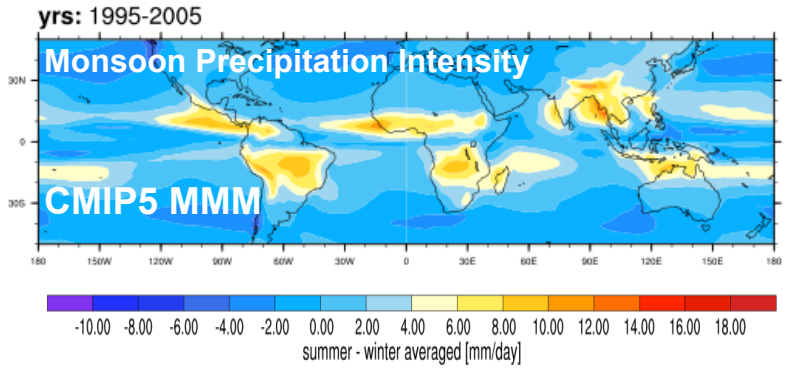
  → can easily be **extended with additional analysis**
- **Community development** under a *subversion* controlled repository
  - Currently ~70 scientists part of the development team from ~30 institutions
  - Allows multiple developers from different institutions to contribute and join
  - Regular releases as open source software (last release version 1.0)
- **Goals:**
  - Improve ESM evaluation beyond the state-of-the-art
  - Routine evaluation of the CMIP DECK and historical simulations as soon as the output is submitted to the Earth System Grid Federation (ESGF)
  - Support of individual modelling centres

The wider climate community is encouraged to contribute to this effort and to join the ESMValTool development team

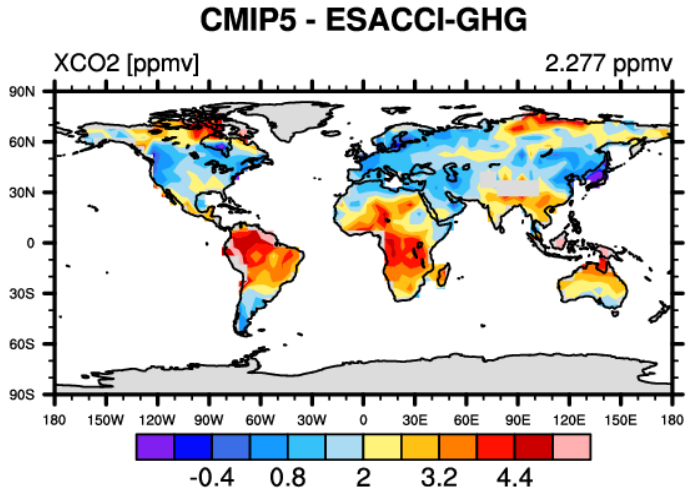
Routine Benchmarking and Evaluation Central Part of CMIP6

ESMValTool tool to produce well-established analyses as soon as CMIP model output is submitted

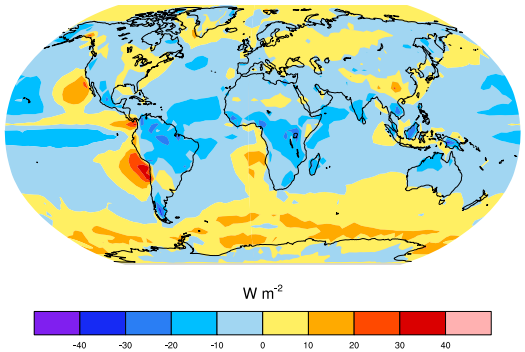
*Similar to **Figure 9.7** of AR5*



Broad Characterization of Model Behavior (including IPCC AR5 Chapter 9 and 12 Diagnostics)



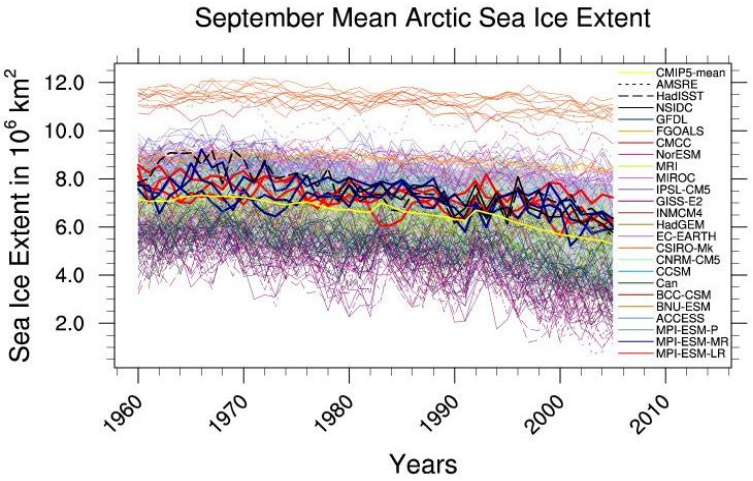
Net Cloud radiative effect against CERES EBAF (2001–12)



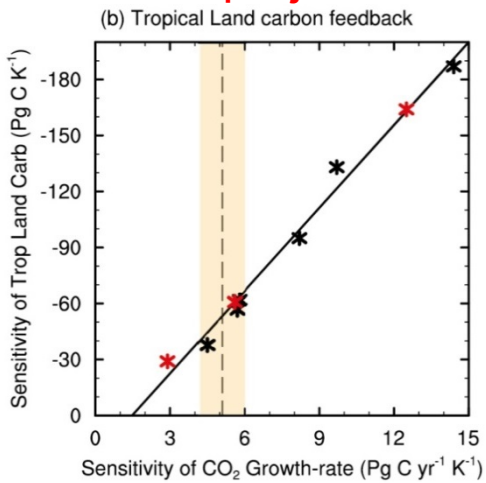
*Similar to **Figure 9.5** of AR5*

Running alongside the ESGF

*Similar to **Figure 9.24** of AR5*



Link to projections



*Similar to **Figure 9.24** of AR5*

ESMValTool
Earth System Model eValuation Tool

Eyring et al., GMD, 2016

The PCMDI Metrics Package (PMP)

https://github.com/PCMDI/pcmdi_metrics

- Emphasizing a diverse suite of summary statistics objectively comparing models and observations across space and time scales, the PMP provides a database for knowledge discovery by researchers and model developers
- Succinct yet comprehensive summaries of model performance
- Open source tools and database built on DOE supported software (UV-CDAT), including hundreds of analysis utilities designed specifically for CMIP conventions
- All code is now publically available on github (Gleckler et al., EOS, May 2016)
- Designed to readily enable the research community to use and contribute new analysis (at initial launch there are ~ dozen collaborators)
- Currently testing at 4 modeling centers (GFDL, NCAR, IPSL and CSIRO)

The PCMDI Metrics Package (PMP)

https://github.com/PCMDI/pcmdi_metrics

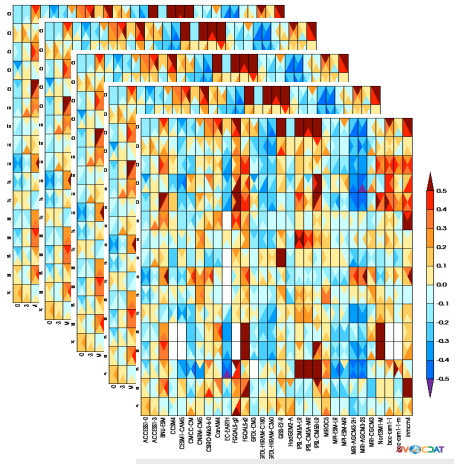
- Emphasizing a diverse suite of summary statistics objectively comparing models and observations across space and time scales, the PMP provides a database for knowledge discovery by researchers and model developers
- Succinct yet comprehensive summaries of model performance
- Open source tools and database built on DOE supported software (UV-CDAT), including hundreds of analysis utilities designed specifically for CMIP conventions

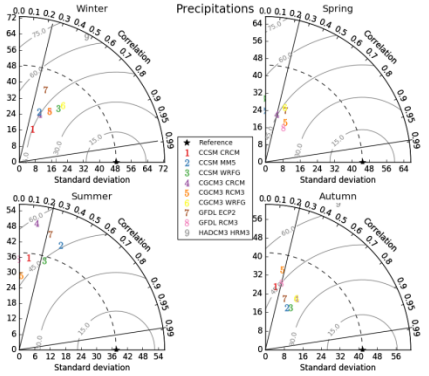
  EXAMPLE: **djf_anom = cdutil.DJF.departures(d)**
- All code is now publically available on github (Gleckler et al., EOS, May 2016)
- Designed to readily enable the research community to use and contribute new analysis (at initial launch there are ~ dozen collaborators)
- Currently testing at 4 modeling centers (GFDL, NCAR, IPSL and CSIRO)

The PCMDI Metrics Package (v1.1x)

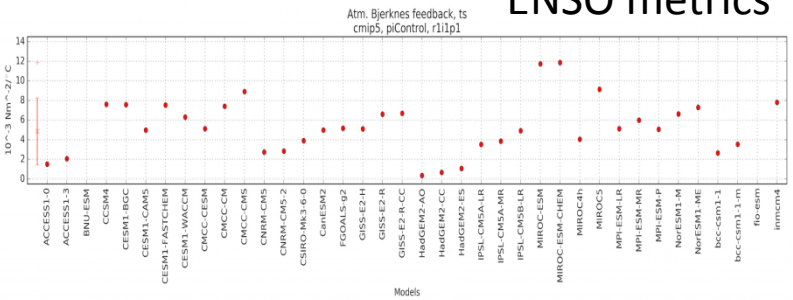
https://github.com/PCMDI/pcmdi_metrics

Taylor and Gleckler Plots:
orthogonal error statistics in space and time

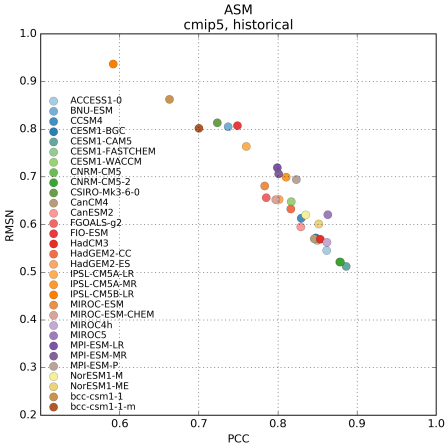




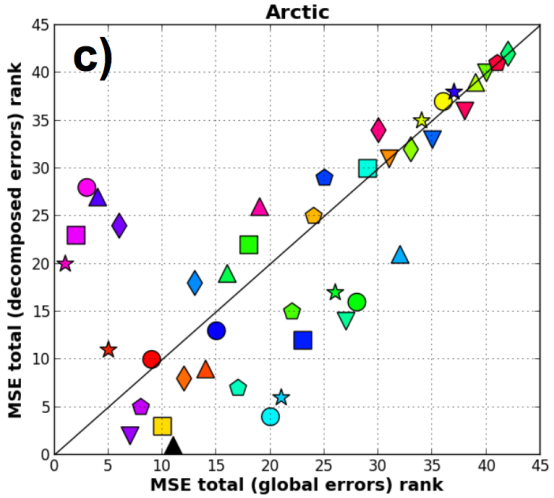
ENSO metrics



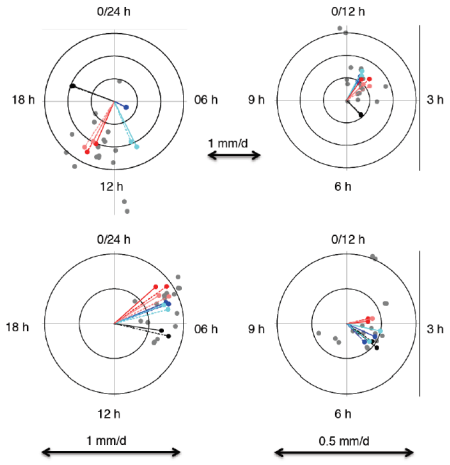
Monsoon Region
Precipitation Indices



Sector scale sea ice

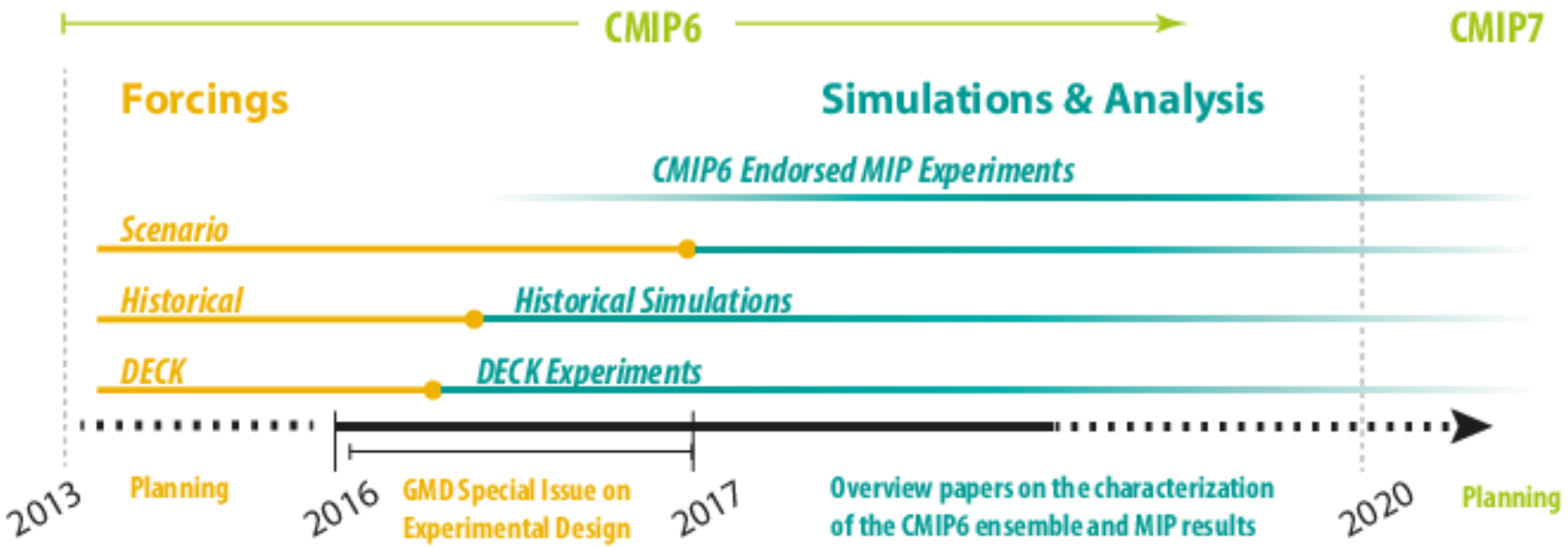


Diurnal cycle of precipitation



Gleckler, P. J., C. Doutriaux, P. J. Durack, K. E. Taylor, Y. Zhang, and D. N. Williams, E. Mason, and J. Servonnat (2016), A more powerful reality test for climate models, *Eos, 97,* doi 10.1029/2016EO051663

CMIP6 Timeline



Eyring et al., Overview of the CMIP6 experimental design and organisation, GMD, 2016

Developing capabilities: shared deliverables in CMIP6

- Collectively, the PMP, ESMValTool and ILAMB packages will be crucial for the systematic benchmarking in CMIP6
- More quickly and openly relay to analysts and modelling centers the strengths and weaknesses of the simulations including the extent to which long-standing model errors remain evident in newer models
- Will be applied to all DECK and relevant CMIP6 simulations as soon as they are published on ESGF, with results provided to modeling groups and research communities
- Emphasizing diagnostics & metrics in the peer-reviewed literature with demonstrated importance in ESM evaluation
- Designed to facilitate community-based contributions for repeat use
- Synthesis papers expected early in CMIP6 analysis phase
- Enable modeling groups to more systematically evaluate simulations during model development - using analysis community expertise

Synergies between developing benchmarking capabilities

- As this new era in CMIP unfolds, having a few tools is a good thing: we will learn more about how to further establish community building blocks (e.g., for CMIP7)

- Some overlap will help verify results and procedures but on the long-term it is important to minimize duplication of efforts and to align tools more closely.

- Coordination of observational data between ESMValTool and PMP is underway, as well as developing longer term strategy on how to organize analysis results. Collaboration with ILAMB will further ensure success of these community-based efforts

Observations - Alignment with CMIP

- obs4MIPs is helping to more closely align a diverse community of data experts with the modeling community (see poster)
- Key to this is ensuring consistency between the data conventions of model output and observational data
- CMOR is used by the CMIP modeling community to ensure conventions are properly applied. CMOR3, to be used for CMIP6, has been generalized to work with gridded observational data
- PMP and ESMValTool will be using CMOR3 to organize their observations to ensure they can be published (searchable) on ESGF

• CMOR = Climate Model Output Rewriter

An boom in accessibility to CMIP analysis – with challenges

- ~500 experiments in CMIP6, evermore complex diagnostic output, volumes of data, etc…
- During CMIP6 we can expect a proliferation readily available analysis that compliments the peer-reviewed literature
- This will transform the CMIP process and accelerate the pace at which we use models to further understand climate change
- Wealth of analysis will need to be managed carefully and traceability will be key
- The WGCM, CMIP Panel and WGNE/WGCM metrics and diagnostic panel will be engaged in ensuring this exciting advancement unfolds in a manner that is in the best interest of advancing the science

Summary

Systematic evaluation of CMIP simulations with community-based capabilities are needed to:

- Advance scientific understanding more efficiently (less re-inventing)
- Facilitate model development (via quick feedback) and benchmarking
- Valuable resource for a variety of demands (assessments, etc.)

- This is a high priority for CMIP6, and CMIP infrastructure and conventions make this challenge tractable

- At this stage some diversity in approaches is helpful (but not too many!)

- The scale, complexity and importance of the task necessitates transparency and accessibility, with coordination between efforts