An update on Metrics and Diagnostics as applied to CMIP and their relevance to WGNE

WGNE 33 (JMA, Tokyo)

Peter J. Gleckler
PCMDI/LLNL

October 11, 2018

LLNL-PRES-948197

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

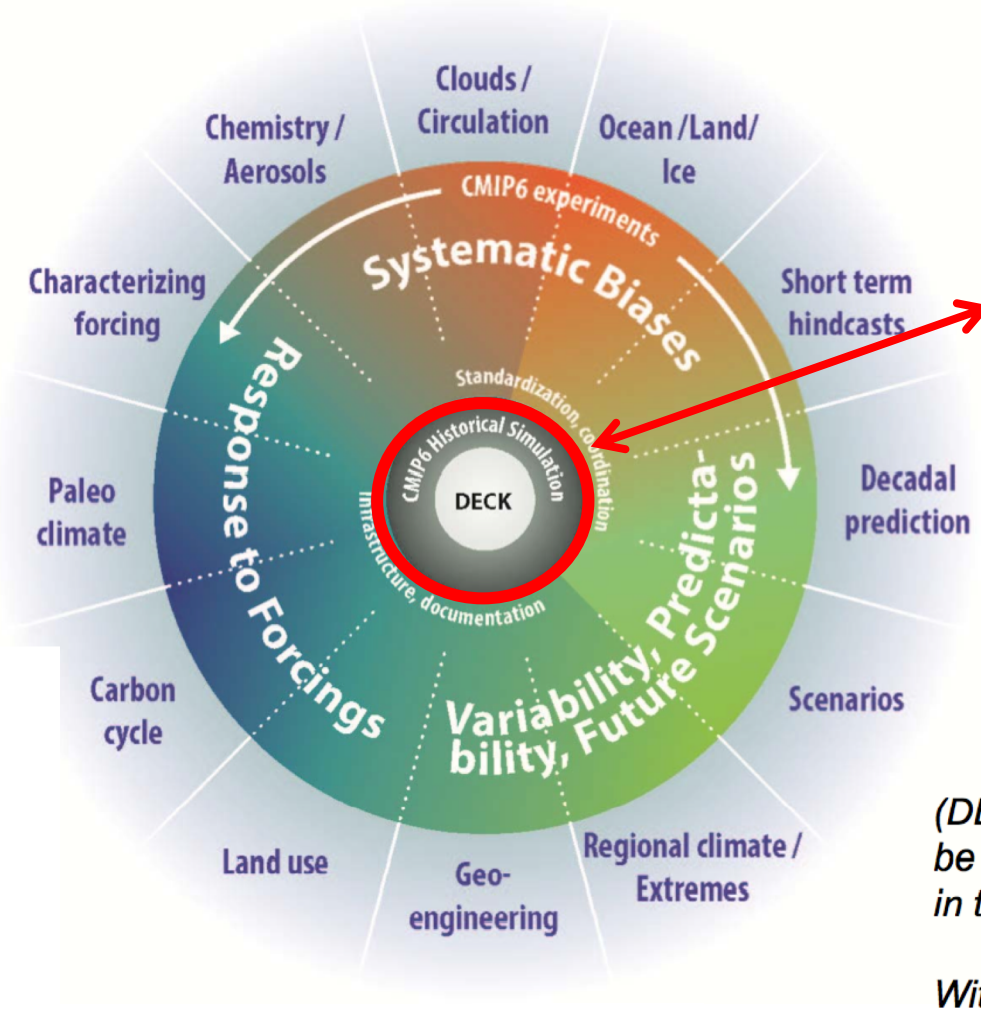
Outline
WGNE and WCRP's role: metrics and diagnostics for climate models

- Background
- WCRP strategic and implementation plan
  - —Opportunity to reassess what is needed and how best to accomplish it
  - —A proposal under discussion
- An area where WGNE might play in important role (hint: precipitation)
- Way forward

The CMIP Diagnosis, Evaluation, Characterization of Klima (DECK)

DECK + Historical: Experiments frequently performed as part of the model development process

More discussion of CMIP tomorrow!



DECK (entry card for CMIP)

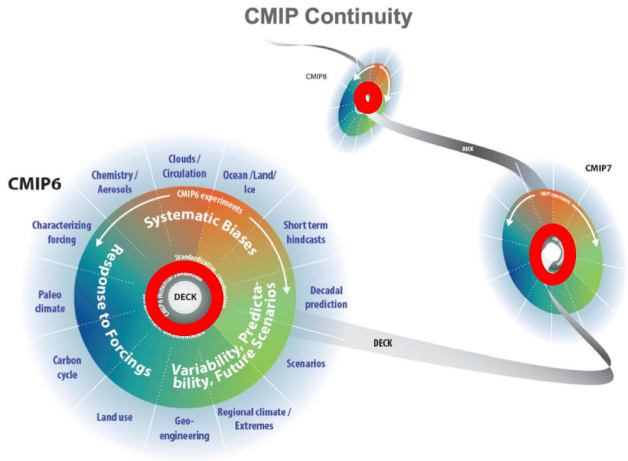
i. AMIP simulation (~1979-2014)
ii. Pre-industrial control simulation
iii. 1%/yr CO_2 increase
iv. Abrupt 4xCO_2 run

CMIP6 Historical Simulation (entry card for CMIP6)

v. Historical simulation using CMIP6 forcings (1850-2014)

(DECK & CMIP6 Historical Simulation to be run for each model configuration used in the subsequent CMIP6-Endorsed MIPs)

With proto-DECK experiments (LMIP,OMIP etc.) in CMIP6 Tier1



Some relevant markers

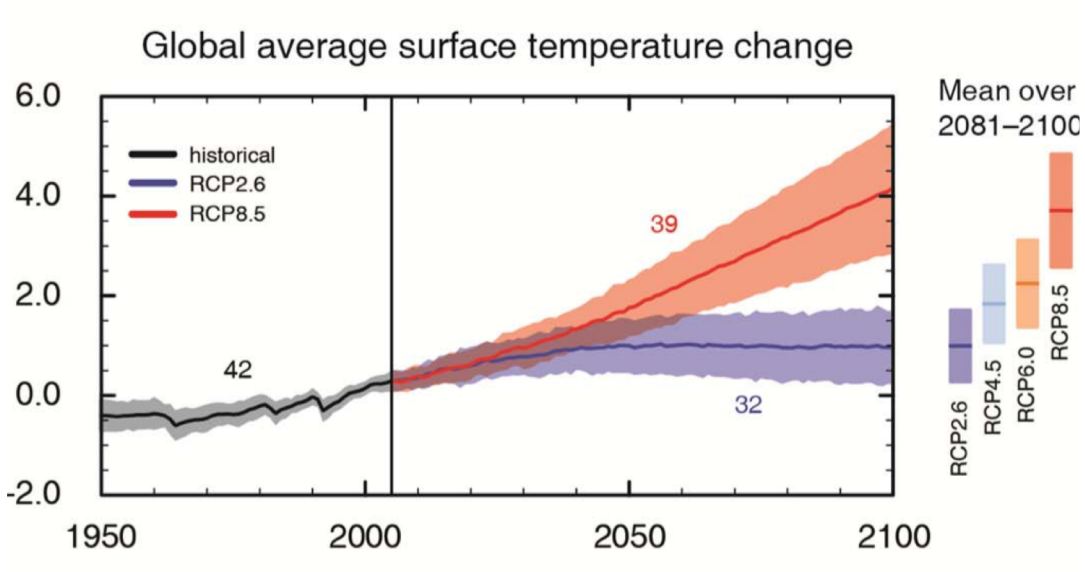
CMIP model evaluation and metrics research

- WGNE encourages group to identify a limited set of metrics for evaluating CMIP class models (2009)
- IPCC expert Good Practice paper on assessing multi-model projections (2010)
- Ample metrics research: new methods, process-oriented, contrasting MME[1] and PPE[2], model weighting, model dependence, tuning and emergent constraints
- A first - model weighting is applied in the IPCC AR5
- CMIP DECK defined, in part, to inspire ongoing benchmarking of models
- Routine model evaluation capabilities being developed by multiple teams (discussed in a few minutes)

[1] MME: Multi-model ensemble
[2] PPE: Perturbed physics ensemble

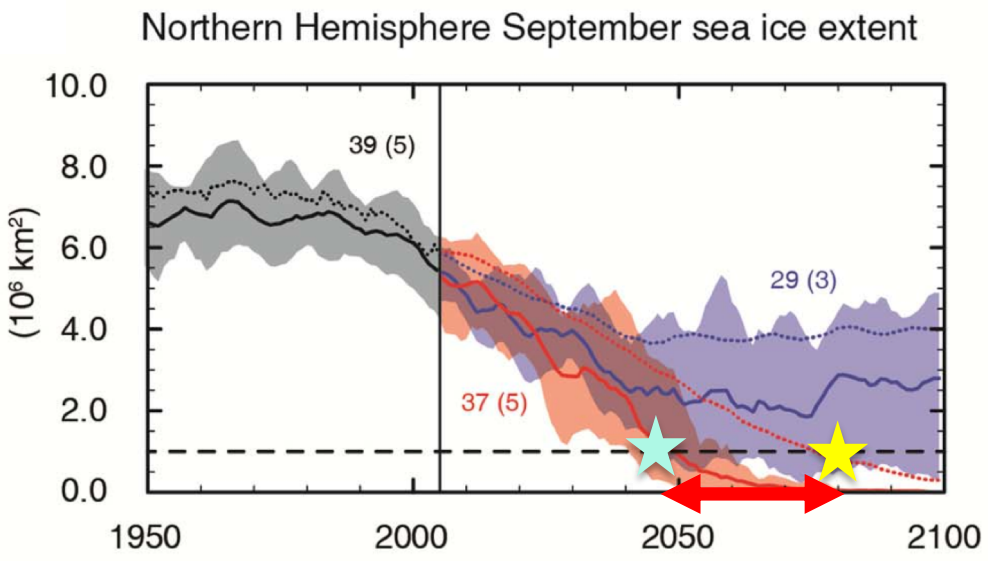
Weighting model projections

Remains an active area of research with important implications



All models treated equally: standard IPPC approach

AR5 WGI Figure SPM.7a



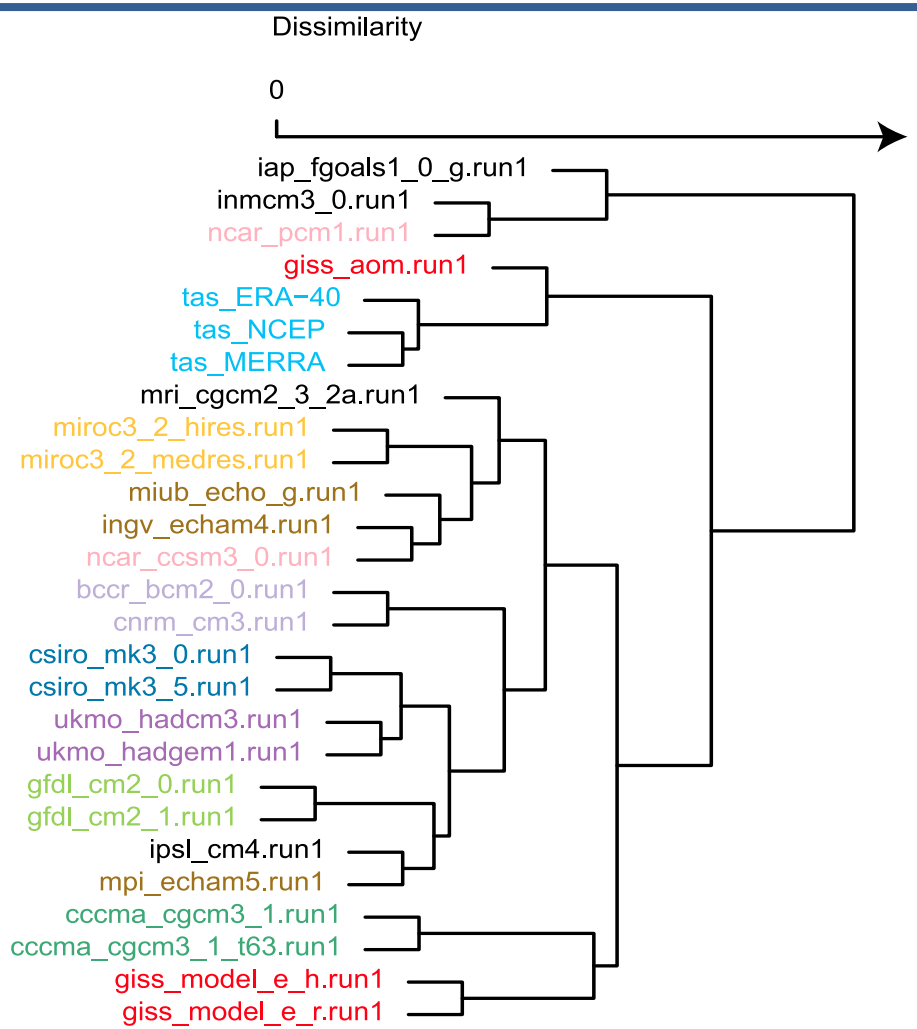
Weighting projections: A first in IPCC

Subset of 5 models averaged together, selected by how well they simulate the present day annual cycle and observed trends (sea ice loss)

AR5 WGI Figure SPM.7b

Model dependence

Masson and Knutti (2011), Knutti (2013), Sanderson et al. (2015), others



Quantifies distance between control runs of two models, accounting for mean state, seasonal cycle, and inter-annual variations

Demonstrates a level of dependence between model pairs

How to use this information in producing multi-model projections? Active area of research; progress is being made

WGNE/WGCM metrics and diagnostics panel (CMDP)

WGNE has a long history of encouraging objective testing of climate models

- WGNE established a group to identify a limited set of performance metrics for climate models (2010)
- Panel expanded and identified as a joint effort with WGCM/WGNE effort (2013)
- Additional scope recommended by WMAC to "include diagnostics" (2016)

Current members selected by relevant and diverse experience, and potential for liaison with key WCRP activities:

Beth Ebert (BMRC) – JWGV/WWRP, WMO forecast metrics

Veronika Eyring (DLR Germany) – WGCM/SPARC/CMIP6, stratosphere, ESMs

Pierre Friedlingstein (U. Exeter) – IGBP, carbon cycle

Peter Gleckler (PCMDI), chair – WGNE/WGCM, atmosphere and ocean

Simon Marsland (CSIRO) – CLIVAR OMDP, WGCM, ocean

Robert Pincus (NOAA) – GEWEX/GCSS, clouds/radiation

Karl Taylor (PCMDI) – WGCM, atmosphere, CMIP

Keith Williams (U.K. Met Office) – WGNE, Transpose AMIP, clouds

Metrics and Diagnostics Panel
Current Terms of Reference

- Foster an environment to advance community-based routine evaluation of climate models
- Coordinate with other WCRP activities that are actively developing diagnostics and performance metrics
- Identify analysis routines and packages that may be of potential use to modeling groups and researchers, and encourage functionality with the CMIP data conventions
- Ensure that well-established capabilities are applied to the CMIP DECK and Historical experiments, with results made readily accessible
- Encourage and facilitate performance metrics research by identifying key areas needing work and possibly organizing workshops
- Progress and terms to be reviewed annually by both the WGNE and the WGCM.

An incomplete listing of community-based capabilities that may be relevant for routine evaluation of CMIP DECK simulations

- ESMValTool (Eyring et al, GMD, 2016)
- PCMDI Metrics Package (Gleckler et al., EOS, 2016)
- Climate Variability and Diagnostics Package (Phillips et al., 2014)
- ILAMB (Luo et al., 2012)
- CFMIP diagnostics (Y Tsushima, 2017)
- TECA (Prabhat et al., 2012)
- ARM Diagnostics package (Zhang et al., 2018)
- MJO task team diagnostics
- NOAA MAPP process-level team
- CLIVAR basin panels
- ….

The WCRP strategic plan is an opportunity to assess priorities
Rethinking the remit of the Metrics and Diagnostics Panel (CMDP)

How the panel has helped:

- Raised the profile - performance testing of climate models
- Inspire new research and the development of evaluation tools

Where progress is lacking:

- Advancing specific scientific methods and topics
- Synergies with the JVWGR

A difficulty is that the existing Panel TOR are too broad

Discussions within the WCRP to reinvent the role of the panel(s)

A community framework for consensus climate model evaluation

A discussion document to be submitted as feedback to the WCRP 2019-2029 implementation plan, including a proposal for a pan-WCRP Model Evaluation Panel (E. Guilyardi, P. Gleckler, V. Eyring, G. Flato, M. Rixen and many others)

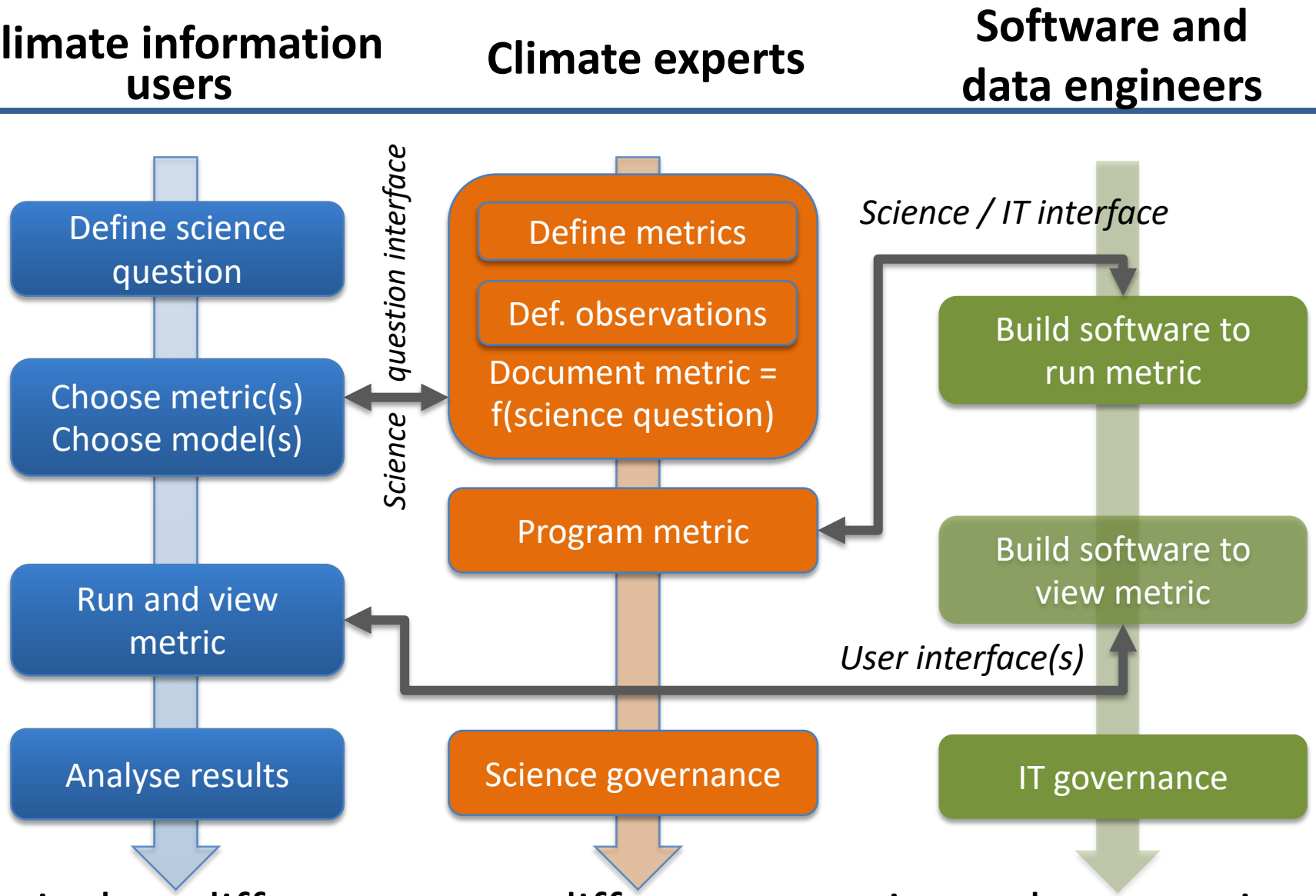
Consider the following possibility:

- A pan-WCRP panel inspires targeted expert teams to define and implement a limited set of model metrics which over time can be revised as the science advances

- A few relevant examples:
  - MJO diagnostics task force
  - CFMIP community diagnostics codes
  - CLIVAR ENSO metrics (in development)

Model evaluation workflow

Courtesy E. Guilyardi



Articulate different actors, different expertise and expectations

Benchmarking simulated precipitation

A possible area where WGNE's oversight could be valuable

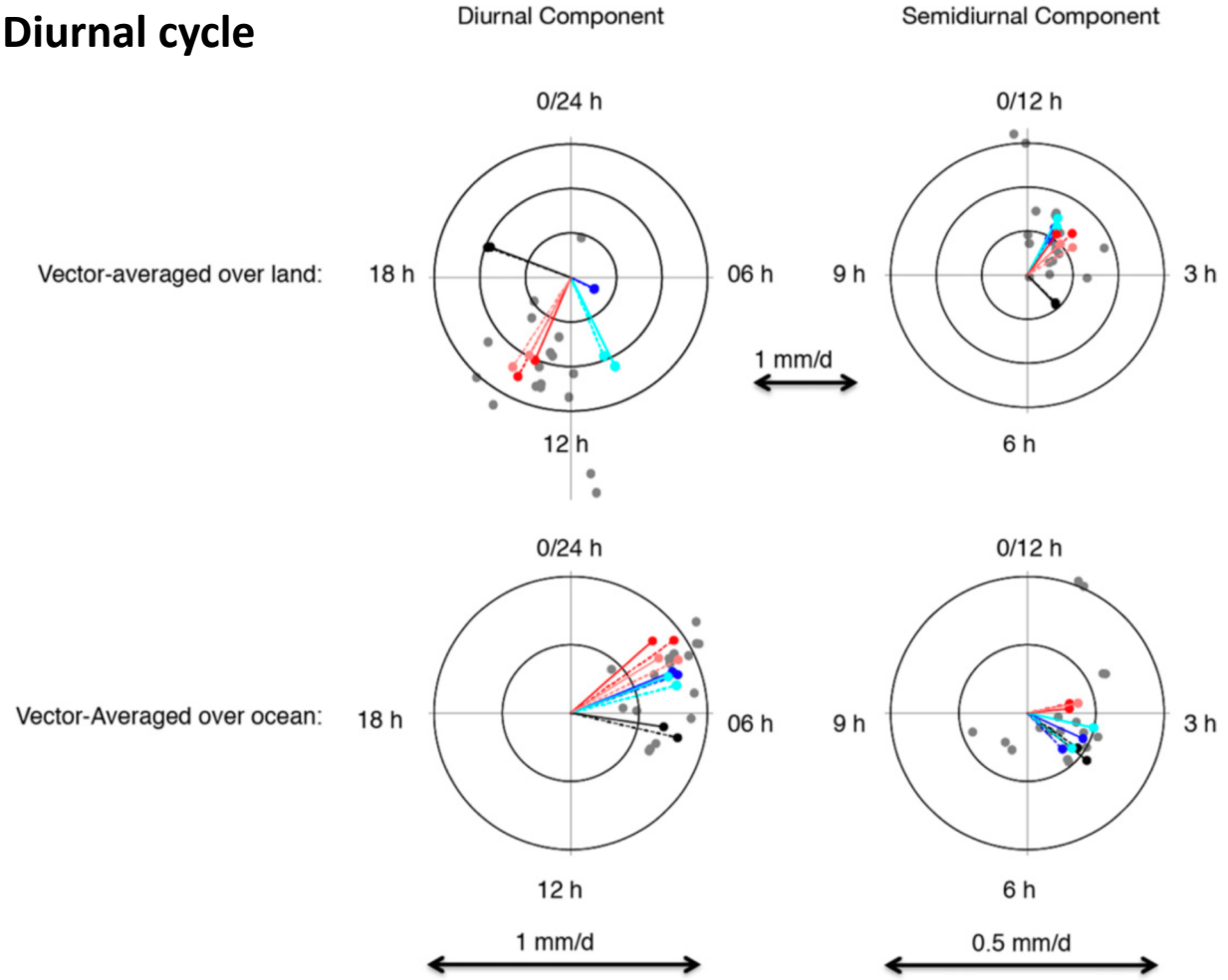
Courtesy C. Jakob

Why precipitation? Because . . .

- It matters to so much more than just our science
- We have a lot of relevant science already happening
- Potential research funders care a lot about it
- Because it's hard to improve (and to measure!)
- Improving it will likely affect many other things in models
- Measuring improvement is more tangible than "reducing uncertainty"
- We need to work together to achieve it

Benchmarking simulated precipitation
A few (of many) examples

Diurnal cycle

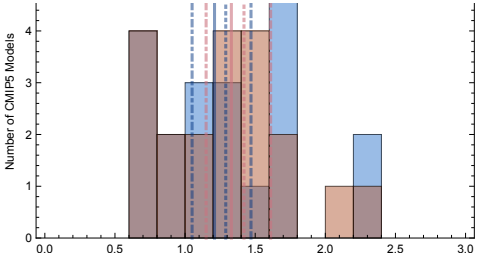


Covey et al., 2016, 2018

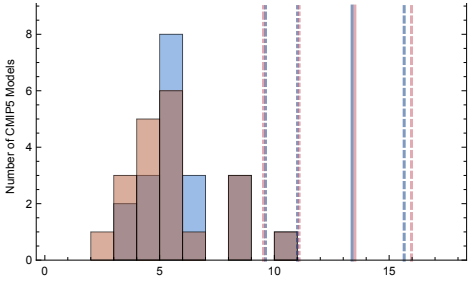
Variance decomposition

CMIP5 models appear to significantly underestimate intermittency w.r.t. GPCP, TRMM and CMORPH

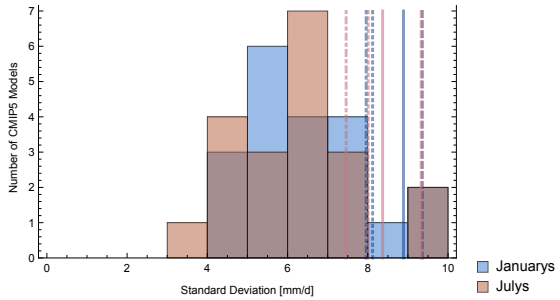
s.d. of Mean Diurnal Cycle



s.d. of irregular sub-diurnal variations



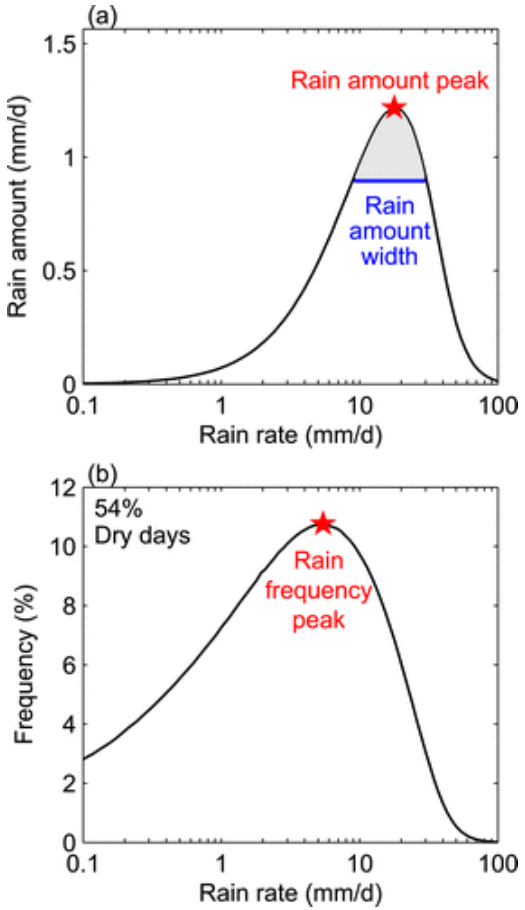
s.d. of daily mean values



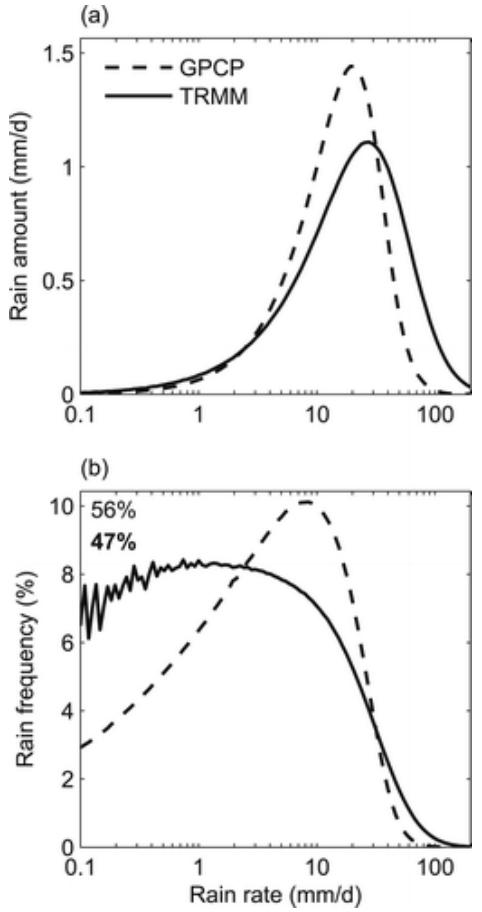
Benchmarking simulated precipitation
A few (of many) examples

Typical daily precipitation

Proposed metrics



Observational discrepancies

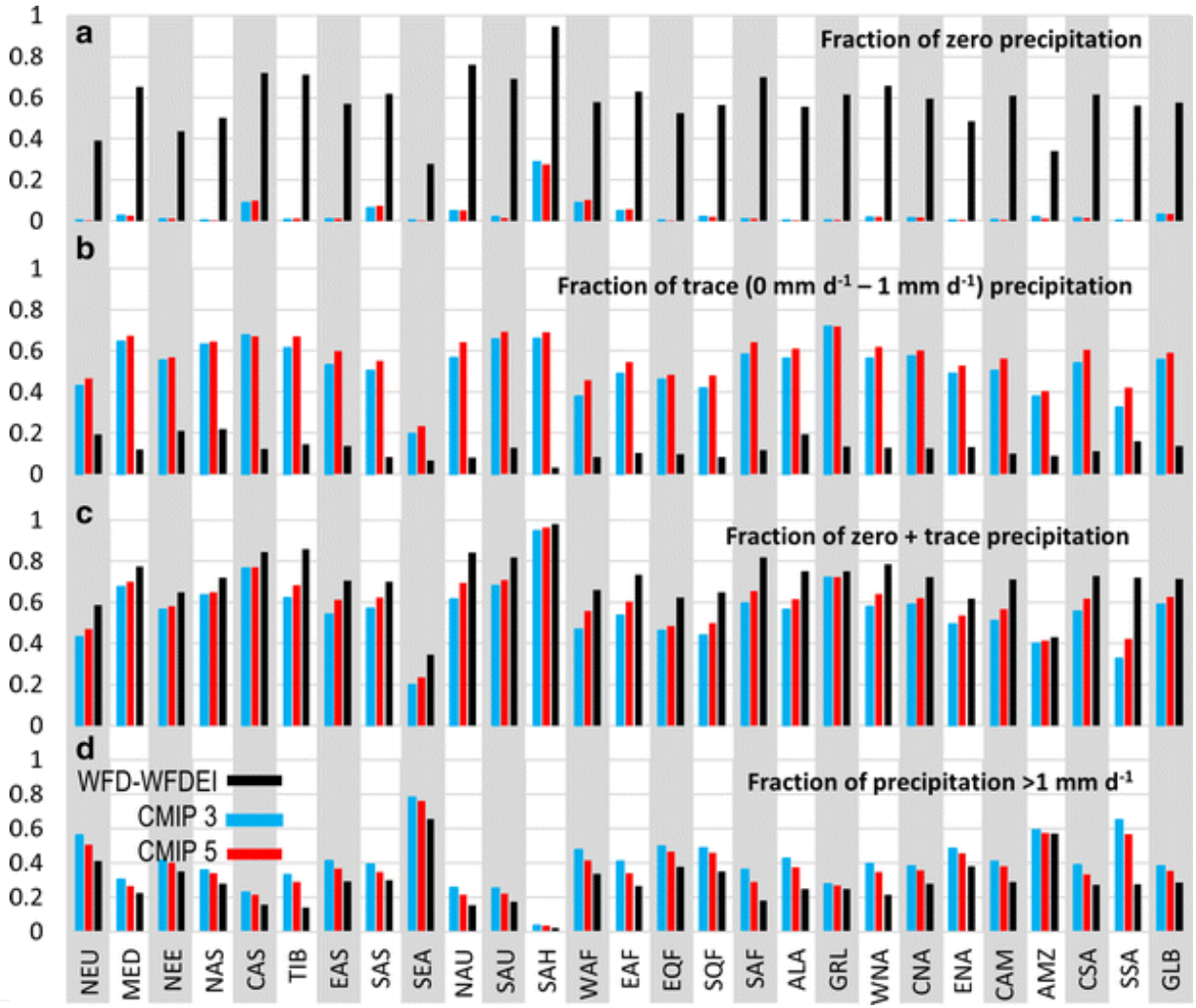


Pedergrass and Deser, 2017

Benchmarking simulated precipitation
A few (of many) examples

Perkins score

$$S_{SCORE} = \sum_{1}^{N} \min\left(Z_m, Z_o\right)$$



Koutroulis et al., 2015

Benchmarking simulated precipitation
Interest is building

- An AGU 2018 town hall is scheduled to get community feedback (led by R. Joseph, P. Gleckler, C. Jakob and A. Pendergras)
- A DOE workshop in spring 2019 is being planned inviting ~20 experts
- Intent is to make some progress first, then establish WCRP connection
- WGNE has expertise that could be very helpful

Benchmarking simulated precipitation

Challenging the modelling community to improve simulated precipitation - How?

- Step 1: An assessment report (and review paper) on the state of the art measured quantitatively
- Step 2: Enable modelers to apply metrics (i.e., code and data provided)
- Step 3: A serious attempt to increase the number of developers in this area achieved by engaging modelling centres and funding agencies.
- Step 4: A repeat of the assessment report in N years, where 5<N<10

Summary

Advancing the use of model metrics for benchmarking climate model improvements

- Substantial progress has been made towards comprehensive objective CMIP model evaluation
- A new way of organizing community efforts is under discussion within WCRP, possibly via an overarching coordinating body, which would lead to changing the makeup and TOR of the WGNE/WGCM CMDP
- With the possible changes, the role of WGNE and WGCM would be more scientifically targeted, rather than trying to address all aspects of metrics research and development
- As discussions continue leading to the March 2019 WGCM meeting, WGNE will be kept engaged in a possible organization transition as well as progress on the precipitation effort