Benchmarking simulated precipitation in CMIP class models and potential synergies with WGNE

WGNE 35 (DWD, Offenbach)

Peter J. Gleckler
PCMDI/LLNL

September 26, 2019

LLNL-PRES-990924

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

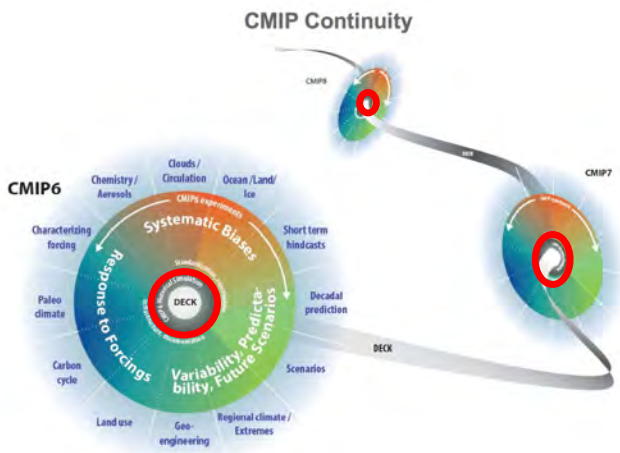
Lawrence Livermore National Laboratory

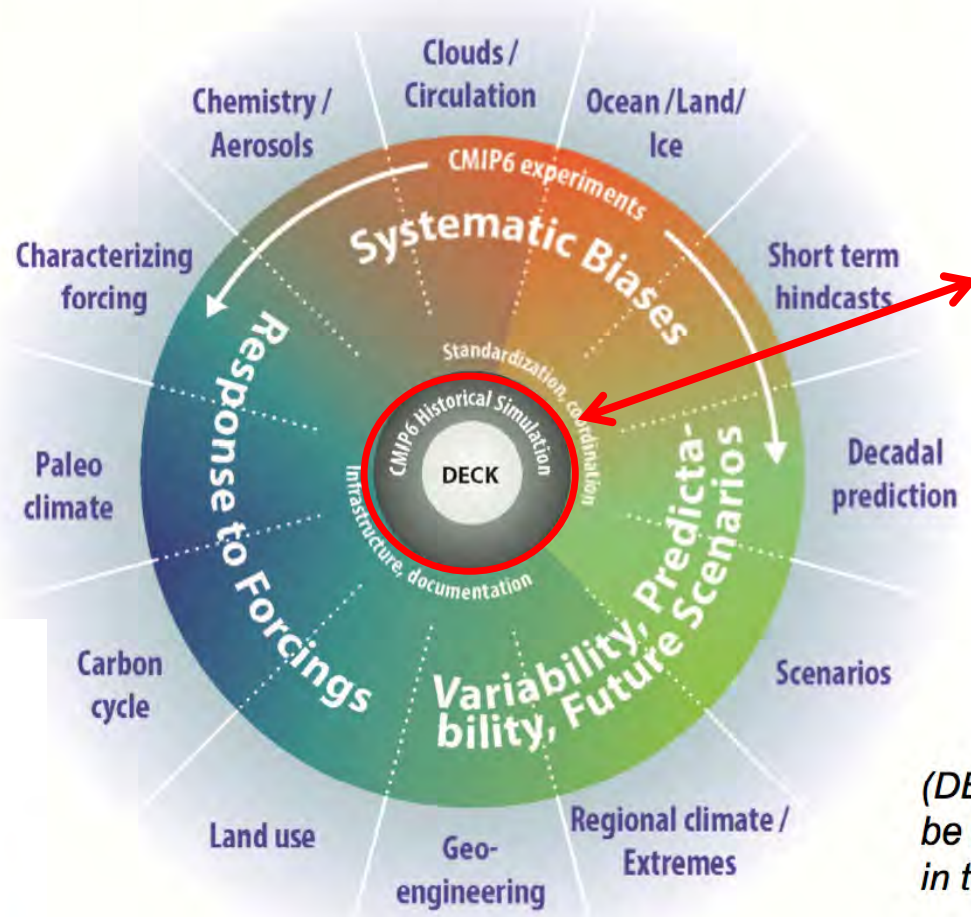
Outline

- Background
- Model evaluation capabilities and their use in CMIP6
- Progress and plans: An initiative to objectively monitor improvements in simulated precipitation

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

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





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

An incomplete listing of community-based capabilities that are relevant for routine benchmarking of CMIP DECK simulations

- *ESMValTool (Eyring et al, GMD, 2016; Eyring et al, GMD, 2019)
- *PMP (Gleckler et al., EOS, 2016)
- *CVDP (Phillips et al., 2014)
- *ILAMB (Collier et al., 2019)
- NOAA MDTF (Maloney, 2019)
- CFMIP diagnostics (Tsushima, 2017)
- TECA (Prabhat et al., 2012)
- ARM Diagnostics package (Zhang et al., 2018)
- MJO Task Force (Ahn et al., 2018; many others)
- CLIVAR basin panels
- ….

These efforts complement CMIP peer-reviewed research publications

*Now being used routinely to provide quick-look results of CMIP6 simulations

Community-based model evaluation and benchmarking in CMIP
WCRP's guiding role

- Joint WGNE/WGCM Metrics and Diagnostic Panel (2012-2017)
  - Raised the profile of objective performance testing of climate models
  - Inspired new research and the development of routine evaluation tools
  - Broad scope of original panel limited potential synergies with JVWGR
- **A more focused approach – targeting simulated precipitation - could help advance research connections between CMIP benchmarking and NWP**
- Possibilities for a new overarching WCRP model evaluation panel under discussion within the longer term context of the WCRP strategic plan (WGCM 2019, E. Guilyardi)

Benchmarking simulated precipitation

Courtesy C. Jakob

Why precipitation?

- 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

Progress:
Advancing the benchmarking of simulated precipitation in CMIP

- Concept introduction to WGCM in 2016 (C. Jakob)
- Original planning discussions with WGNE in 2018 (P. Gleckler)
- 2018 Fall AGU DOE Town Hall: Improving models through the benchmarking of simulated precipitation
- DOE Precipitation Workshop July 1-2, 2019

Townhall & workshop team:
Christian Jakob (WCRP community engagement)
Angie Pendergrass (precipitation benchmarks)
Peter Gleckler (metrics and implementation)
Ruby Leung (advancing process-oriented tests)
Renu Joseph (DOE involvement)

Precipitation Metrics Workshop

https://climatemodeling.science.energy.gov/news/doe-host-precipitation-metrics-workshop

- Inspired by the lack of objective and systematic benchmarking of simulated precipitation
- Community input via DOE 2018 AGU Town Hall and international modeling working groups
- Date/venue: July 1-2, 2019 in Rockville, Md

Identify targets for improvement

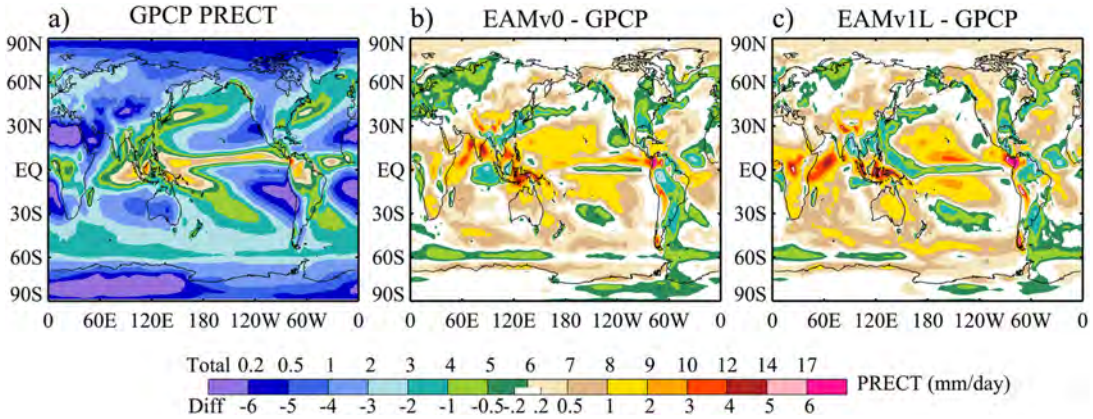
Team of experts identifies useful measures for gauging how well models simulate precipitation

Develop capability to gauge model quality

Baseline metrics incorporated into a model evaluation capability and used to assess current models

Improve simulated Precipitation

Modelers provided with metrics capability to serve as a target for improving newer model versions



Xie et al., 2018
https://doi.org/10.1029/2018MS001350

Establishing a pathway to help guide modelers

- Select a limited set of established benchmarks and develop a strategy for implementing them in a model evaluation capability
- Define how to use this capability for baseline evaluation
- Address the multiscale nature of precipitation, including the existence of model errors at all scales scales
- Identify key research areas where exploratory work can yield more in-depth and informative metrics to include
- Challenge the modeling community to use the expert groups' evaluation metrics as a guide to improve their models; quantify improvement in the next generation of models

DOE workshop
Participants with a diverse spectrum of expertise

Workshop attendees included

- Model developers interested in improving simulated precipitation
- Observational experts liaising with international teams
- Experts in model analysis of precipitation including the mean state, a broad range of variability and event characteristics including extremes, precipitation distribution rates
- Practitioners gauging model-obs agreement with performance metrics
- Experts in impact-related and use-inspired metrics
- Scientists involved in research topics where established metrics are lacking but desired (fronts, tropical cyclones, atmospheric rivers, etc.)

Outcomes

- An agreed upon list of established baseline precipitation characteristics that is being finessed and will be used to assess current generation ESMs
- A summary of topics that hold promise for developing more in-depth metrics encompassing a broad range of processes and phenomena
- Formation of two working groups: 1) to develop/implement the baseline metrics, and 2) bring together state-of-the-art precipitation analysis

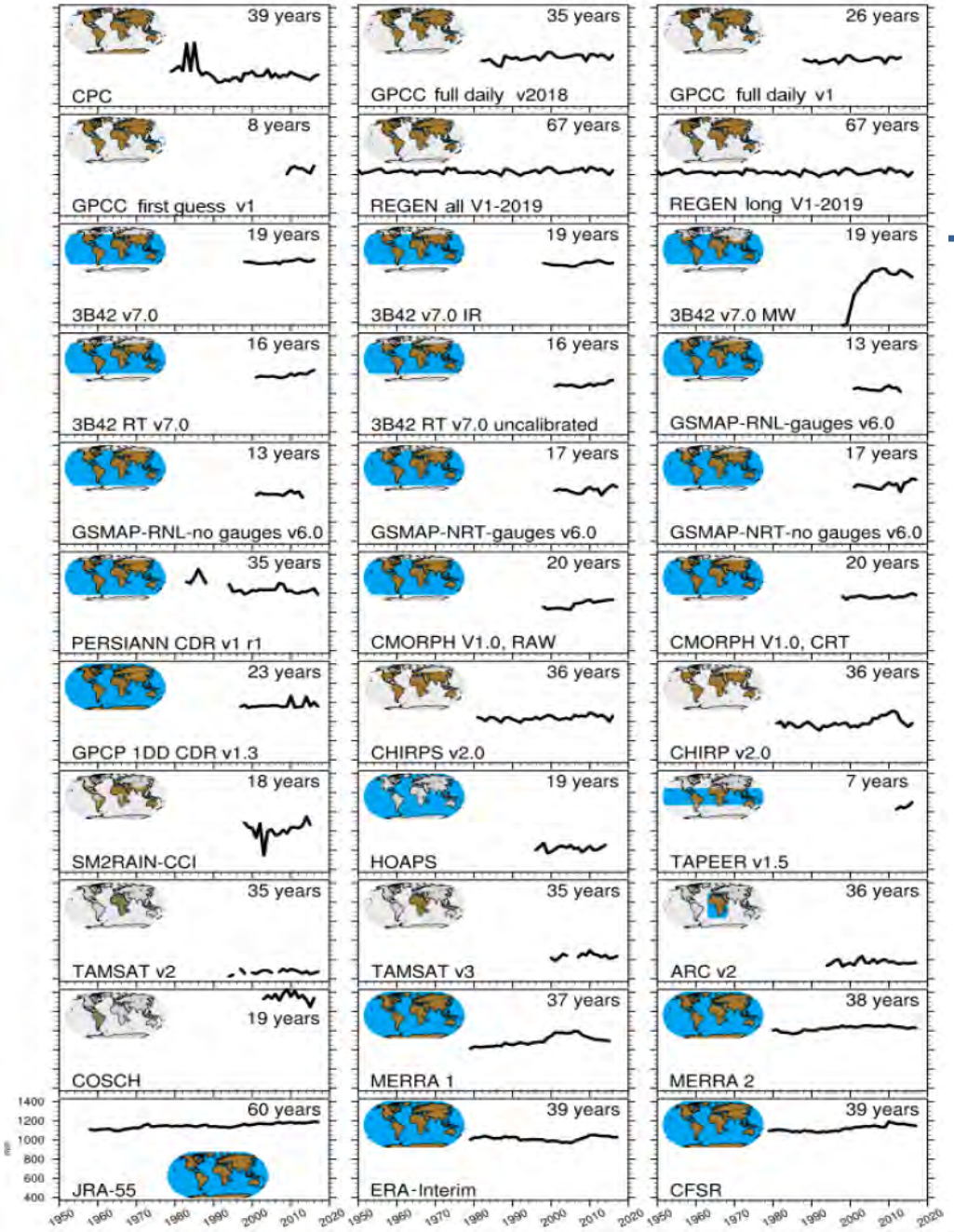
- Outline and plans for DOE & BAMS workshop reports
- Plans for papers advanced by both groups including CMIP analysis
- Strategy for a repository of all codes and data developed by the project to enable community use with all current and future generations of ESMs

GEWEX Data Assessment of Products Work in progress

Time series of annual total daily precipitation (mm) averaged over each dataset domain as shown on the embedded maps in the panels

The name of the dataset and number of years available are indicated in each panel.

FROGS: a daily 1° × 1° gridded precipitation database of rain gauge, satellite and reanalysis products (Roca et al., 2019)

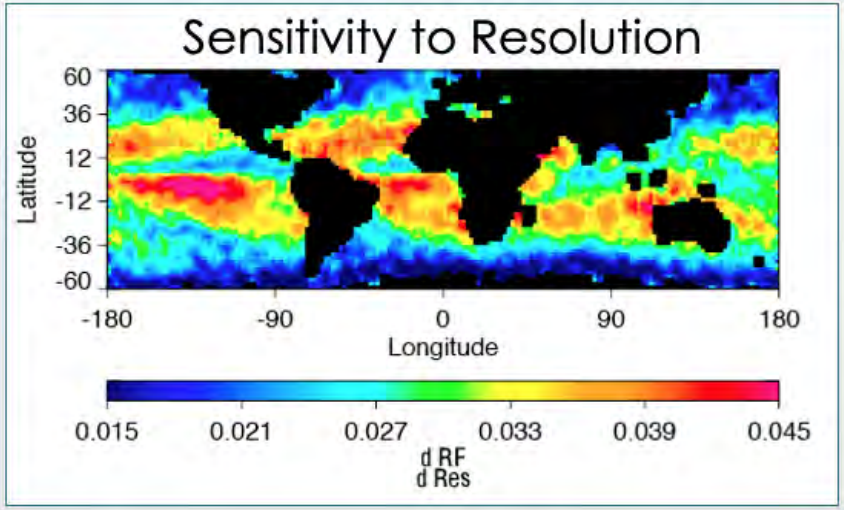


Roca, et al.: FROGS: a daily 1° × 1° gridded precipitation database of rain gauge, satellite and reanalysis products, Earth Syst. Sci. Data, 11, 1017–1035, https://doi.org/10.5194/essd-11-1017-2019, 2019.

Challenges associated with observational deficiencies

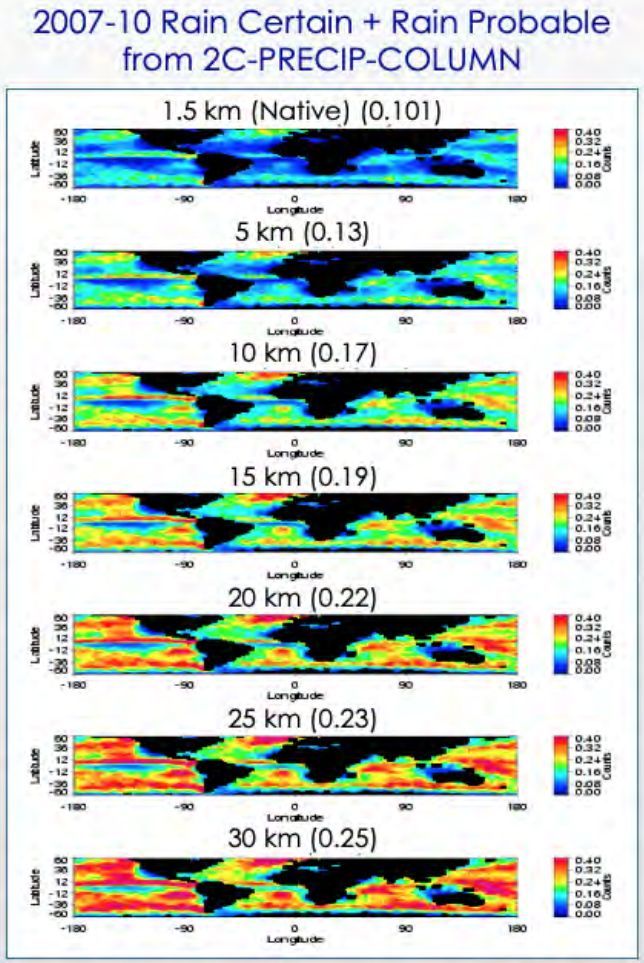
- Example thresholds for precipitation occurrence and phase. The frequency of precipitation depends strongly on the scale.
- On smaller scales, precipitation frequency generally decreases with increasing spatial resolution as seen clearly with CloudSat observations.

Rain Probability Depends on Scale



- The effects of spatial resolution on rainfall occurrence vary significantly with rainfall type.

Courtesy T. LeCuyer



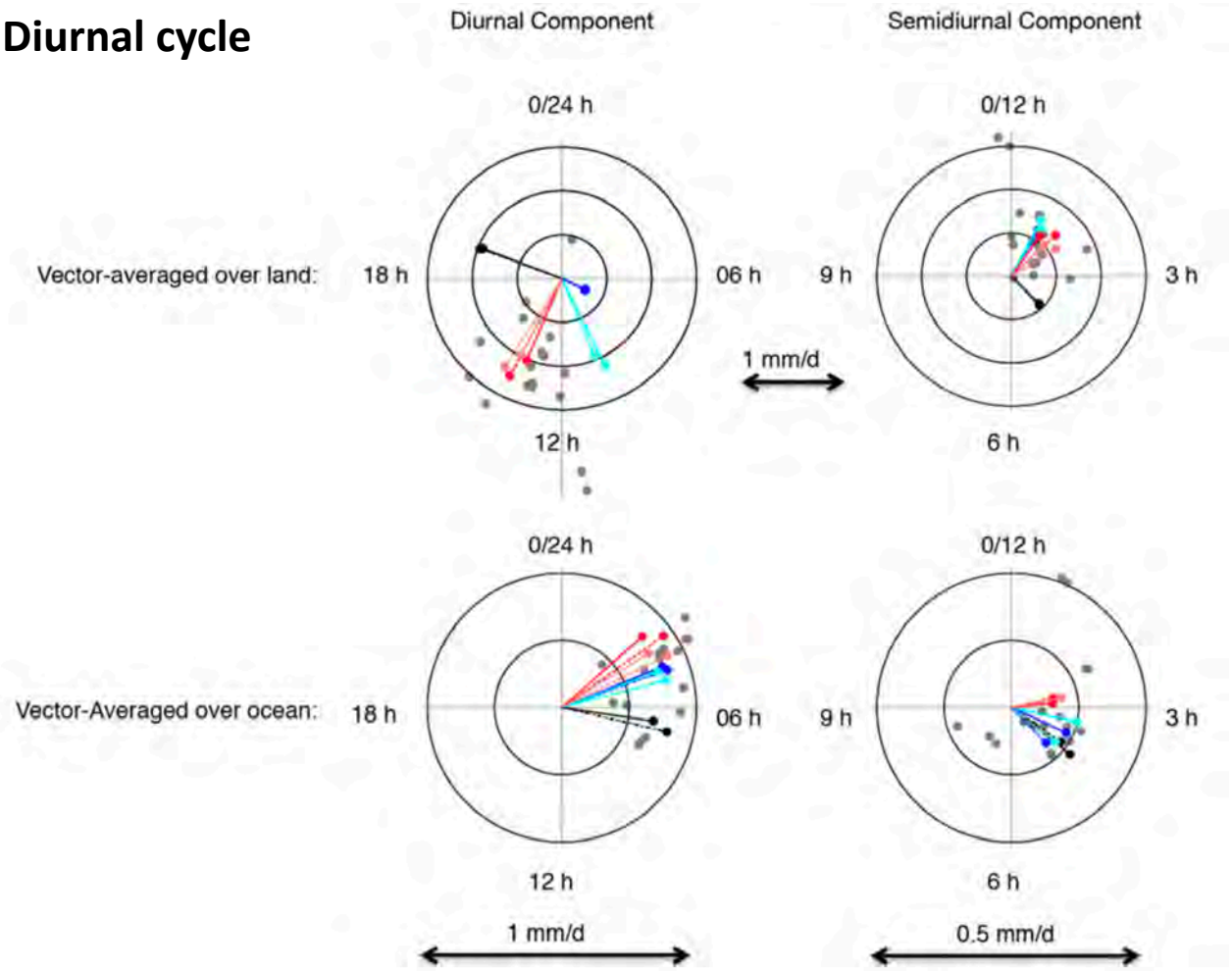
Baseline metrics: Tiers 1 and 2
Resulting from workshop

- **Spatial distribution of mean state**
  - RMS error / MAE of mean state
  - Pattern correlation
  - Monthly mean snow amount
- **Seasonal cycle**
  - Amplitude+phase of seasonal cycle (first two harmonics)
  - or: Monthly score (RMS error)
- **Variability across timescales**
  - Standard deviation at different timescales
    - Daily, weekly / synoptic, intraseasonal, interannual, ENSO
    - Absolute and relative
    - Seasonal breakdown
  - Diurnal cycle – phase and amplitude
- **Intensity / frequency distributions**
  - Simple Daily Intensity Index (SDII)
  - Unevenness (number of days for half of annual precip)
  - Mean and variance of daily precip
    - Cutoff precip rate
    - Power law scale
  - Perkins score (goodness of fit) - various moments
  - Fraction of precipitating days
- **Extremes**
  - Rx1day
  - Rx5day
  - 20-y return values (from GEV)
  - Rx3h
  - Seasonal breakdown
- **Drought (lack of precip)**
  - Frequency of SPI spells below a threshold
  - Consecutive Dry Days

Courtesy Angie Pendergrass (NCAR)

Benchmarking simulated precipitation
A few examples

Diurnal cycle

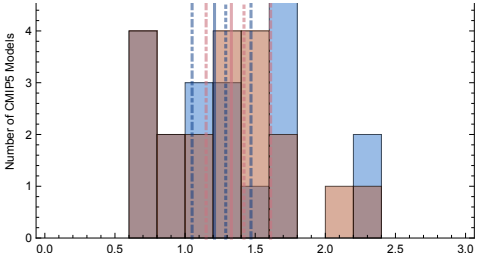


Covey et al., 2016, 2018

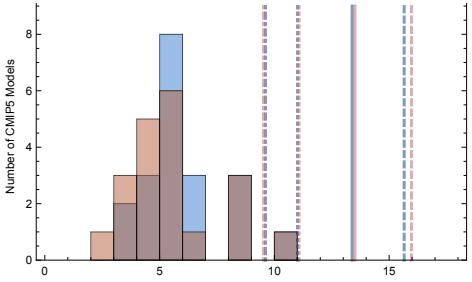
Variance decomposition

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

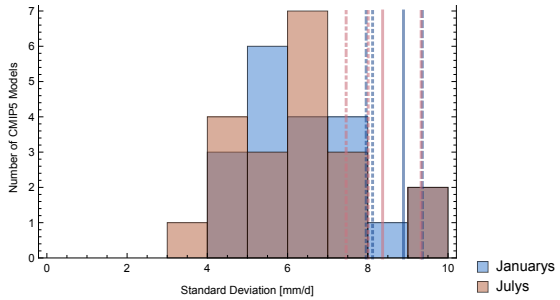
s.d. of Mean Diurnal Cycle



s.d. of irregular sub-diurnal variations

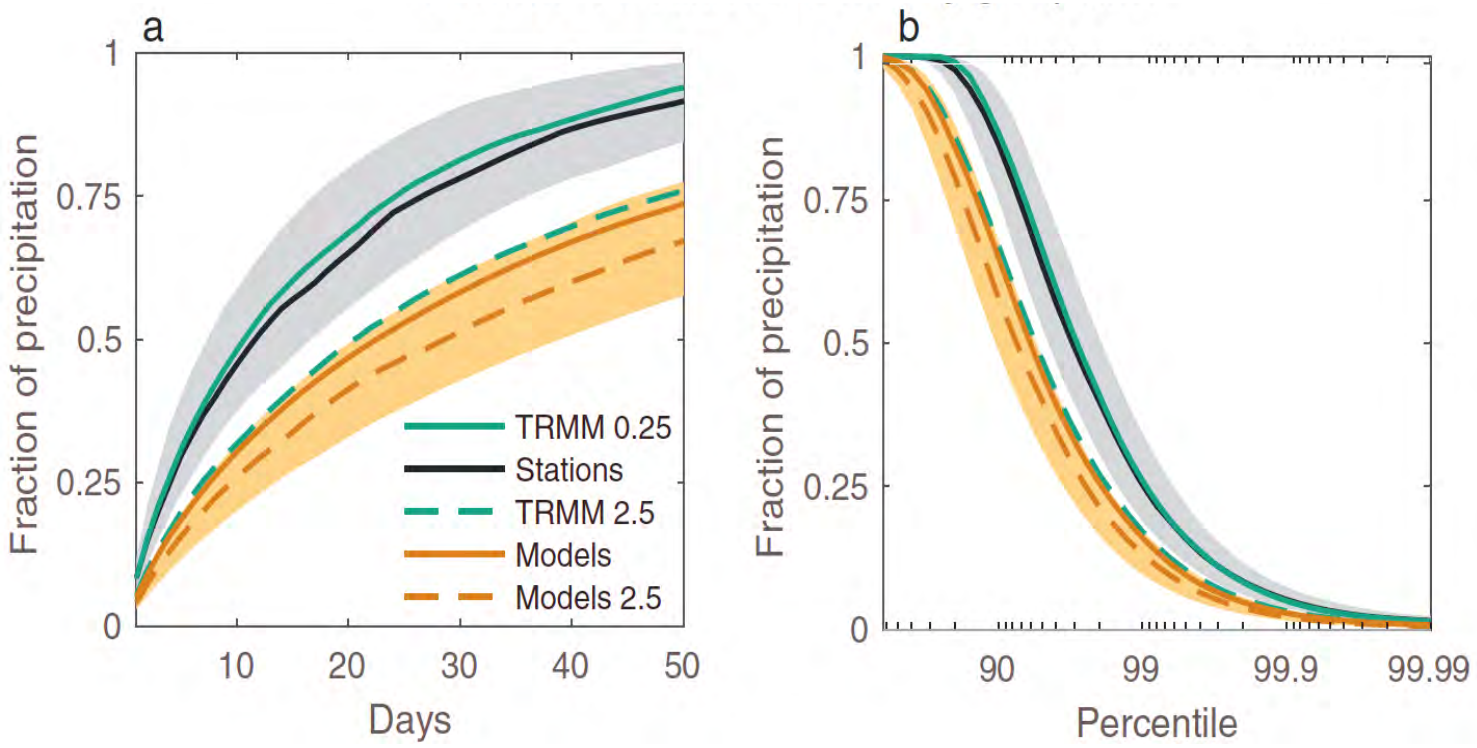


s.d. of daily mean values



Benchmarking simulated precipitation
A few examples

Models underestimate unevenness, even when resolution is accounted for



Pendergrass and Knutti (2018) GRL

Exploratory metrics
Team led by R. Leung (PNNL)

- Purpose: motivate the need for exploratory metrics and the objectives, and demonstrate the value of process-oriented metrics using some examples
- Model output: low and high resolution simulations from DECK historical and HighResMIP
- Metrics ready for demonstration
  - Coherence in space and time
  - Frontal precipitation
  - Top 10 precipitation events
  - Convection onset
  - Orographic enhancement
  - Monsoon
  - MCS precipitation
  - MJO precipitation
  - AR diagnostics
- Form small groups to collaborate on the above metrics

Scope: Baseline metrics
CMIP model experiments and output

- Must be able to be calculate for CMIP6 DECK + Historical simulations with standard output
  - piControl
  - AMIP
  - Historical
  - Data request: monthly, daily, and 3h mean precip, monthly prsn

- Although the initial target is CMIP class models, connections with the JWGFVR could substantially strengthen the effort

Summary I:
Benchmarking simulated precipitation – How?

- Step 1: An assessment report and review paper objectively gauging the current generation of models
- Step 2: Enable modelers to apply metrics (i.e., provide code and data)
- 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 with the next generation of models

Summary II:
Benchmarking simulated precipitation

- Multiple potential synergies within WCRP:

  GDAP, GEWEX GASS, etc.

- JVWGR involvement and expertise could make effort useful for NWP centres