

An update on the WGNE/WGCM Climate Model Metrics Panel

Members selected by relevant and diverse experience, and potential to liaison with key WCRP activities:

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

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

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

Peter Gleckler (PCMDI), chair – WGNE, **atmosphere**

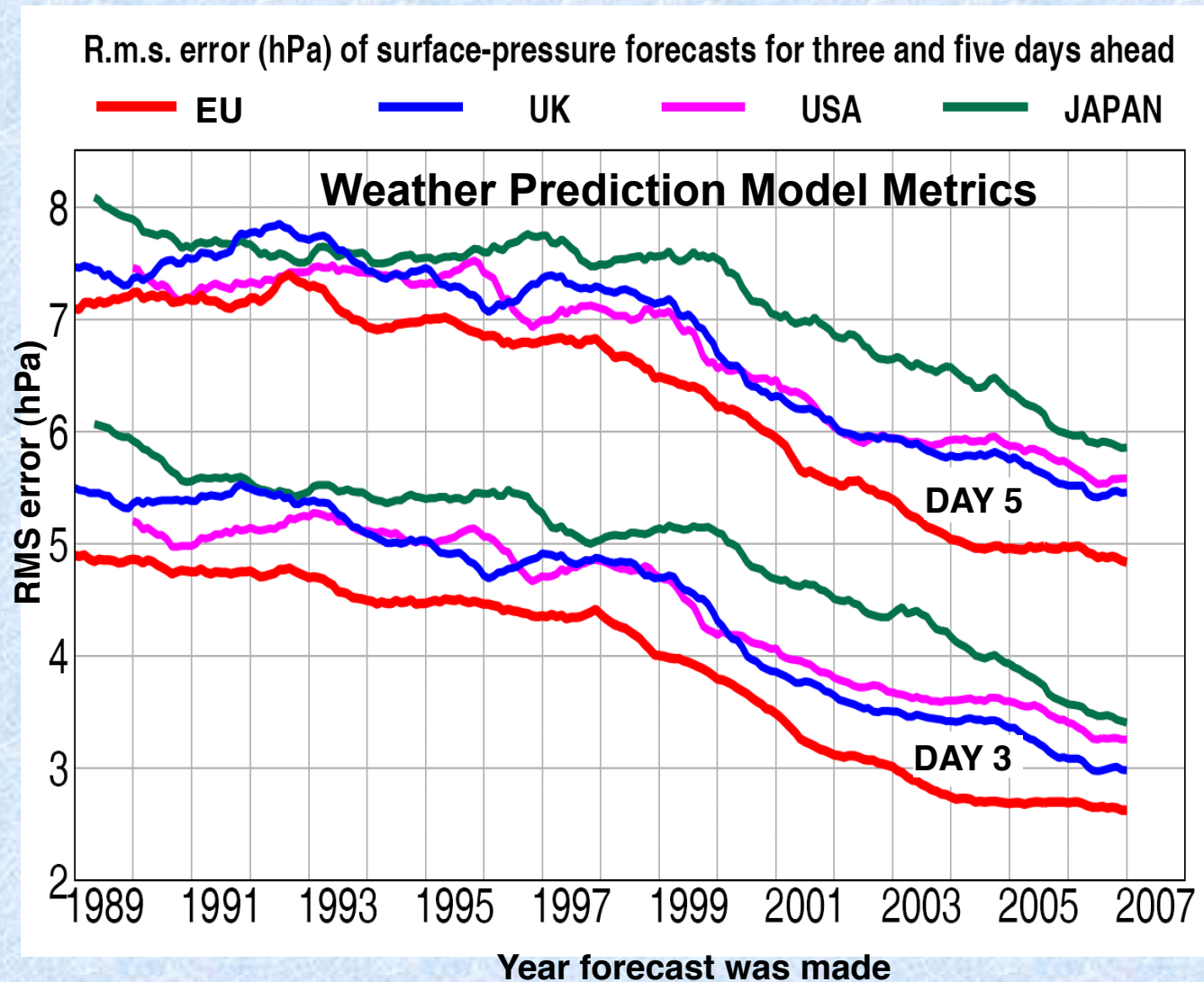
Robert Pincus (NOAA) – GEWEX/GCSS, **clouds/radiation**

Karl Taylor (PCMDI) – WGCM, **CMIP5**

Helene Hewitt (U.K. Met Office) – WGOMD, **ocean and sea-ice**



Monitoring evolution of model performance: Example from Numerical Weather Prediction



The climate modeling community does not yet have routine performance metrics

Courtesy
M. Miller,
ECMWF



What this effort is and is not about...

- Quantifying model agreement with observations, with a broad perspective, not necessarily identifying the causes of model errors
- Assessing different aspects of model skill, but not combining them into an overall measure of model performance
- Providing a useful complement to in-depth diagnosis, not a last word on model performance
- Focusing on *performance metrics* (comparison with observations), not *projection reliability metrics*



Questions motivating routine benchmarks for climate models

- Of direct concern to the WGNE/WGCM metrics panel:
 - Are models improving?
 - Are some models more realistic than others?
 - What do models simulate robustly, and what not?
- Related research drivers, but not (currently) the panel's focus:
 - How does skill in simulating observed climate relate to projection credibility?
 - Can we justify weighting model projections based on metrics of skill?



What opportunities are there to construct climate model performance metrics?

- Model's externally "forced" responses on a range of time-scales:
 - ▶▶▶ Diurnal cycle
 - ▶▶▶ Annual cycle
 - ▶▶▶ Volcanic eruptions, changes in solar irradiance, ...
- Model's "unforced" behavior (weather, MJO, ENSO, NAO, PDO ...)
- Evaluate model representation of individual processes and co-variability relationships
- Test model ability to solve the "initial value" problem
- Examine how well models perform with added complexity



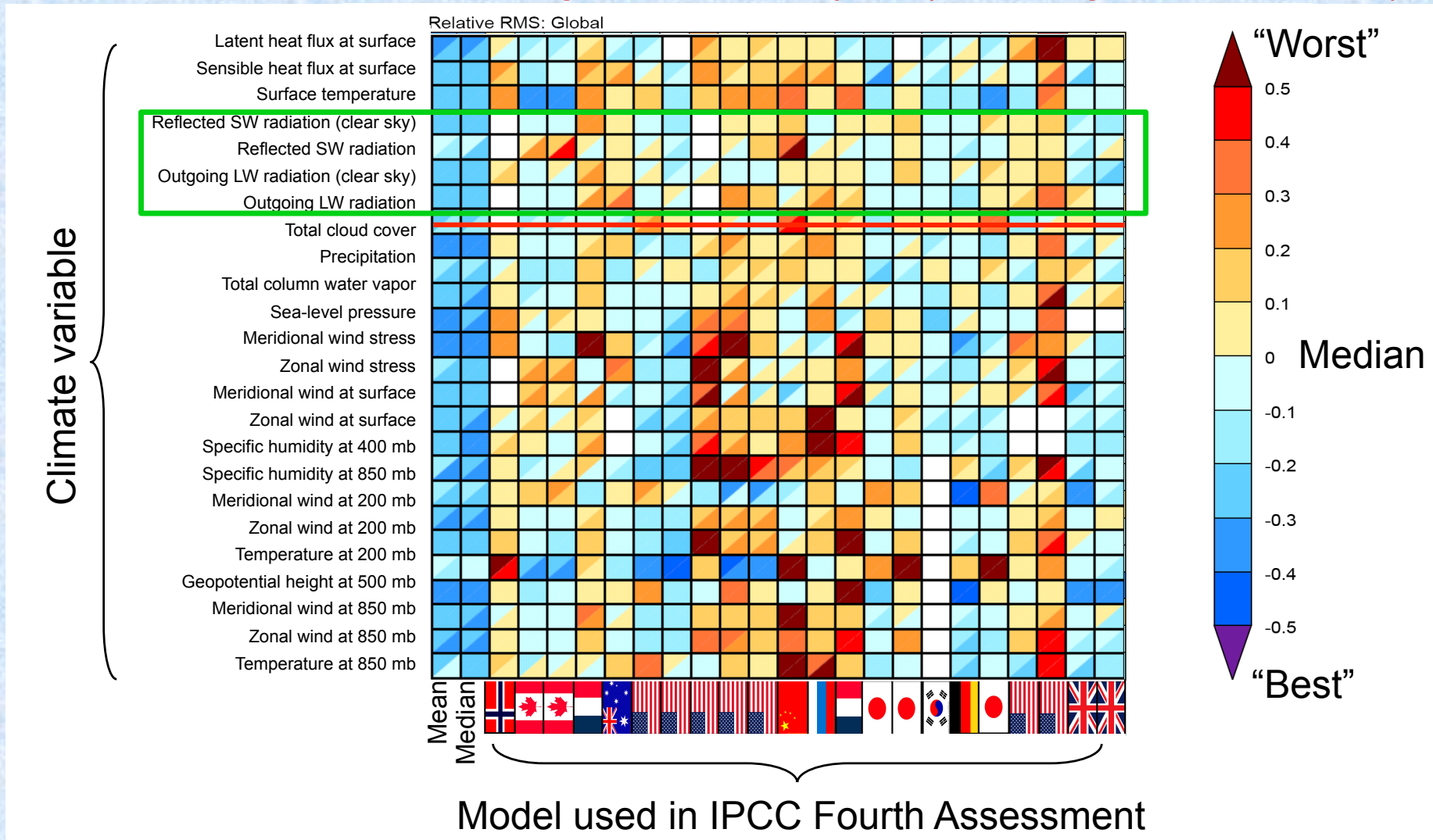
Targeting WCRP/CMIP5 benchmarks experiments

- Panel focus on CMIP5 expts where comparing to observations is most relevant
- Primary (initial) focus:
 - Historically forced AOGCMs, physical aspects of historical ESMs, and AMIP simulations
- But the panel is considering metrics for:
 - Historical ESM emission driven (e.g., metrics for [CO₂])
 - Initial condition experiment: Transpose AMIP



Evaluating how well climate models simulate the annual cycle: A "Performance Portrait" of relative errors

Relative RMSE in Climatological Annual Cycle (including spatial pattern)



Examination of redundancies in metrics results

Yokoi et al., 2011: Application of Cluster Analysis to Climate Model Performance Metrics, J. Appl.Metr.Clim

TABLE 2. Members of the seven clusters for the *K*-means clustering. The mean-bias metrics ($|b_m|$) and the centered-RMSE metrics (c_m) are indicated by italic and boldface type, respectively.

Cluster	Metrics
A	U20c, U50c, U85c, V20c, V50c, V85c, T50c, T85c, Z50c, Z85c, Q50c, Q85c, Tsfc, SLPc, OLRc, CLDc, PRCc
B	<i>U20b, U50b, Z20c, Q30b, Q30c, Q50b, SHFc</i>
C	<i>T50b, T85b, Z20b, Z50b, Tsfb</i>
D	<i>T20b, T20c, OSRc, CLDb, LHFc</i>
E	<i>OLRb, OSRb, PRCb</i>
F	<i>U85b, Z85b, SLPb</i>
G	<i>Q85b, SSTb, SSTc</i>

- Similar metrics to previous studies (e.g., Murphy et al. 2004, Gleckler et al 2008)
- Examine centered RMSE and bias in two cluster analysis methods
- Consistent results: ~7 clusters, with a mix of mean bias and centered-RMSE metrics

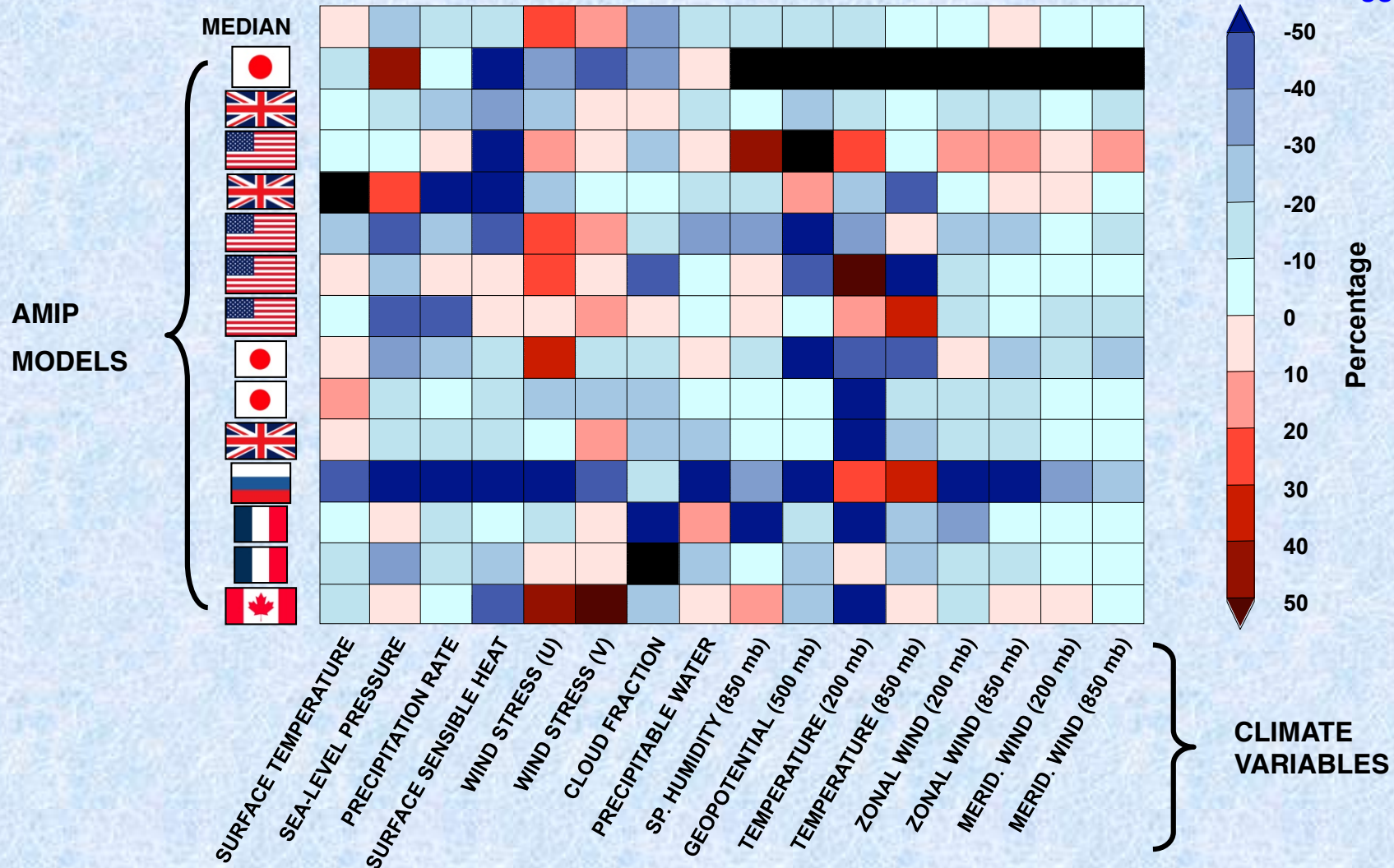


Tracking changes in model performance

What will we find between CMIP5 and CMIP3?

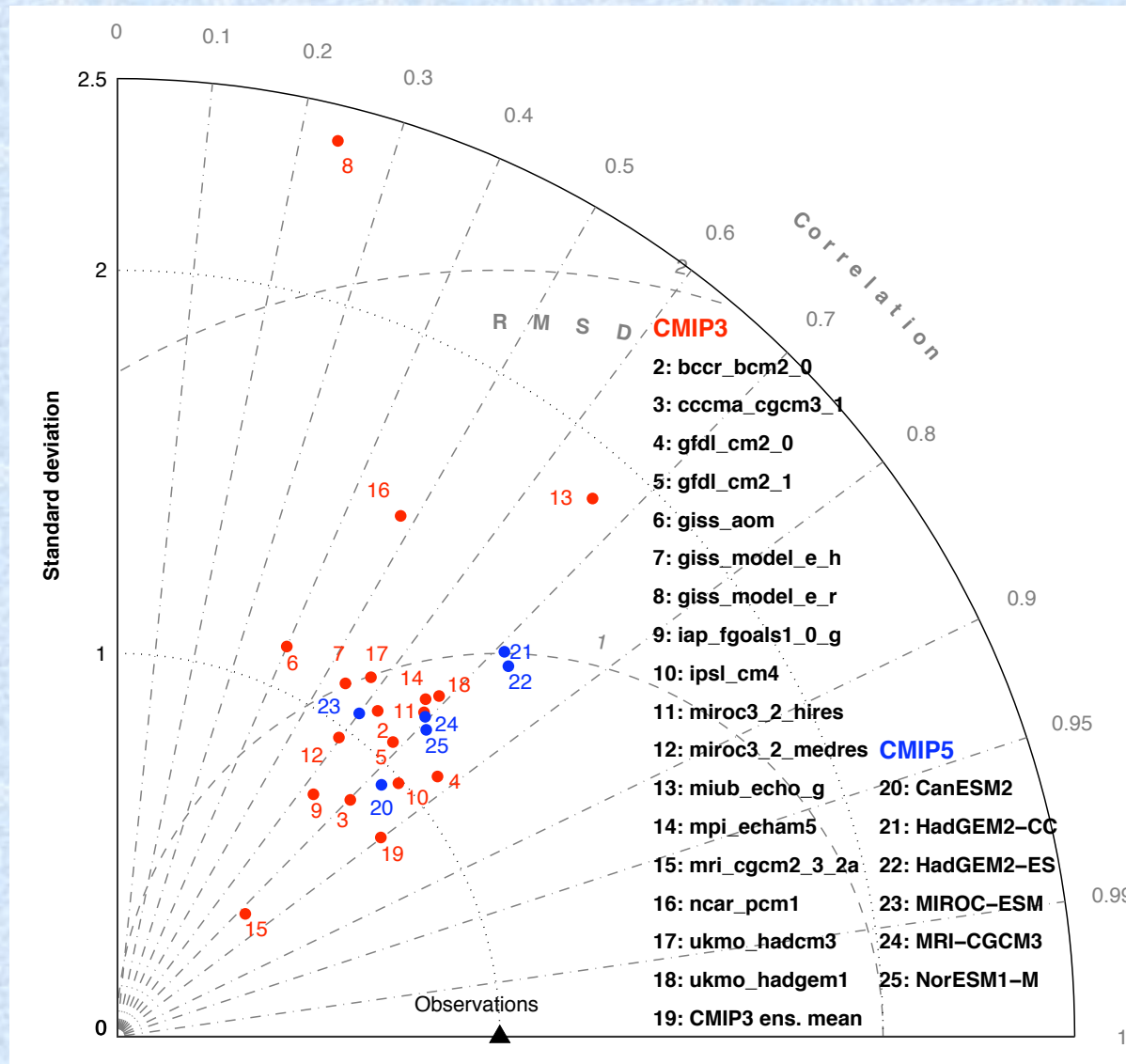
Percentage change in total error (AMIP2-AMIP1)

Global, all seasons



Some early results from CMIP5

Annual cycle Taylor diagram of dynamical sea-surface height



Quantity not extensively analyzed

High quality obs: merged altimetry AVISO product

Integrates dynamical and steric annual cycle changes

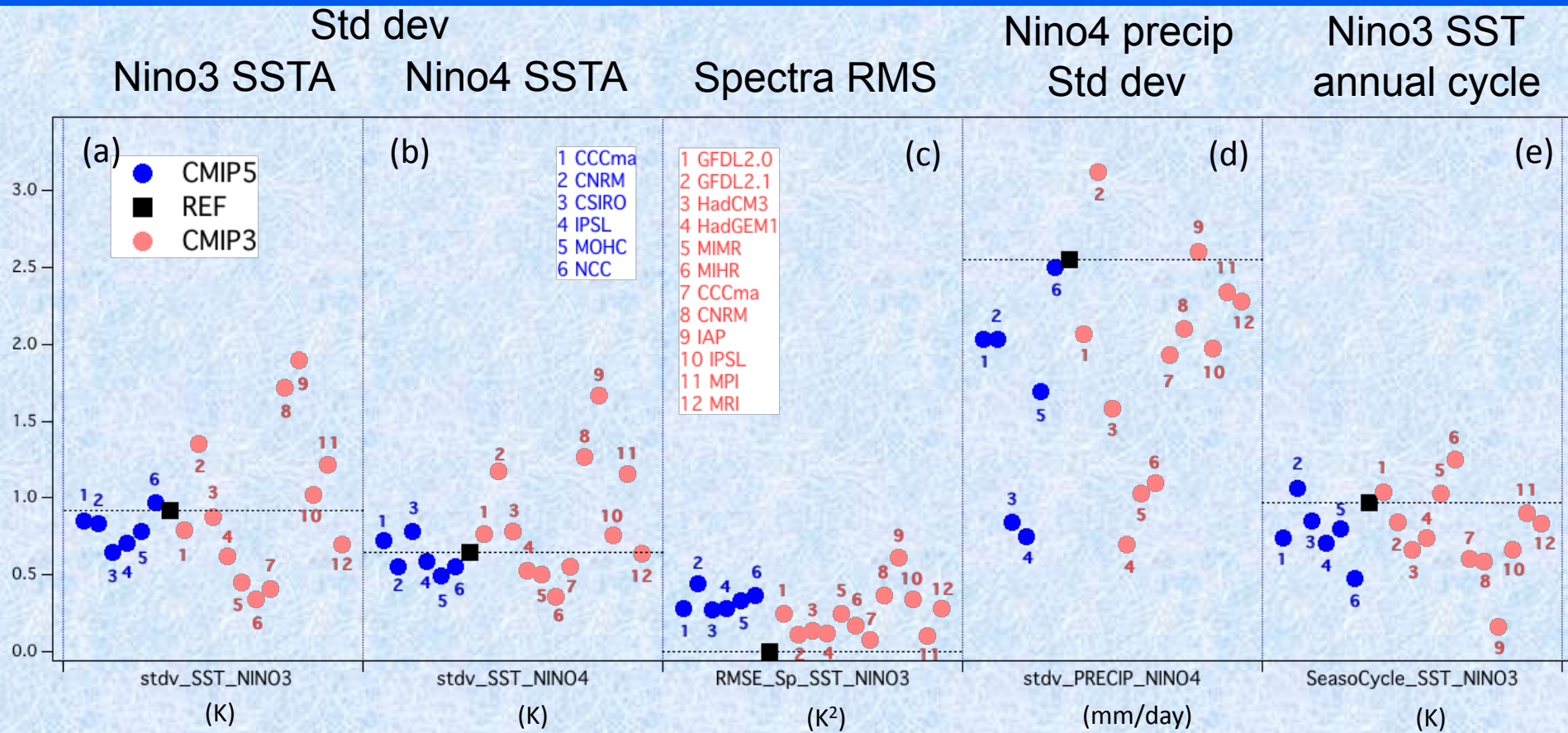
Correlations are relatively low

Too early to tell if CMIP5 results are demonstrably better



Courtesy Felix Landerer, JPL

First look of ENSO in CMIP5 using CLIVAR Pacific Panel Metrics



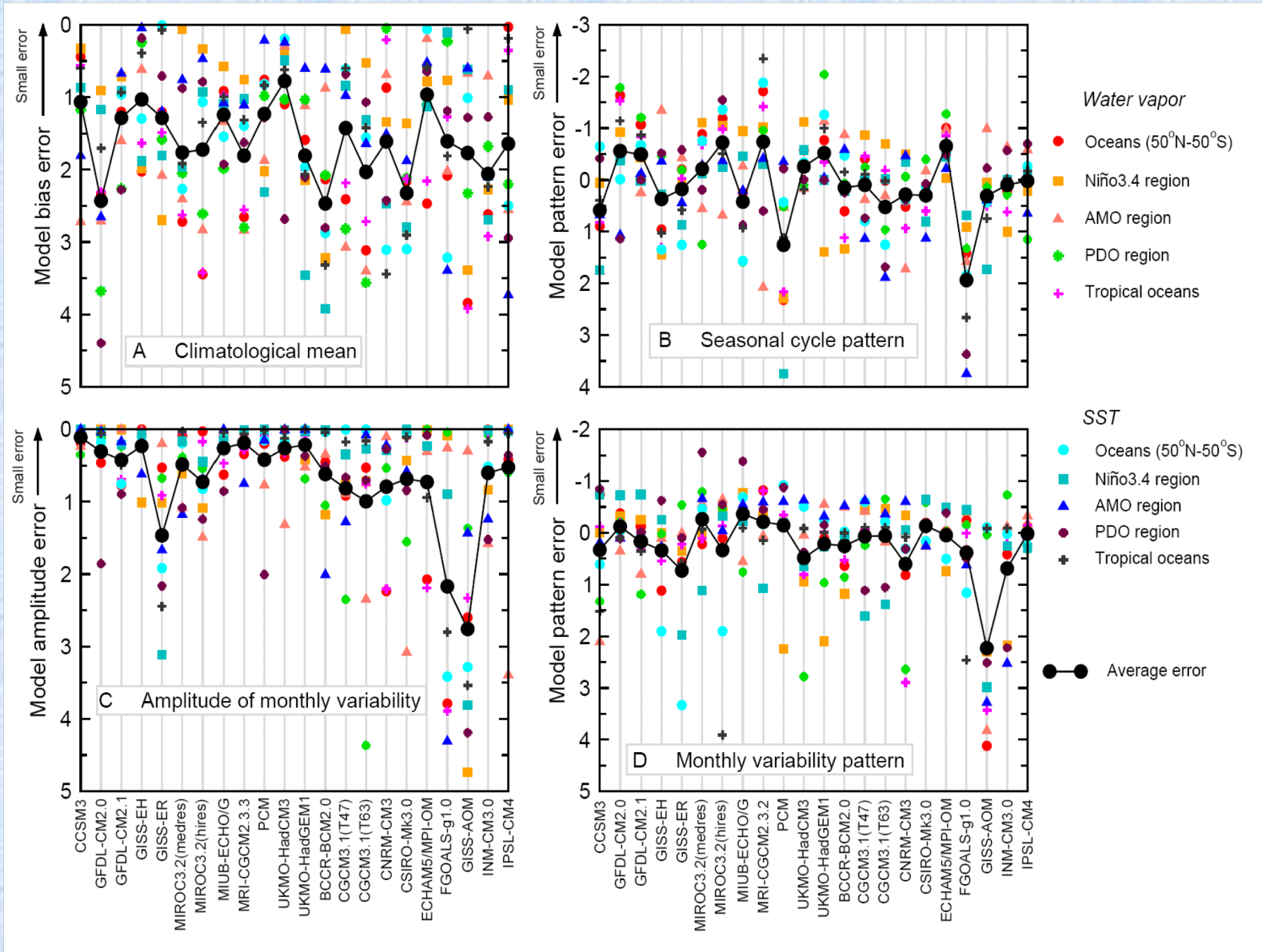
- Some improvement of NINO3 and 4 interannual SST variability (a and b)
- No clear trend for (c) ENSO spectral characteristics, (d) precipitation response and (e) SST annual cycle

Caveat: only 6 CMIP5 models

Courtesy E. Guilyardi, H. Bellenger (LOCEAN/IPSL), A. Wittenberg (GFDL)



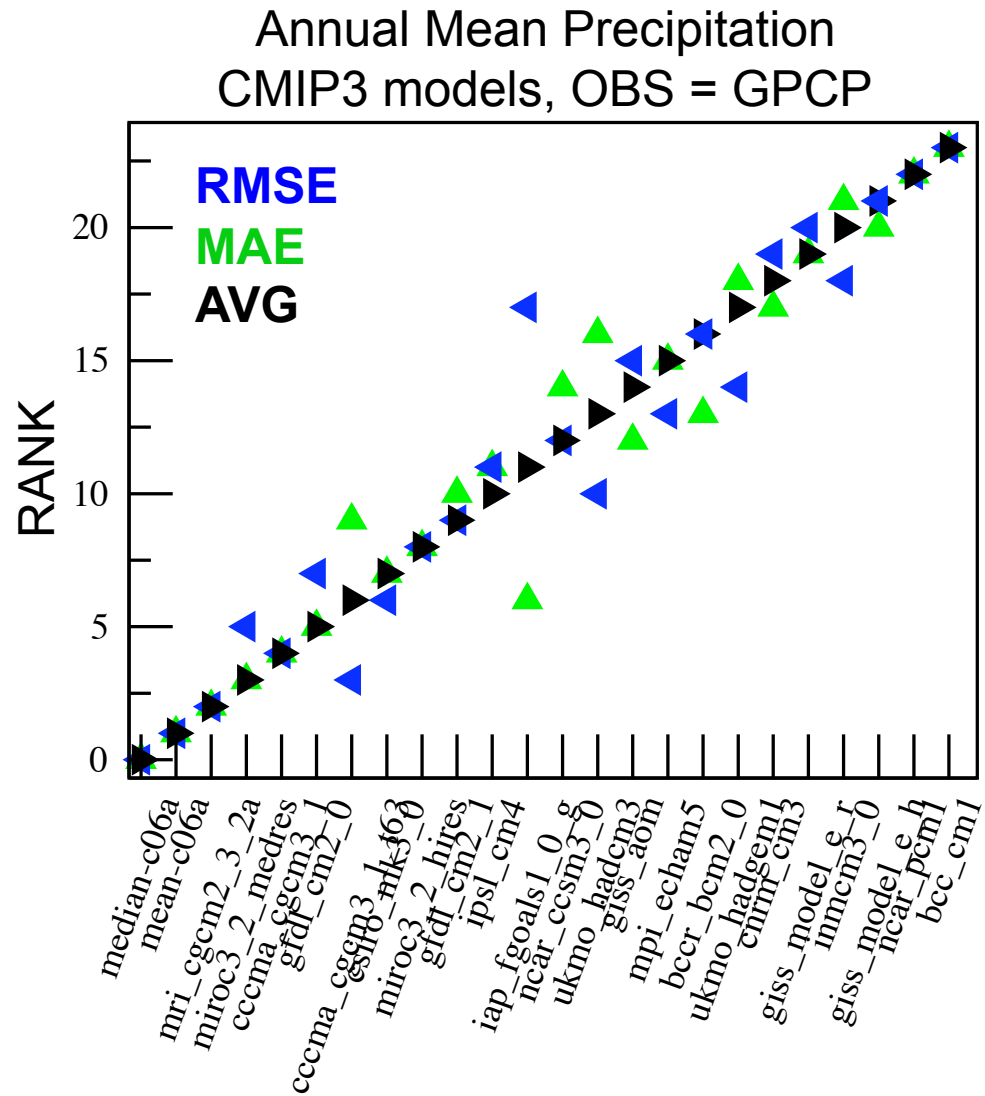
Gauged by simple metrics, the structure of relative model errors is complex



Santer et al., *PNAS*, 2009



What difference does the choice of metric make?



- Choice of metrics can impact rank
- Outliers (good/bad) robust to choice (in this example)

Better to be aware of how results are impacted by choice of metric than to rely on a single score

Metrics panel terms of reference (working version)

- **Identify a limited set of basic climate model performance metrics**
 - based on comparison with observations
 - well established in literature, and preferably in widespread use
 - easy to calculate, reproduce and interpret
 - covering a diverse suite of climate characteristics
 - large- to global-scale mean climate and some variability
 - atmosphere, oceans, land surface, and sea-ice
- **Coordinate with other WCRP/CLIVAR working groups**
 - Identify metrics for more focused evaluation (e.g., modes of variability, process level)
 - Striving towards a community based activity by coalescing expertise
- **Justify and promote these basic metrics in an attempt to**
 - establish routine community benchmarks
 - facilitate further research of increasingly targeted metrics
- **Ensure that these metrics are applied in CMIP5 and widely available**



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

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)



"OBS4MIPS"

- Launched by JPL/PCMDI, with NASA products being the test case
- Guidelines being developed for documenting information about observations of particular importance for model evaluation (e.g., uncertainty estimates)
- Data is 'technically aligned' with CMIP5 model output, i.e., in data structure and metadata
- Traceability is important for the metrics panel effort - having a clear record of exactly which observational product *version* is used



OBS4MIPS: Striving for more traceability in the observational products used for model evaluation

The screenshot shows the ESG-PCMDI Gateway website. The browser address bar displays `://pcmdi3.llnl.gov/esgset/home.htm`. The page header includes navigation links: [Home](#), [Data](#), [Account](#), [About](#), [Contact Us](#), and [Login](#). The main content area is titled "ESG Gateway hosted by the Program for Climate Model Diagnosis and Intercomparison".

Search: Datasets for: [] Search Start Over
To conduct a search, select a category from the pull down menu and/or enter free text into the the text box.

Search Categories

- Project
 - > CMIP5
 - > obs4MIPS
- + Institute
- + Model
- + Experiment
- + Frequency
- + Product
- + Realm
- + Variable
- + Ensemble

Welcome to PCMDI

PCMDI

The Program for Climate Model Diagnosis and Intercomparison (PCMDI) was established in 1989 at the Lawrence Livermore National Laboratory (LLNL), located in the San Francisco Bay area. Our staff includes research scientists, computer scientists, and diverse support personnel.

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 observations 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 putative global climate change.

Quick Links

- [Getting Started Guide](#)
- [Create Account](#)
- [Browse Catalogs](#)
- [Search for Data](#)
- [CMIP5 Home](#)

ESG Data Gateways

- [ESG-BADC Gateway](#)
- [ESG-NCAR Gateway](#)
- [ESG-NCI Gateway](#)
- [ESG-NERSC Gateway](#)
- [ESG-ORNL Gateway](#)
- [ESG-WDCC Gateway](#)
- [NASA JPL Gateway](#)

Quick View

- [Data Overview](#)

Status of the CMIP5 Archive

- 7/7/2011: NCC datasets are now available to all users.
- 7/19/2011: PCMDI data server will be down for maintenance 7/19 16:30 PST. It is expected back online 7/20 17:00 PST.
- 7/20/2011: PCMDI data server is back online.
- 7/20/2011: Because of a processing fault affecting the MOHC HadGEM-ES rcp85 data from 2080 onwards, this data has been withdrawn for the time being. They expect to provide us with corrected data in a matter of weeks

PCMDI

A few other possibilities...

Expertise exists to guide the panel in expanding metrics...

- CFMIP committee
- Transpose AMIP steering committee
- Working Group on Ocean Model Development (WGOMD)

Lacking connections...

- Key variability indices
- Diurnal cycle of precipitation (verified with TRMM)



WGNE/WGCM Climate Model Metrics Panel

Search Titles Text

pjg Settings Logout
FrontPage

[QuickLookMetrics](#) » [metricspanelGroup](#) »
[QuickLookMetrics2011](#) » [OtherMetrics](#) »
[FrontPage](#)

[RecentChanges](#) [FindPage](#) [HelpContents](#) [moreinfo](#) **[FrontPage](#)**

[Edit \(Text\)](#) [Edit \(GUI\)](#) [Info](#) [Unsubscribe](#) [Add Link](#) [Attachments](#) [More Actions:](#)

The WGNE/WGCM Climate Model Metrics Panel is an ad-hoc group that has been jointly established by the [Working Group on Numerical Experimentation \(WGNE\)](#) and the [Working Group on Coupled Modeling \(WGCM\)](#). A primary objective of this panel is to identify and promote a limited set of frequently used performance metrics in an attempt to establish community benchmarks for climate models, and to facilitate research & development of increasingly in-depth objective measures via coordination with other activities of the [World Climate Research Program \(WCRP\)](#).

This effort will provide a quantitative summary of model agreement with observations for several routinely examined aspects of the simulated climate. The limited set of results is expected to compliment a diverse range of more in-depth model diagnosis efforts. [More Information >>](#)

Initial set of WGNE/WGCM metrics (v2011)

- [Criteria](#)
 - [Working version of the WGNE/WGCM metrics \(v2011 is currently under development\)](#)
 - [Quick-look metrics results applied to CMIP5 \(and earlier phases of CMIP where data permits\) - in preparation](#)
 - [Download the panel's quick-look metrics package here - in preparation](#)
-
- [Community-wide diagnostic and performance metrics code repository - in preparation](#)
 - [Other climate model and NWP performance metrics activities](#)
 - [Panel members and terms of reference](#)
 - [Related references](#)

Enabling community contributions: Wiki repository

- A repository is planned for scientists to contribute metrics and diagnostics codes
- Two categories are envisioned:
 - Contributions overseen by the panel, required to follow certain guidelines, (e.g., documentation). An example here is code provided for cloud property metrics following Williams & Webb (2008)
 - Open contributions - minimal restrictions



Next steps engaging with research community

- Feedback from WGNE and WGCM (this week)
- Modeling groups to be given opportunity to comment on wiki / selected metrics
- Metrics panel wiki (to be made public in early 2012):
 - Discussion of metrics, their limitations, panel goals
 - Posting metrics results for all CMIP3 and CMIP5 simulations
 - Openness: all observations, codes and documentation made public
 - A resource for CMIP5 analysts and possibly modelers
 - Pointers to and possibly discussion of relevant metrics research
 - Repository of community contributions



The metrics panel effort going public

- Poster at WCRP OSC (session C34)
- Presentation at the 5th International Verification Methods Workshop (Melbourne, December 5-7)
- Wiki going public
- Possible description of panel effort in EOS (e.g., to advertise repository)
- A publication is planned to highlight CMIP5 / CMIP3 comparisons (in time for July 31 2012 deadline for AR5)



Several possible directions for the metrics panel

Panel's list is gradually augmented and annually reviewed by WGNE and WGCM, along with the panel's terms

OR

Panel tasked to finalize its metrics list, publish CMIP5 results, and disband

