

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, **CCMI, ESMs**

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

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

Simon Marsland (CSIRO) – WGOMD, **ocean**

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

Karl Taylor (PCMDI) – WGCM, **CMIP5, atmosphere**

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

Metrics panel terms of reference (working version)

Identify a limited but diverse set of climate model performance metrics

- based on comparison with observations
- well established in literature, and preferably commonly used
- easy to calculate, reproduce, interpret and are fairly robust
- 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., variability modes, 'process' level)
- striving towards a community based activity by coalescing expertise

Justify and promote these basic metrics in an attempt to

→ establish **routine** performance benchmarks

→ facilitate further **research** of increasingly targeted metrics

Identifying routine metrics

Basic mean state and annual cycle:

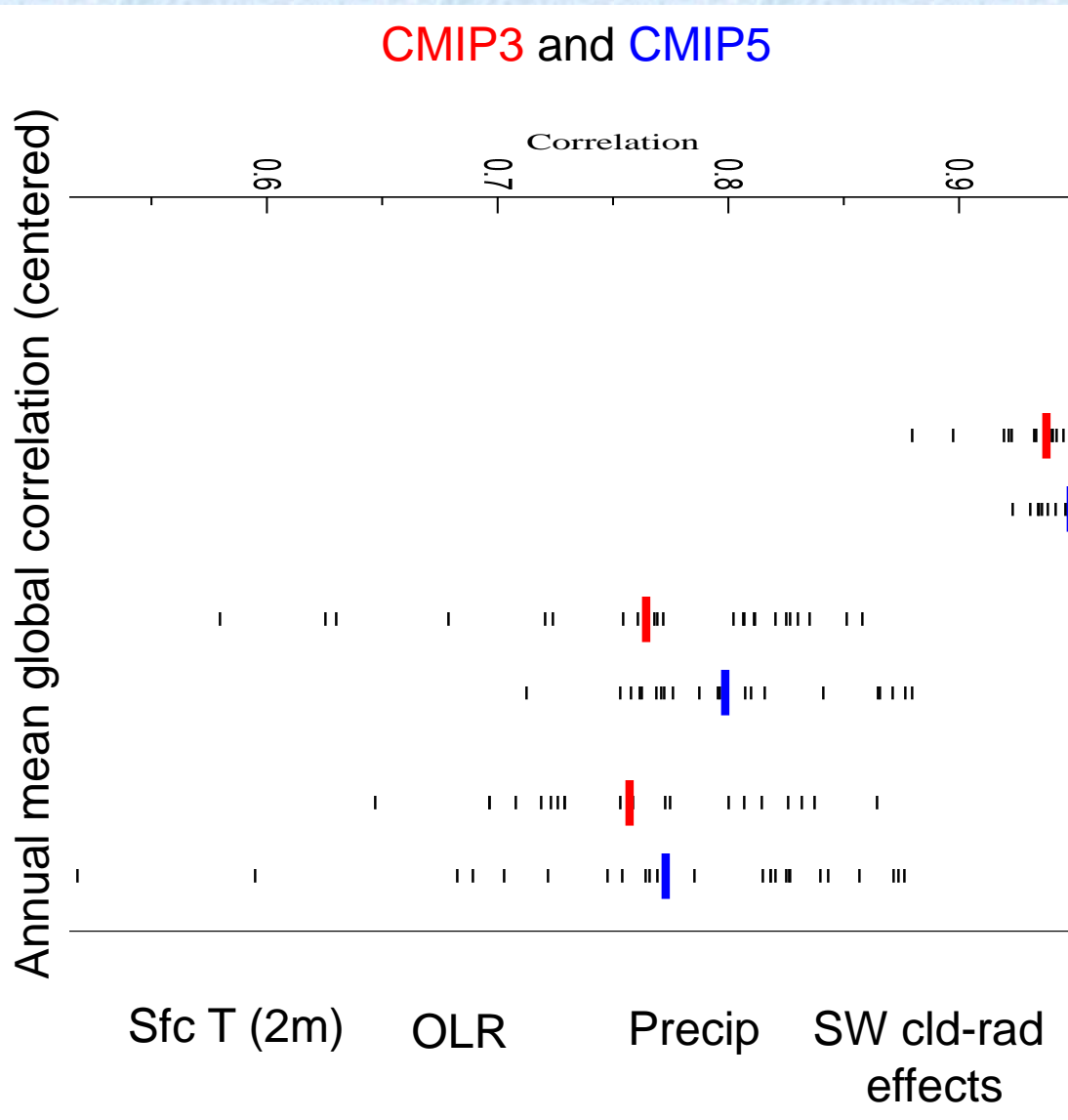
- Large- to global- scale evaluation (global, tropical, NH/SH extra-tropics)
- 20 year climatologies: Annual and seasonal means
- Routine metrics: bias, centered RMSE, MAE, correlation, S.D.
- Field examples: OLR, T850, precip, SST, SSH, sea-ice extent
- Observations: multiple for most cases

Towards an extended set of metrics, coordinating with other working groups (in progress):

- ENSO (CLIVAR Pacific Panel)
- Monsoons (CLIVAR AAMP)
- MJO (YOTC Task force)
- CFMIP committee
- WGOMD
- Carbon cycle in emission-driven ESMs (ILAMB)
- Chemistry-Climate (CCMVal, CCMI) . . .

Tracking model performance

Incremental improvement from CMIP3 to CMIP5



Ensemble average results in CMIP5 are incrementally better than CMIP3

Likely due to a reduction in spread, with fewer poor outliers

Examining redundancies in mean state metrics

Yokoi et al., 2011: *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>

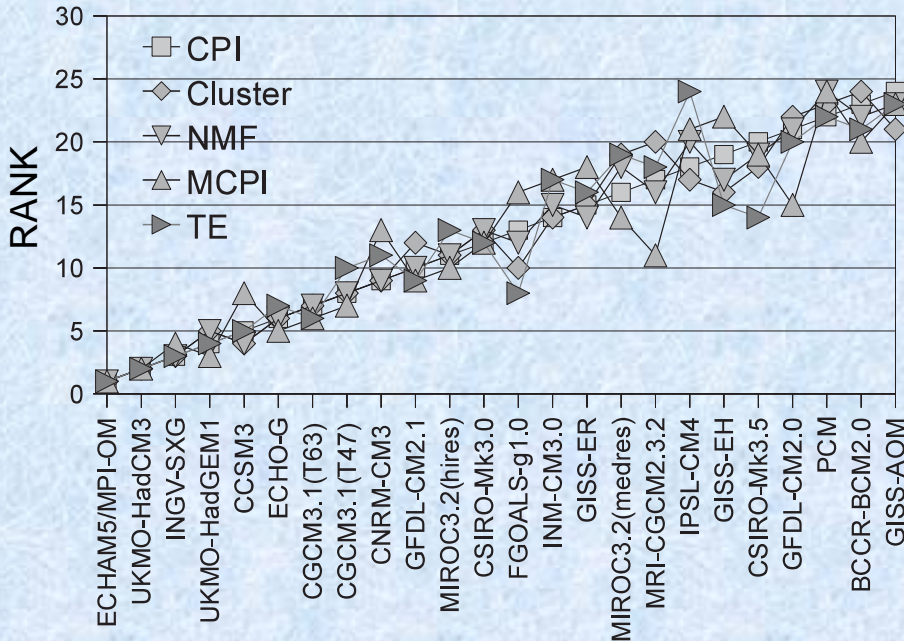
- Similar metrics to previous studies (e.g., Murphy et al. 2004, Gleckler et al 2008, Pincus, 2008)
- Compare results from two cluster analysis methods
- Methods yield similar results : ~7 clusters, with a mix of mean bias and centered-RMSE metrics

Summarizing mean climate performance

Nishii et al., 2012, *JAMS*

An index based on total ATM “total energy” yields similar results to other, more comprehensive measures (e.g., CPI).

At this stage the panel is not advocating overall skill scores, but there is now evidence that at some level results are robust to how such indices are being constructed



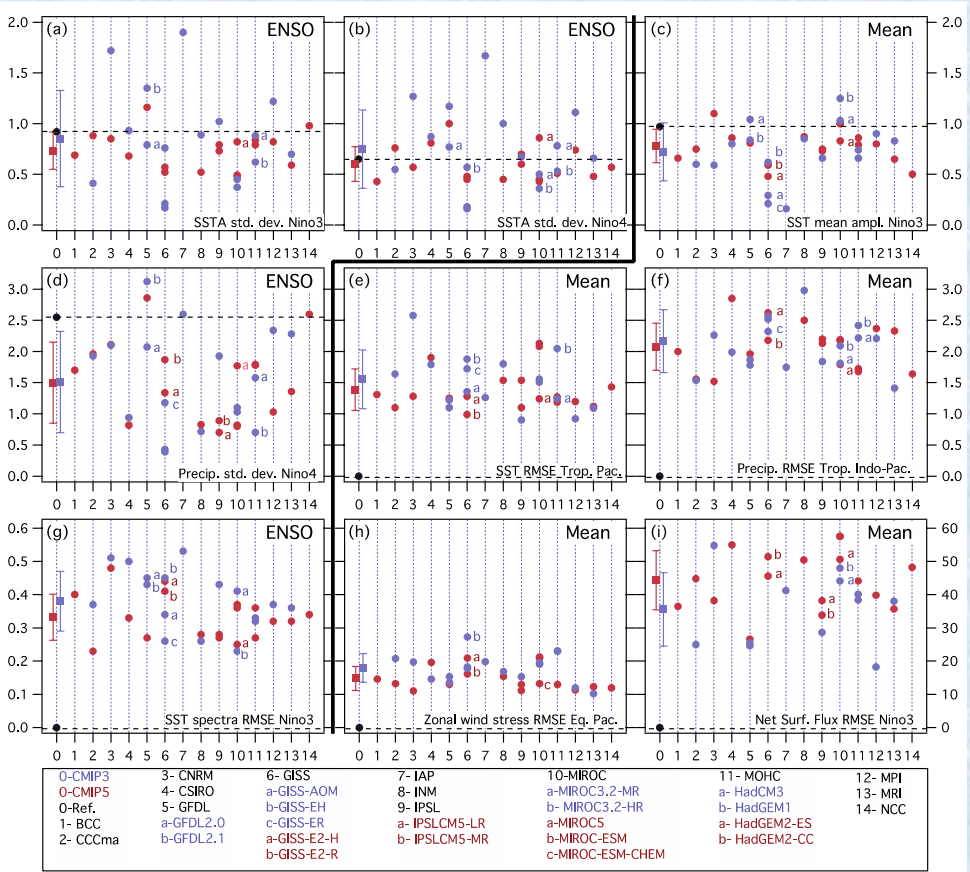
$$TE = \frac{1}{2} \iint \left\{ u'^2 + v'^2 + \frac{C_p}{T_r} T'^2 + RT_r \left(\frac{p'_s}{p_r} \right)^2 + \frac{L^2}{C_p T_r} q'^2 \right\} dA dp$$

ENSO metrics

Recommendations from CLIVAR's Pacific basin panel metrics group

A first look at ENSO in CMIP5

Eric Guilyardi^{1,2}, Hugo Bellenger¹,
 Mat Collins³, Samantha Ferrett³,
 Wenju Cai⁴, Andrew Wittenberg⁵



0- CMIP3	3- CNRM	6- GISS	7- IAP	10- MIROC	11- MOHC	12- MPI
0- CMIP5	4- CSIRO	a- GISS-AOM	8- INM	a- MIROC3.2-MR	a- HadCM3	13- MRI
0- Ref.	5- GFDL	b- GISS-EH	9- IPSL	b- MIROC3.2-HR	b- HadGEM1	14- NCC
1- BCC	a- GFDL2.0	c- GISS-ER	a- IPSLCM5-LR	a- MIROC5	a- HadGEM2-ES	
2- CCCma	b- GFDL2.1	a- GISS-E2-H	b- IPSLCM5-MR	b- MIROC-ESM	b- HadGEM2-CC	
		b- GISS-E2-R		c- MIROC-ESM-CHEM		

Use of pre-industrial control runs is recommend rather than historical runs for sampling considerations

Tropical mean state and inter-annual variability measures (std. dev.)

Cloud related metrics?

Some examples:

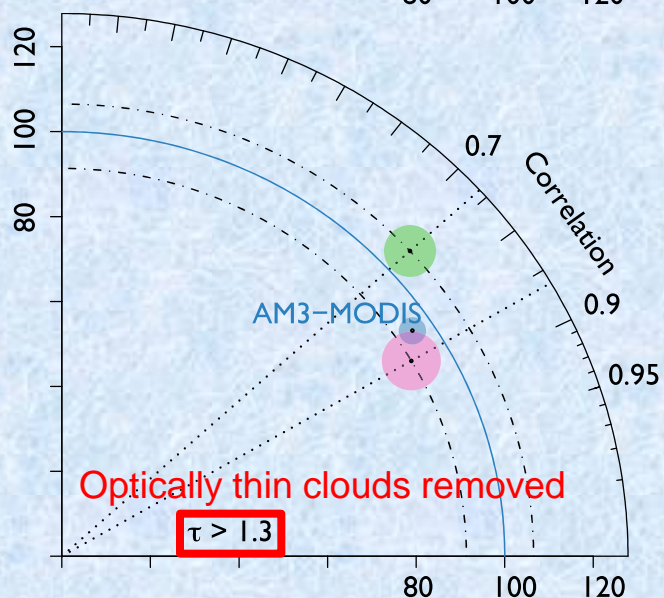
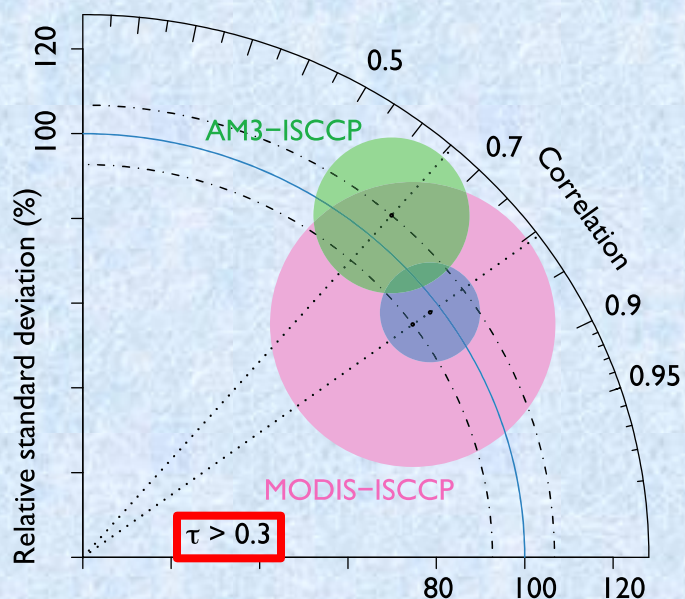
- Pincus et al. (2008)
CMIP3 cloud evaluation, low-order error measures, no obs proxy
- Williams and Webb (2009)
CFMIP1, evaluation using ISCCP proxy against observed canonical subsets
- Jiang et al (2012) and Li et al (2012)
CMIP5 LWP and IWP using A-train observations, no obs proxy
- Klein et al (*submitted*)
CFMIP1 + CFMIP2, evaluation of cloud-radiative impact using ISCCP proxy

Bottom line: Active area of research makes it difficult at this stage to identify metrics that meet the panel's criteria

Cloud property metrics – some recent progress

Reconciling Simulated and Observed Views of Clouds: MODIS, ISCCP, and the Limits of Instrument Simulators (Pincus et al., 2012)

- Biases between the two observational datasets (pink radius) is much larger than either model estimate
- Biases greatly reduced when optically thin clouds ($\tau < 1.3$) are excluded
- Improved understanding of ISCCP/MODIS differences can lead to more robust approaches to model evaluation



Some example challenges

- Sampling limits robustness with all longer time scale variability (not just ENSO)
- The MJO task force has recommended (and published) a few first order measures to meet the metrics panel's objectives, but the analysis is much more complex (i.e., not transparent) than our other metrics. Other, more complicated metrics can be anticipated.
- Large scale sub-surface ocean metrics are not well established for coupled models. While observations are the predominant limitation, there are other obstacles (e.g., ocean grids). For now the panel is focused on surface analysis, but CLIVAR has been making attempts to foster ocean metrics and the panel expects to engage increasingly with the WGOMD.

WGNE/WGCM Climate Model Metrics Panel

Titles

Text

pjj 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](#)

Current Priorities

- Strengthen the Panel's wiki so that it becomes recognized as a useful resource
- Make public a database/code of standard metrics results from all CMIP 3/5 models. This will enable modeling groups, if interested, to incorporate into their development process an ability to examine how their model compares to others
- Prepare manuscript synthesizing metrics panel results for CMIP 3 & 5
- Advance a repository for metrics/analysis codes
- Consider a workshop dedicated to performance metrics, 6-18 months after the March 2013 WGNE systematic errors workshop?