

# Community-Based Performance Metrics for Climate Models

Presented for

**Peter Gleckler**

by

Karl E. Taylor

Program for Climate Model Diagnosis and Intercomparison (PCMDI)  
Lawrence Livermore National Laboratory

Presented at the  
Climate and Earth System Modeling PI Meeting

Washington D.C.  
22 September 2011

# An update of the

## WGNE/WGCM\* Climate Model Metrics Panel

---

Members appointed based on relevant and diverse areas of expertise, 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**

\*WGNE – Working Group on Numerical Experimentation

WGCM – Working Group on Coupled Modeling

# 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 this effort **is**(and **is not**) about...

---

- Aims to provide a collection of metrics assessing a broad variety of model performance characteristics
  - ▶▶▶ Not to develop a single “figure of merit”
- Aims to quantify model consistency with observations
  - ▶▶▶ Not necessarily identify the causes of model errors
- Aims to establish a relatively small suite of standard metrics expected to be of interest for several development cycles of models
  - ▶▶▶ But not completely frozen



# Metrics panel terms of reference

---

- **Identify a limited set of basic climate model performance metrics**
  - based on comparison with observations
  - well established in literature
  - easy to calculate, reproduce and interpret
  - covering a diverse suite of climate characteristics
    - large- to global-scale mean climate and variability
    - atmosphere, oceans, land surface, and sea-ice

# Metrics panel terms of reference

---

- **Identify a limited set of basic climate model performance metrics**
  - based on comparison with observations
  - well established in literature
  - easy to calculate, reproduce and interpret
  - covering a diverse suite of climate characteristics
    - large- to global-scale mean climate and variability
    - atmosphere, oceans, land surface, and sea-ice
- **Justify and promote these basic metrics in an attempt to**
  - establish routine community benchmarks
  - facilitate further research of increasingly targeted metrics

# Metrics panel terms of reference

---

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

# Metrics panel terms of reference

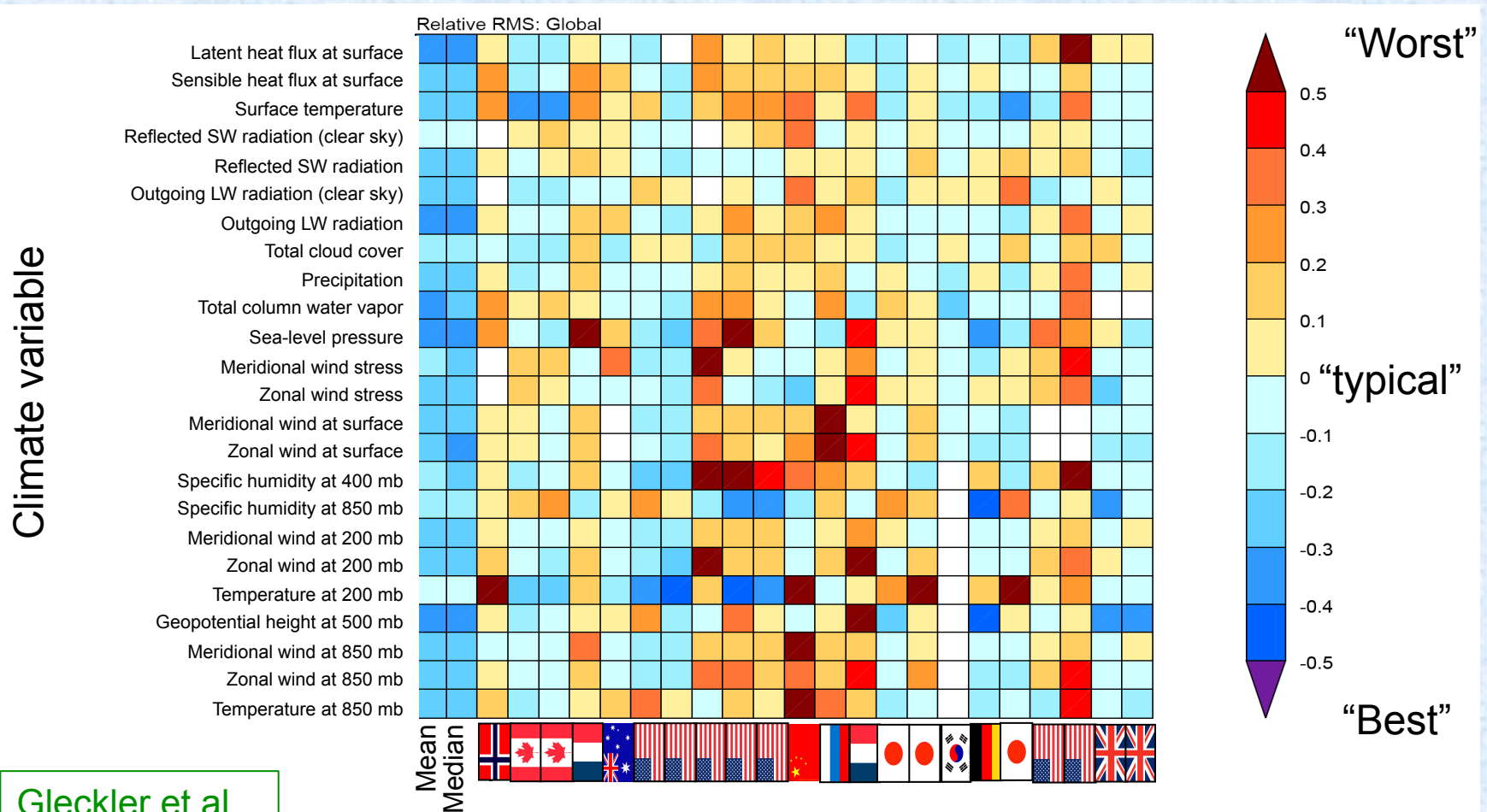
---

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



# A performance metric is a (statistical) measure of agreement between models and observations

## Relative RMSE in Climatological Annual Cycle (including spatial pattern)



Gleckler et al.,  
JGR, 2008

Models used in IPCC Fourth Assessment

# Current status: Focus is on a limited set of metrics to be periodically reviewed and augmented

---

## Climatological 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 (in progress) with other working groups:

- ENSO (CLIVAR Pacific Panel)
- MJO (YOTC Task force)
- Monsoons (CLIVAR AAMP)
- Carbon cycle in emission-driven ESMs (ILAMB)
- Coordination with other working groups is planned...

(e.g., GCSS/CFMIP and WGOMD)

# Next steps: Further engagement with research community

---

- Panel's efforts to be reviewed next month by both WCRP modeling working groups (WGNE and WGCM)
- Modeling groups to be given opportunity to comment on selected metrics
- Metrics panel wiki (to be made public before 2012):
  - Discussion of metrics, their limitations, panel goals
  - Posting metrics results for all CMIP3 and CMIP5 simulations
  - Traceability: all observations, codes and documentation made public
  - A resource with pointers to and discussion of relevant research





---

Some scratch slides....

# Targeting WCRP benchmarks experiments

---

- Panel focus is on WCRP/CMIP5 experiments where comparing to observations is most relevant
- Primary focus:
  - Historically forced AOGCMs, physical aspects of historical ESMs, and AMIP simulations
- But the panel will consider metrics for:
  - Historical ESM emission driven (e.g., metrics for [CO<sub>2</sub>])
  - Initial condition experiment: evaluation in "NWP mode" (AMIP<sup>T</sup>)