

# An update on the WGNE/WGCM Climate Model Metrics Panel

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

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

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

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

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

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

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

Karl Taylor (PCMDI) – WGCM/CMIP, **CMIP5**

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

# Origins...

- Effort initiated by the WGNE, with aspirations of developing a set of performance standards for climate models
- Formally connected with with WGCM in 2011
- Initial plans were to identify a short list of basic performance metrics, apply them to CMIP simulations, and make results widely available - not working directly with modeling groups

# What happened next...

Plenty of reaction, ranging from “this is a great idea” to “will be misused and is not helpful”

The metrics panel has been observant of shifting attitudes during an active period of related research including:

- Perspectives on the pros/cons of climate model metrics
- Use of multivariate performance indices, tracking performance changes
- Comparison of the error structure in MME and PPE ensembles
- Gauging model independence, exploration of model weighting
- Processed-oriented metrics
- Emergent constraints

# Priority of the panel re-oriented to ensure its efforts are useful to modeling groups

- Benefit to modeling groups submitting simulations to CMIP1-5 has always been limited by community research becoming less relevant to modelers once they move on to their next generation model.
- Modeling groups would learn more if they could assess the relative strengths and weakness of their model *during the development process*.
- The metrics panel, with support from PCMDI, will be providing modeling groups with a package that enables them to “quick-look” compare their newer model versions with all CMIP3 & CMIP5 models.

# The intent of the package

- Enable a limited “quick-look” evaluation *prior to* the substantial commitment of ESGF publishing of new CMIPX simulations
- First steps beyond difference maps provide useful quantitative information to an overall assessment of performance
- Enable modelers to more readily identify weaknesses and set their own “minimum bar”
- Highlighting weaknesses may help identify priorities for model development

# Benchmarking climate model performance

Some desired characteristics of performance metrics (**panel criteria**):

- A useful quantification of model error (with respect to observations)
- Well established in the literature and widely used
- Relatively easy to compute, reproduce and interpret
- Fairly robust results (**hence the emphasis on mean climate**)
- Well suited for repeated use

Some important distinctions and the **metrics panel emphasis**:

- **Routine** versus research (**the panel is not doing research**)
- **Broad** versus targeted
- **General purpose** versus application specific
- **Useful** versus most important

# The metrics panel package

## Nuts and bolts

- Includes code, documentation, carefully selected observations and a database of results for all CMIP3 & 5 simulations
- Built on a stripped down version of PCMDI's UV-CDAT which is based on python. ESMF regridding is built in so that data on ocean grids can be interpolated to regular lon x lat grid.
- Alpha version currently under platform portability testing – offering in 2013 to all modeling groups participating in CMIP

# Initial set of routine metrics for benchmarking

- Bias, pattern correlation, centered RMSE and mean abs error for global, tropical, and extra-tropical seasonal climatologies
- Package includes climatologies from multiple sources, primarily near-global satellite data and reanalysis
- Currently includes: Upper air temp and winds (200 and 850hPa), geopotential (500hPa), surface air temp, winds and humidity, TOA radiative fluxes and CRE, precipitation, precipitable water, SST, SSH, surface salinity, ...
- Additional possibilities under consideration in the near term: Global Monsoon Precipitation Index (CLIVAR AAMP), ENSO metrics (CLIVAR Pacific Panel), annual cycle of sea-ice extent, Mauna Loa [CO<sub>2</sub>], ...



# Next steps . . .

- Enabling all interested modeling groups to test package and provide feedback on possible improvements
- The intent is to gradually include an increasingly diverse set of metrics, taking into account input from modeling groups
- WGCM18: Assess the usefulness of the package and possible future directions

# “Won’t modelers just tune to these metrics?”

- The routine benchmarks in this package are *not* the most important measures of model performance. They have little relation to climate sensitivity.
- If the package has a sufficiently diverse set of metrics, tuning to them would be futile.
- The package does not include more advanced tests performed during systematic model evaluation. Explicit tuning would almost certainly result in disappointment.

## Additional benchmarking possibilities

- Longer term: Exploring pathways to develop more comprehensive community-based diagnostic (code) contributions (next talk – Veronika)
- Organized assessment of CMIP model dependence (raised by B. Stevens and R. Pincus). Would the metrics panel play a role?

## Why is the panel focused on routine metrics rather than attempting to identify the most important performance tests?

- A priority of the panel is to help create a communal environment for routine benchmarking of all climate models – for this we need to start somewhere
- Little consensus on what are the most important tests or even which questions to address\*
- “process-oriented” and “emergent constraints” seem attractive possibilities, but there are few examples that meet the panels criteria (in routine use, robust results, etc)

\* As a group, the panel is not engaged in doing science