Standardisation of model evaluation using modelevaluation.org

+

Weighting or sub-selection of ensemble members

Gab Abramowitz

gabriel@unsw.edu.au

UNSW Sydney / ARC Centre of Excellence for Climate Extremes

ARC CENTRE OF EXCELLENCE FOR CLIMATE EXTREMES

UNSW AUSTRALIA

Welcome to ModelEvaluation.org

ModelEvaluation.org is a web application for evaluating and benchmarking computational models. Browse menus or create an account to begin.

How does it work?

ModelEvaluation.org is supported by a range of funding and research coordination bodies, including:

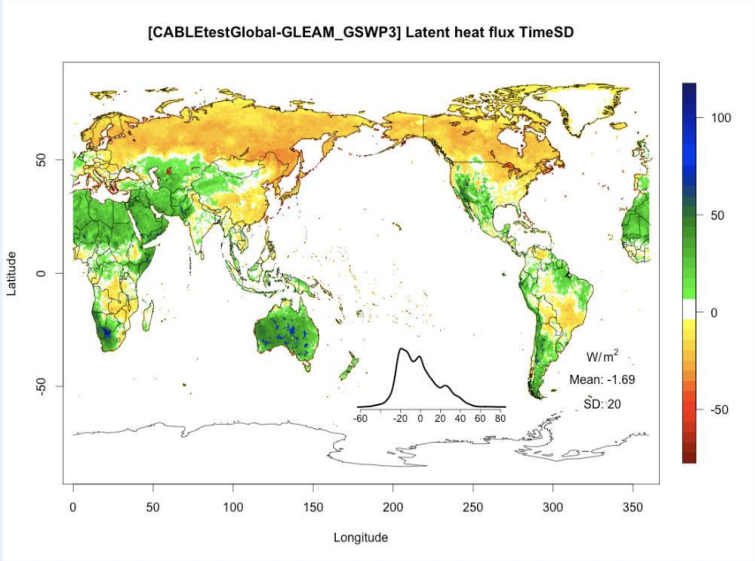
ARC CENTRE OF EXCELLENCE FOR CLIMATE EXTREMES

GEWEX

OzEWEX

NCI

TERN



Workflow in modelevaluation.org - simple description

Experiment in modelevaluation.org

Model outputs from other users

Analysis package

python

Experiment description

Driving data set(s)

Evaluation data set(s)

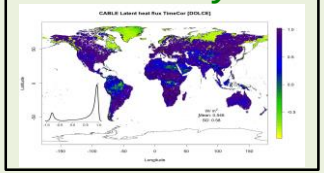
Analysis of model output produces plots, tables

View experiment protocol, download driving data

Run model in local computing environment

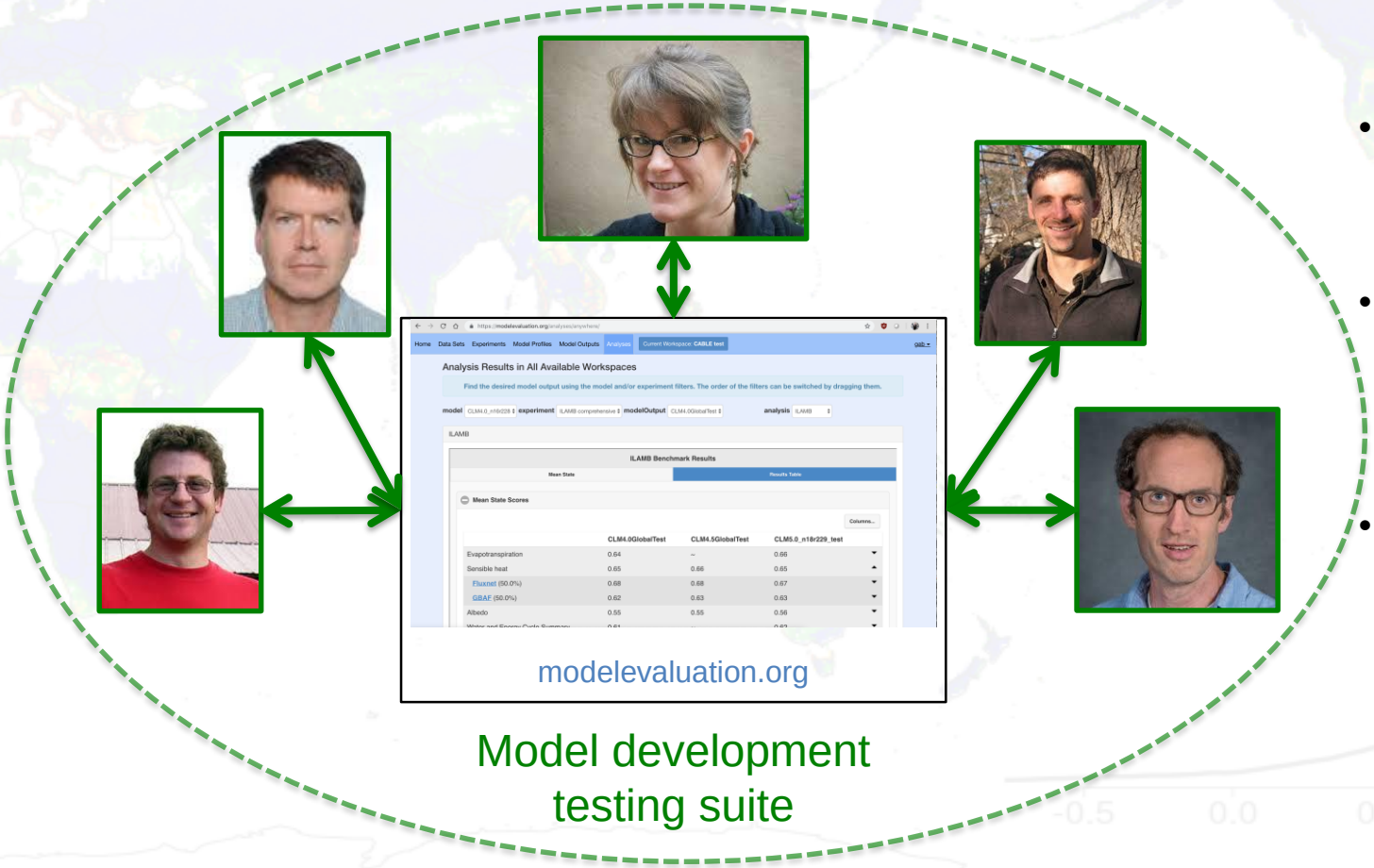
'Upload' model output

View analyses



Participant user

Participating in Experiments – for model development

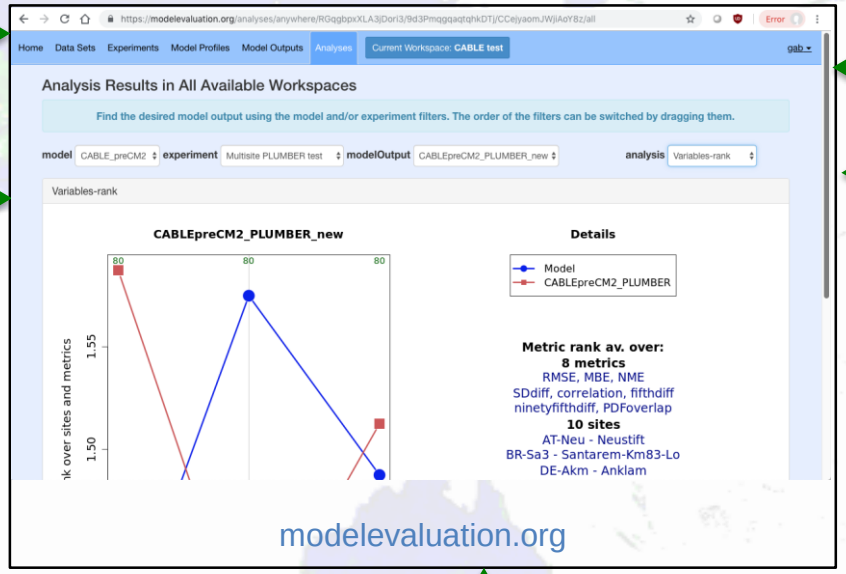


- Workspace only accessible to a team of users
- Import (or develop) experiments useful for model development
- Fast, shared evaluation results with meta data

Participating in Experiments – as part of a MIP

JULES
Joint UK Land Environment Simulator

ORCHIDEE
LAND SURFACE MODEL



modelevaluation.org

CLM
COMMUNITY LAND MODEL

NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

- Fast, shared reproducible evaluation results with meta data
- New contributions analysed automatically / repeated at any time
- Evaluation data can be hidden

CSIRO
ARC CENTRE OF EXCELLENCE FOR CLIMATE EXTREMES

Model Intercomparison environment

Standardisation of model evaluation using modelevaluation.org

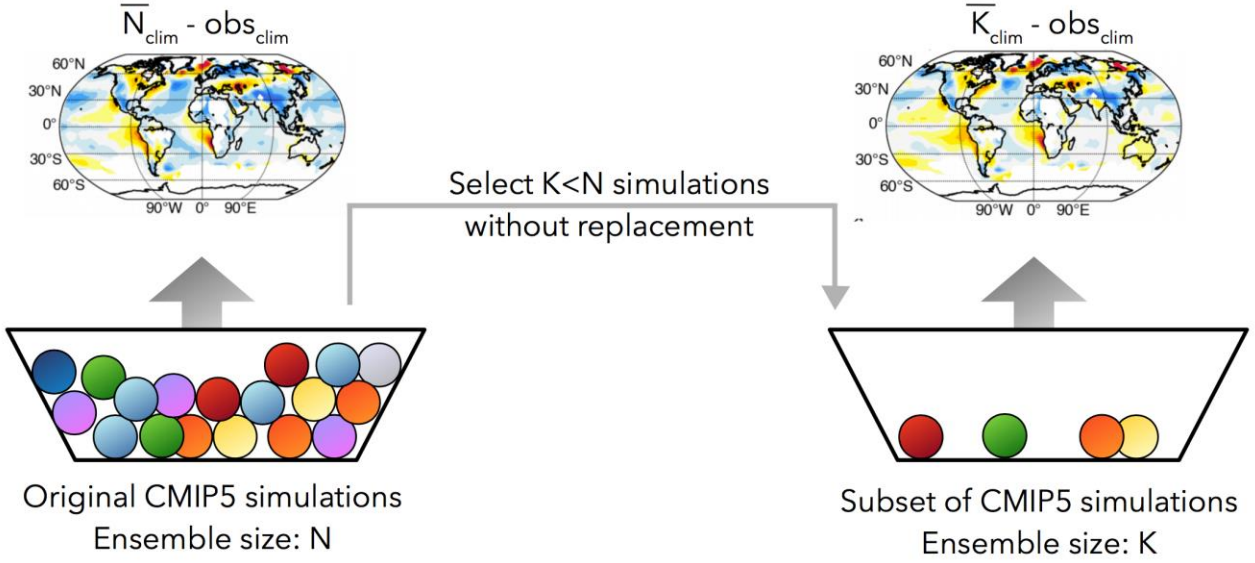
+

Weighting or sub-selection of ensemble members

Ensemble sub-sampling to account for dependence

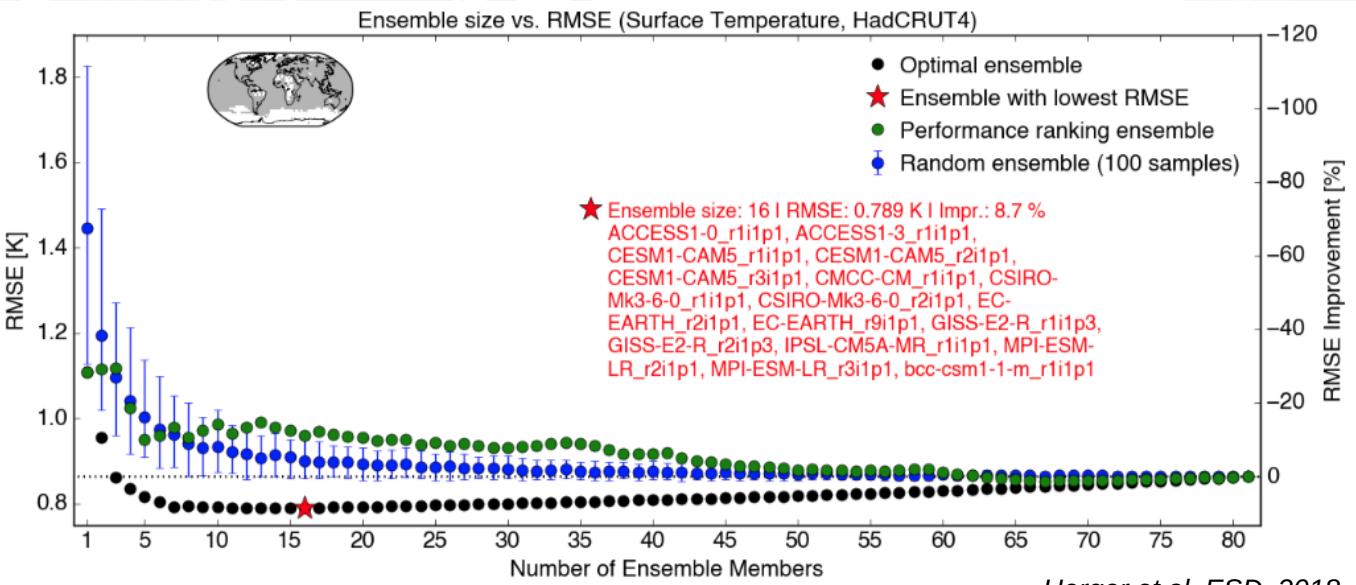
Three ensemble sub sampling approaches:

1. Random sampling of *K* simulations from a pool of *N* (100 times)
2. Choose the best performing *K* simulations (in terms of climatology)
3. Choose the *K* simulations whose mean has minimum RMSE in climatology against obs – account for dependence in regional biases



Ensemble size vs RMSE

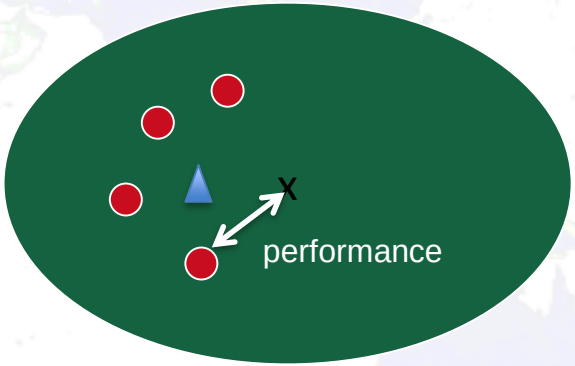
- Choosing the optimal ensemble is non-trivial – choosing K=40 (of N=81) means there are 212,392,290,424,395,860,814,420 possible ensembles

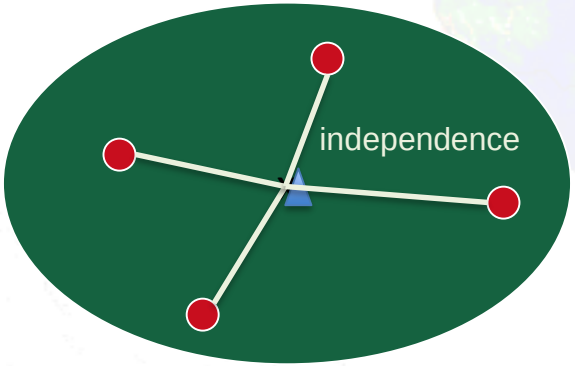


Herger et al, ESD, 2018

- Choosing the best performing models does not imply the best performing ensemble mean – dependence degrades the mean

Dependence is at least as important as performance





Framing dependence

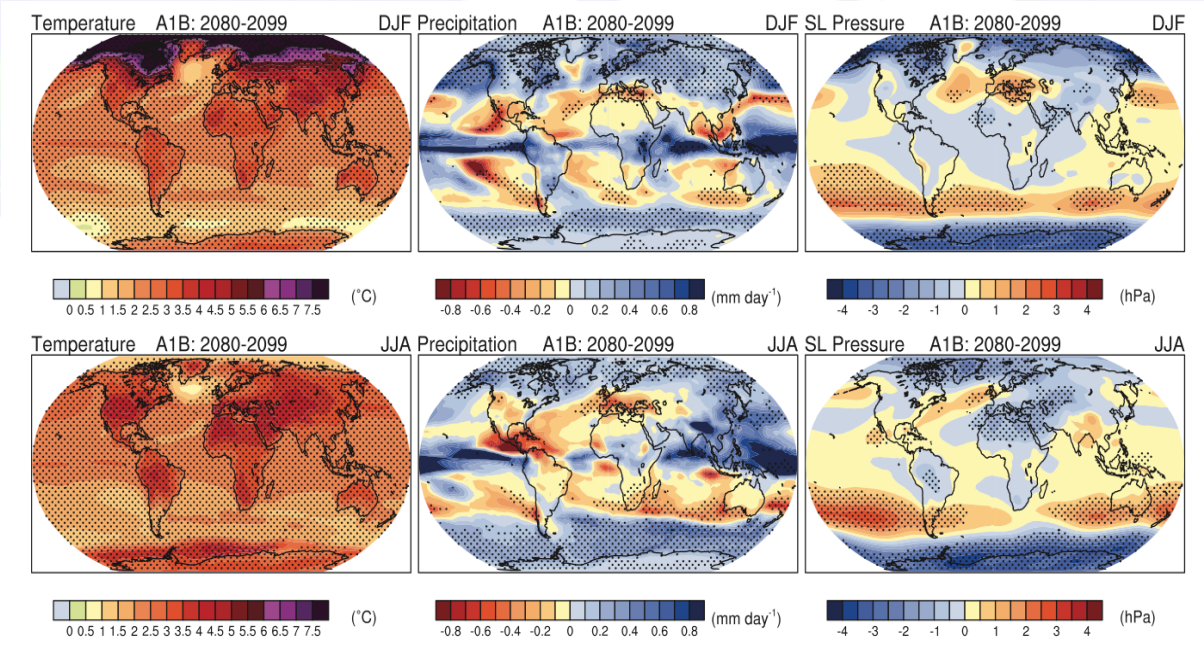


Figure 10.9. *Multi-model mean changes in surface air temperature (°C, left), precipitation (mm day^{-1}, middle) and sea level pressure (hPa, right) for boreal winter (DJF, top) and summer (JJA, bottom). Changes are given for the SRES A1B scenario, for the period 2080 to 2099 relative to 1980 to 1999. Stippling denotes areas where the magnitude of the multi-model ensemble mean exceeds the inter-model standard deviation. Results for individual models can be seen in the Supplementary Material for this chapter.*

Is agreement a sign of robustness?

Standardisation of model evaluation using modelevaluation.org

+

Weighting or sub-selection of ensemble members

Gab Abramowitz

gabriel@unsw.edu.au