Multi-objective optimization (Pareto-optimal) analysis

This process-oriented diagnostic (POD) package demonstrates a multi-objective approach to assess global climate model (GCM) performance evaluated by multiple objective functions considering trade-offs in model performance for these different objective functions. Two examples are illustrated in this POD by quantifying model accuracy in simulating boreal winter mean precipitation for the California and the South American Monsoon regions, respectively, and their associated large-scale circulation and tropical SST pattern in a targeting GCM (i.e., GFDL-CM4) along with multiple CMIP6 model simulations. A so-called Pareto-optimal analysis (Langenbrunner and Neelin 2017a; 2017b) approach is applied to identify the set of optimal model sub-ensemble sets across measures against three different objective functions. This approach can be used to determine the best model subset(s) with top performances for desired measures, which can be useful for various studies, for example, on future climate projection or regional dynamical downscaling using large-scale forcing from global model simulations.

Contact information:

- Xianan Jiang (UCLA; xianan@ucla.edu)
- PI: J. David Neelin (UCLA; neelin@atmos.ucla.edu)
- Current developer: Xianan Jiang
- Contributor: Baird Langenbrunner

POD structure:

This POD package consists of the following functionalities:
(1) The POD driver script ("Pareto_optimal.py").
(2) A script to set analysis parameters, e.g., the targeting model name, analysis domains for evaluation of multiple objective functions ("pareto_calculation_parameters.py").
(3) A script to read pre-processed observational and CMIP6 model output ("input_data.py").
(4) A script to process winter mean patterns based on the targeting model ("process_model_data.py") with the core codes based on NCAR NCL ("m_win_mean.ncl").
(5) The main POD script to calculate 2D and 3D Pareto-optimal fronts ("calculate_data_for_figure2_and_figure3_k1to3.py").
(6) A script to plot winter mean biases based on CMIP6 multi-models and the targeting model with comparison to their observed counterparts ("figure1_script_present_climate.py").
(7) A script to plot 2D and 3D Pareto-optimal analysis ("figure2_script_k3.py").
(8) A script to plot spatial patterns of various objective functions corresponding to a minimum RMSE over the analysis region for each objective function ("figure3_pattern_pareto_front_minimum_rmse_k3")

All scripts of this package can be found under "<CODE_ROOT>/diagnostics/Pareto_optimal/", and pre-processed observational and CMIP6 model data under "<OBS_DATA_ROOT>/Pareto_optimal/"

Required Programming language and libraries:

The package was written using python 3.9.1. The required packages are ncl, netCDF, numpy, scipy, itertools, cartopy, matplotlib.

Required model output variables:

The following monthly mean model fields are required:

3D variables (lat-lon-time):

- Precipitation rate (PR; units: kg m^{-2} s^{-1})
- Sea surface temperature (TOS: units: degC)

4D variable (time-height-lat-lon; only 200hPa and 850hPa are used in this POD):

- Zonal wind (UA; units: m s^{-1})

Two examples of analyses provides in the POD:

This POD provides two examples of the multi-objective optimization analysis by focusing on the winter mean precipitation over the California and South American Monsoon regions, respectively, and their associated large-scale circulation/SST patterns, with the GFDL-CM4 from the CMIP6 archive as the targeting analysis model.

These two pre-defined analysis examples can be selected with the parameter "exp_name" in the script "pareto_calculation_parameters.py".

exp_name='CA' - to run diagnostics focusing on CA winter precipitation
exp_name='SAM' - to run diagnostics focusing on the South American Monsoon

References:

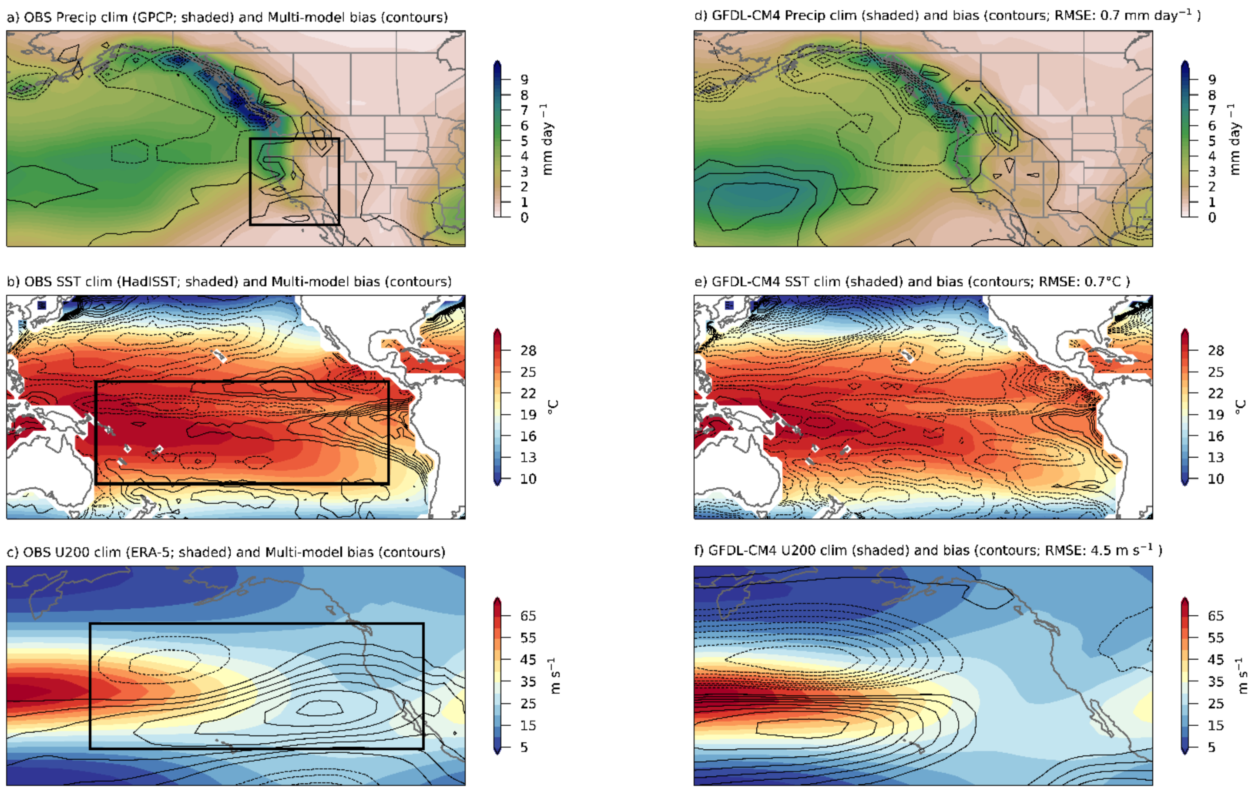
Langenbrunner, B. and J. D. Neelin, 2017a: Pareto-Optimal Estimates of California Precipitation Change. *Geophys. Res. Lett.*, **44**, https://doi.org/10.1002/2017GL075226, 12,436-12,446.

Langenbrunner, B. and J. D. Neelin, 2017b: Multiobjective constraints for climate model parameter choices: Pragmatic Pareto fronts in CESM1. *Journal of Advances in Modeling Earth Systems*, **9**, https://doi.org/10.1002/2017MS000942, 2008-2026.

Descriptions of the POD output:

This diagnostic POD will produce three figures, which are described in the following based on the experiment case with multi-objective analysis associated with winter precipitation over the California region.

Fig. 1 Precip, U200, SST biases in CMIP6 multi-model and the targeting GCM (i.e., GFDL-CM4) simulations and compared to the observed counterparts.



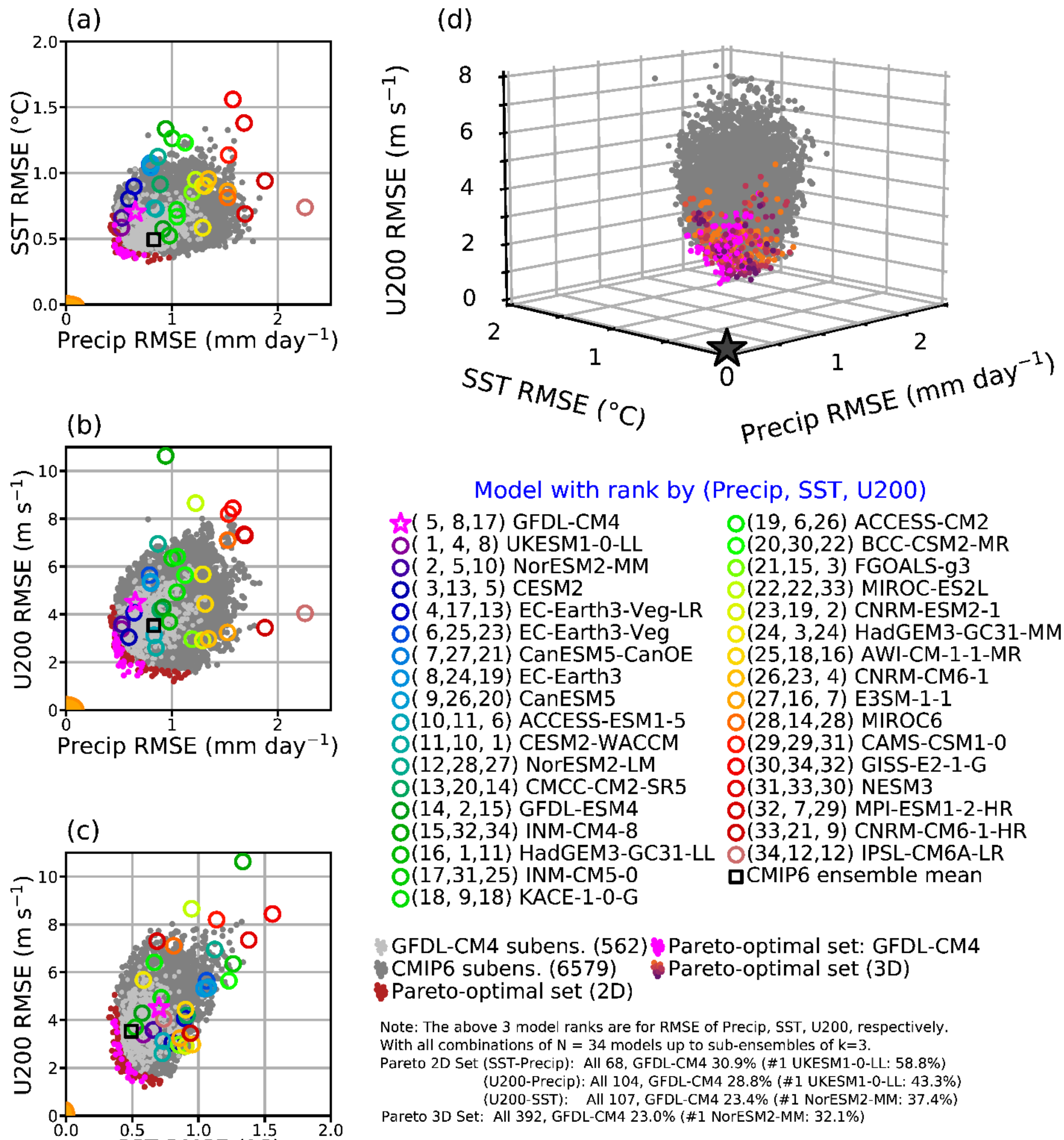
All fields are for boreal winter or austral summer seasonal mean (Dec-Jan).
Contours (zero-lines omitted) with intervals of 0.5 mm day^{-1} for Precip, 0.2°C for SST, and 1.0 m s^{-1} for U200, respectively.
Multi-model simulations are from the CMIP6 project.
Rectanglar boxes represent analysis regions for the three values.

Left column: Shaded patterns represent the *observed* climatological winter mean (Dec-Feb) a) precipitation (GPCP), b) SST (HadISST), and c) U200 (ERA-5) for the period of 1981-2010. Contours are for model biases in simulating these three climatological fields based on multi-model mean of 33 CMIP6 GCMs (see Fig. 2 for details of CMIP6 GCMs) and the targeting GCM.

Right column: Similar as for the left column, but the shading and contours are for winter climatological mean and biases of the targeting GCM.

The three boxes on the left column denote the analysis regions for the three objective functions for this experiment case, including precipitation over the California region, SST over tropical Pacific Ocean, and U200 over extra-tropical Pacific associated with subtropical jet stream that strongly influences wave activity and moisture advection that have been linked to CA precipitation. Refer to Langenbrunner and Neelin (2017a) for details.

Fig. 2 2D and 3D Pareto-optimal analyses for Precip, SST, and U200 focusing on CA.

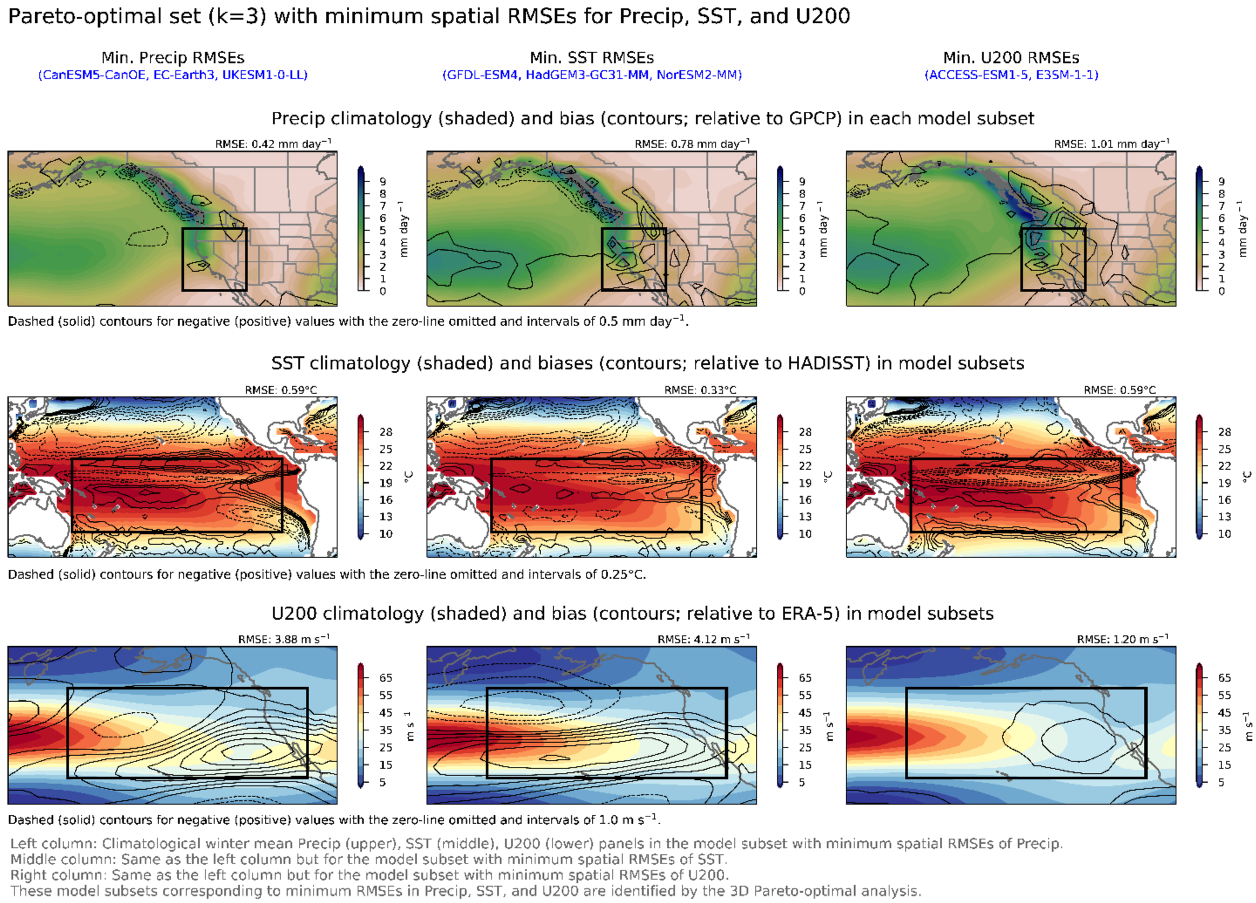


Left column: 2D Pareto-optimal analyses (a-c). The spatial root-mean-square error (RMSE) of fields from each CMIP6 model and the target model (i.e., GFDL-CM4) are first calculated over their corresponding domain (box regions in Fig. 1). The RMSEs based on each individual model simulations are plotted on 2-D planes consisting of each pair of the three variables (SST, Precip, U200; i.e., the objective functions) with color circles for CMIP6 models and the magenta "star" sign for the target GCM. RMSEs for ensemble mean fields based on all CMIP6 GCMs and the targeting model (GFDL CM4) are represented by the black square. These thus provide an evaluation of how the winter mean patterns of these three objective functions are simulated in individual GCMs along with all model ensemble mean, with the ranking for each objective function (total 3 ranking scores) shown in parentheses in the legend and sorted by the rank of mean precipitation RMSEs.

Then RMSEs for the three objective functions are calculated based on ensemble mean from all possible model combinations that consists of up-to unique “k” models from the total 34 GCMs (k=3 for this analysis), and are plotted by dark gray dots in the figure. For a total model number N=34, there are 6579 possible combinations of model sub-ensembles. Among these total combinations, model subsets (total 562) in which the targeting model (GFDL-CM4) is included are labelled by the light gray dots. Members of model subsets with minimum RMSEs on each 2-D objective function spaces can then be identified as optimal subsets following a “Pareto-optimal” analysis approach described in Langenbrunner and Neelin (2017a; 2017b), i.e., the Pareto fronts (shown as red dots), which accounts trade-offs between the two objective functions in a)-c). A “shorter” Pareto front (small extent in the objective function space) implies that constraints are in accord, whereas a “longer” front indicates a broader set of trade-offs that require careful examination. The pareto front members that the targeting model (GFDF-CM4) is included are labeled by mangeta dots.

Right column: A 3D Pareto-optimal analysis for Pareto-optimal solutions on the three objective function spaces using a similar approach for 2D analyses.

Fig. 3 Spatial patterns of variables in the model subset with minimum RMSEs evaluated by each of the three objective functions



The third figure shows spatial distribution of climatological winter mean patterns (shading) and biases related to the observations (contours) of each variable, i.e., PRECIP (top row), SST (middle row), U200 (bottom row), based on sub-ensemble mean of the model subset with the minimum RMSEs in CA precipitation (left column), tropical SST (middle column), and U200 (right column) over their corresponding evaluation regions denoted by the black boxes. The corresponding model subsets are identified by the 3D Pareto-optimal analysis in Fig. 2, with individual models in the subset listed in blue fonts.