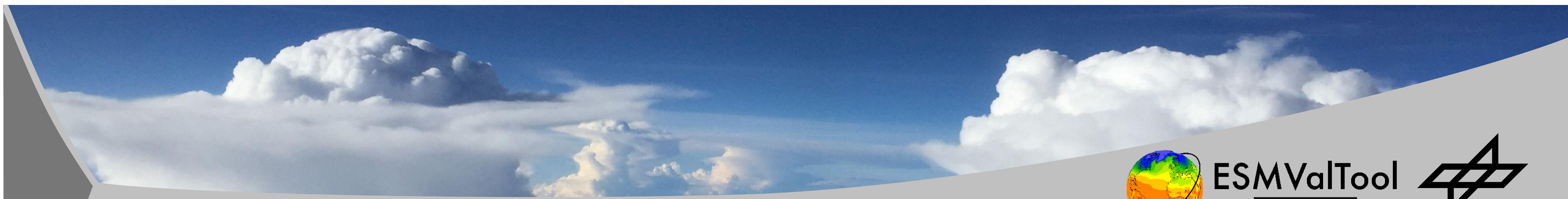


Monitoring and Benchmarking of Earth System Model Simulations with ESMValTool

Manuel Schlund[1], Axel Lauer[1], Lisa Bock[1], Birgit Hassler[1], Patrick Jöckel[1], Lukas Ruhe[1,2]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

EGU24-8947

Sharing is encouraged

Paper: **Lauer et al., GMD (in prep.)**

Abstract

Earth system models (ESMs) are important tools to improve our understanding of present-day climate and to project climate change under different plausible future scenarios. For this, ESMs are continuously improved and extended resulting in more complex models. Particularly during the model development phase, it is important to continuously monitor how well the historical climate is reproduced and to systematically analyze, evaluate, understand, and document possible shortcomings. Putting model biases relative to observations into the context of deviations shown by other state-of-the-art models greatly helps to assess which biases need to be addressed with higher priority. Here, we introduce a new capability of the Earth System Model Evaluation Tool (ESMValTool) to monitor running or benchmark existing simulations with observations in the context of results from the Coupled Model Intercomparison Project (CMIP).

About ESMValTool

- Open-source community-developed diagnostics and performance metrics tool for the evaluation and analysis of ESMs
- Fast and easy routine evaluation of ESMs including provenance records for all results (traceability and reproducibility)
- Multitude of generalized preprocessor functions (e.g., regridding, statistics, multi-model analysis, etc.)
- Extensive documentation (user guide, peer-reviewed papers, tutorial)
- Well-established analysis based on peer-reviewed literature
- Listed by CMIP7 Model Benchmarking Task Team as one of the model benchmarking and evaluation tools

Righi et al. (2020): **Technical overview**
Eyring et al. (2020): **Large-scale diagnostics**
Lauer et al. (2020): **Emergent constraints and future projections**
Weigel et al. (2021): **Extreme events, regional and impact evaluation**
Schlund et al. (2023): **Evaluation of native ESM output**

Scientific Documentation

Benchmarking metrics

In addition to assessing **absolute values** and **biases** of variables, the following **metrics** are available (calculated over arbitrary dimensions relative to a reference data set, usually observational products):

Root mean square error (RMSE)

- Calculation: $RMSE = \sqrt{\sum_{i=1}^{N} w_i (X_i - R_i)^2}$
- X: data, R: reference, N: number of elements over which RMSE is calculated, w: normalized weights ($\sum_{i=1}^{N} w_i = 1$)
- Range: $[0, \infty]$, smaller is better
- Measures the total deviation of data vs. reference as sum of individual deviations in each point

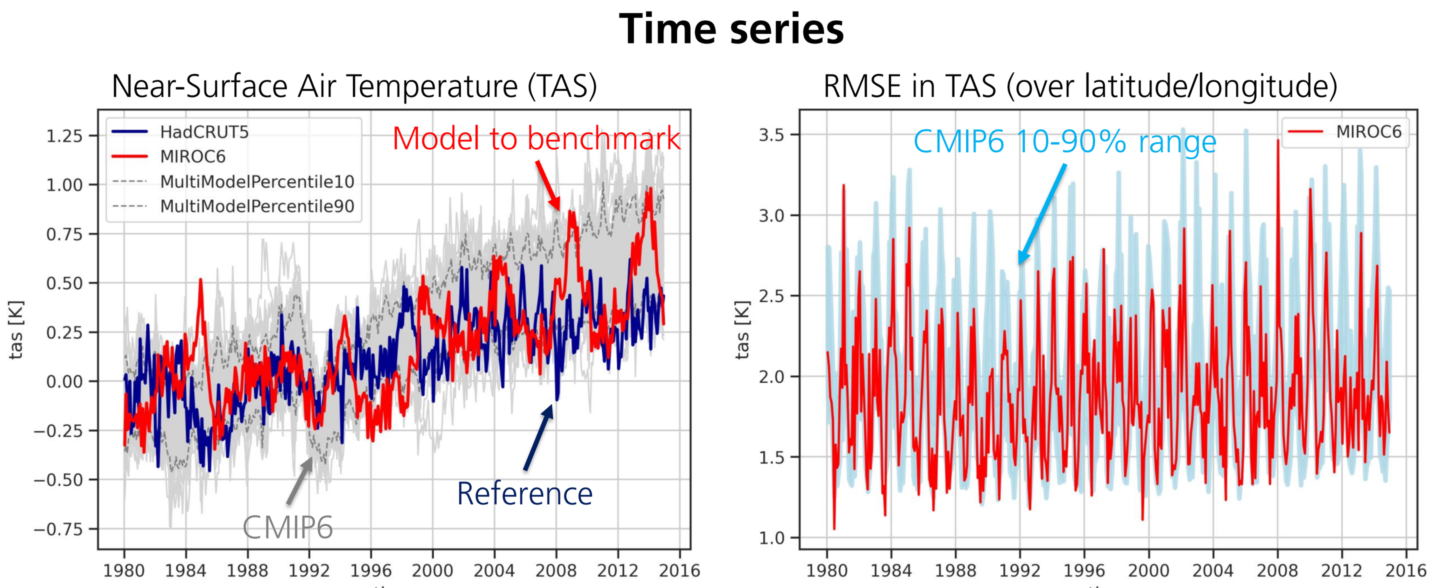
Pearson correlation coefficient (r)

- Calculation: $r = \frac{\sum_{i=1}^{N} w_i (X_i - \bar{X})(R_i - \bar{R})}{\sqrt{\sum_{i=1}^{N} w_i (X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N} w_i (R_i - \bar{R})^2}}$
- X: data, R: reference, N: number of elements over which r is calculated, w: normalized weights ($\sum_{i=1}^{N} w_i = 1$)
- Range: $[-1, 1]$, higher is better
- Measures the linear correlation between data and reference (can be positive or negative)

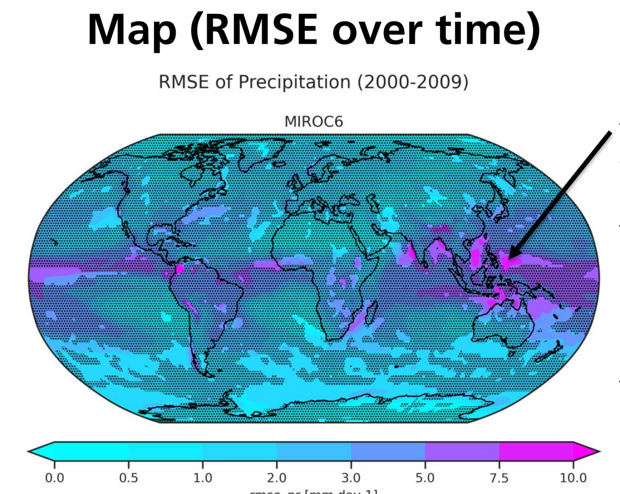
Earth mover's distance (EMD)

- Calculation: $EMD = \min_{\gamma \in \mathbb{R}_+^{n \times n}} \sum_{i,j}^{n} \gamma_{ij} |x_i - r_j|$
- γ: joint probability distribution of the binned data x and reference r that minimizes the transportation cost ("optimal transport matrix"), n: number of bins
- Range: $[0, \infty]$, smaller is better
- Measures similarity between two probability distributions (data and reference): minimum amount of work needed to transform one distribution (= "pile of earth") into another

Time series

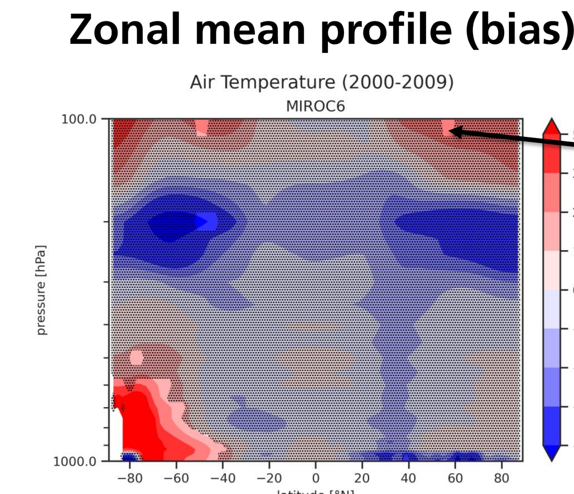


Map (RMSE over time)



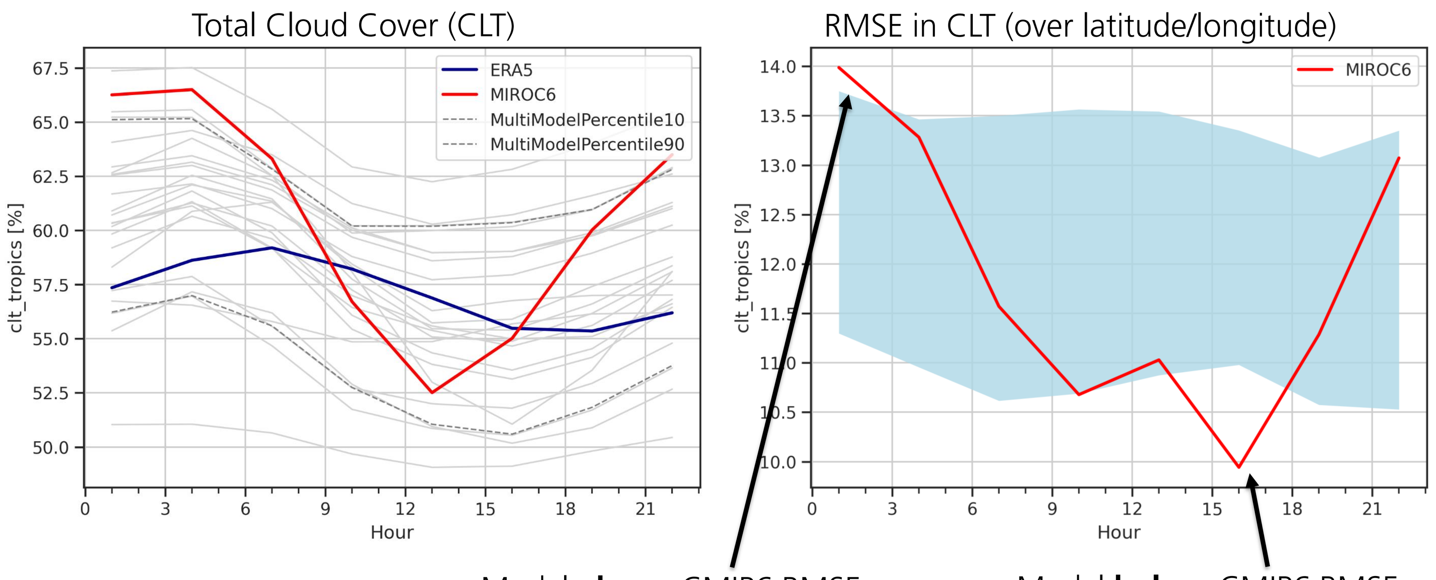
Areas without stippling: Model RMSE is **larger** than 90[th] percentile of CMIP6 RMSE (i.e., noticeably **worse** performance)

Zonal mean profile (bias)



Areas without stippling: Absolute model bias is **larger** than maximum of absolute 10[th] and 90[th] percentile of CMIP6 bias (i.e., noticeably **worse** performance)

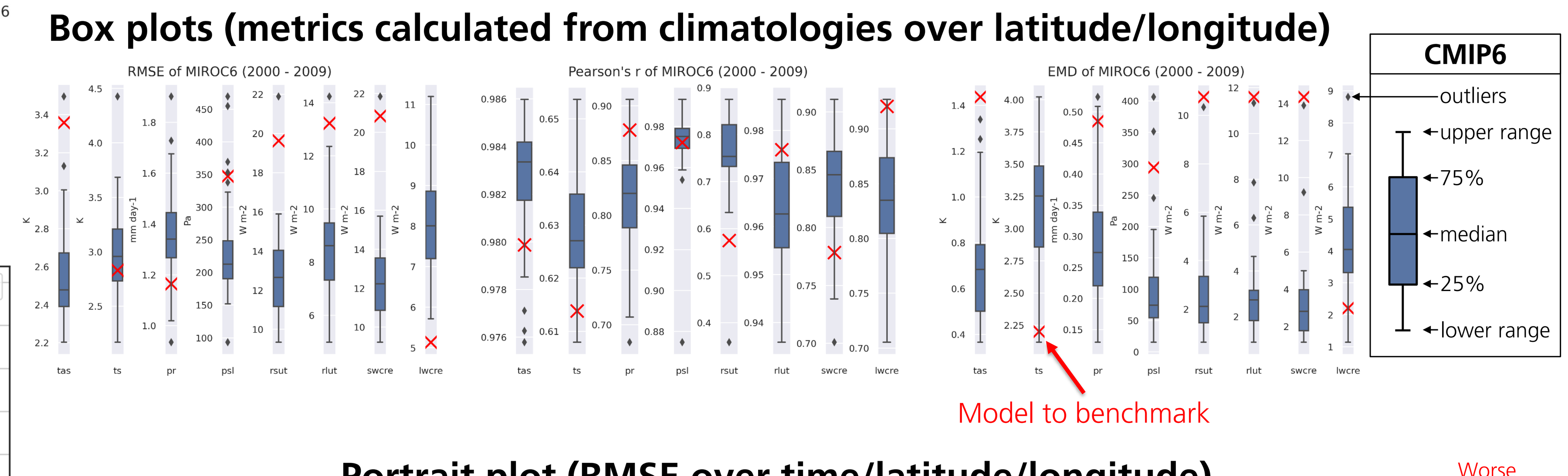
Diurnal cycle



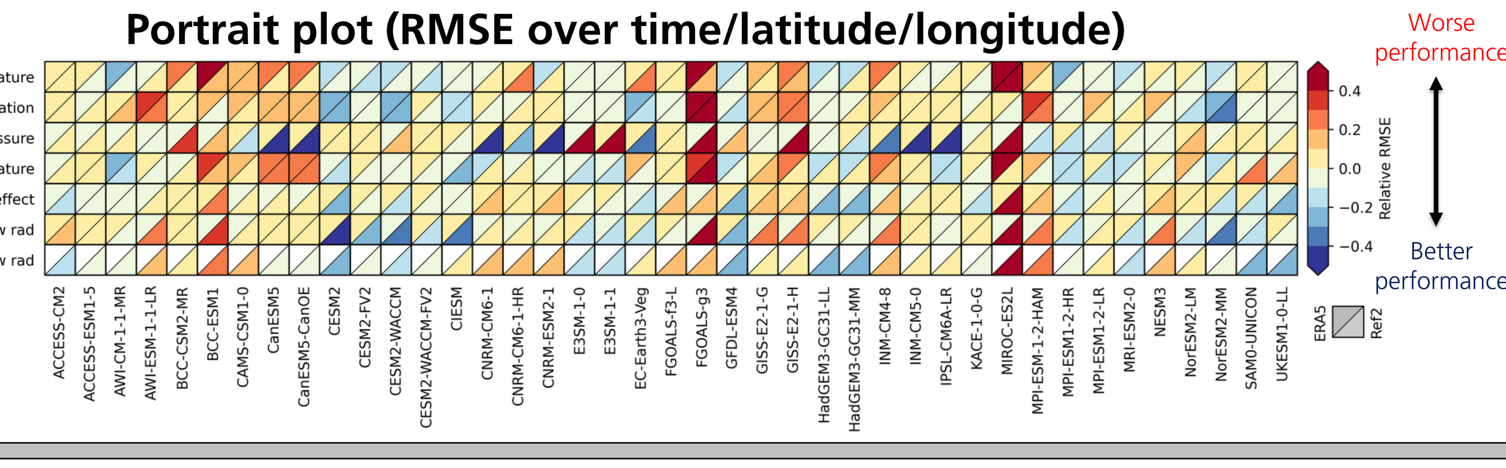
Model **above** CMIP6 RMSE range: noticeably **worse** performance

Model **below** CMIP6 RMSE range: noticeably **better** performance

Box plots (metrics calculated from climatologies over latitude/longitude)



Portrait plot (RMSE over time/latitude/longitude)



Further Resources

- **Website**: esmvaltool.org
- **GitHub**: github.com/ESMValGroup
- **Documentation**: docs.esmvaltool.org/en/latest
- **Tutorial**: tutorial.esmvaltool.org
- **X**: twitter.com/ESMValTool

Contact: manuel.schlund@dlr.de