

Using machine learning to distinguish km-scale climate models and observations on a regional scale

Maximilian Meindl¹, Aiko Voigt¹, and Lukas Brunner²

¹University of Vienna, Department of Meteorology and Geophysics, Vienna, Austria (maximilian.meindl@univie.ac.at)

²University of Hamburg, Research Unit for Sustainability and Climate Risks, Hamburg, Germany

The use of machine learning (ML) for climate science has attracted considerable attention within the last few years. A number of recent studies have used ML to extract information from global climate data (e.g. regional downscaling), predict future states of the climate system and evaluate models against observations. In particular, *Brunner and Sippel (2023)* showed that low-resolution global climate models and observations can reliably be distinguished based on the global distribution of daily temperature, even after removing the mean model bias. ML is thus able to isolate fundamental differences between models and observations even in the presence of substantial internal variability. This raises the questions of whether ML can also distinguish between model and observational data on a regional scale, whether ML is as successful for km-scale models as for coarse-resolution models, and whether more complex bias correction methods reduce the success of ML.

To answer these questions, we use daily temperature fields over Austria, a topographically very complex domain. As training data, we use 200 different, randomly drawn days from each of the 13 ÖKS15 bias-corrected EURO-CORDEX models with an output resolution of 1km, resulting in 2600 samples labeled “model” which are matched by the same number of random days labeled “observation” from the SPARTACUS observation dataset. We use the binary classification approach to distinguish between the two classes of models versus observations. A logistic regression classifier is trained to determine the probability that a daily temperature field belongs to one of the two classes. In order to evaluate the ML algorithm subsequently, all days from the out-of-sample 10-year period 2005-2014 are used as test data.

The ML algorithm succeeds in correctly identifying the overwhelming majority of the test data for the setup used, resulting in an accuracy of 99%. The results remain consistent even when a different sample of 2x2600 random training days is used. In contrast to more complex classifiers, such as a convolutional neural network (CNN), the learned coefficients from the logistic regression allow insights into the spatial patterns that are crucial for distinguishing between models and observations. While the performance of climate models is typically evaluated on climatological timescales, our results highlight that such classifiers can be used to identify patterns of structural model biases. Our method hence offers a computationally efficient approach for model evaluation, especially when handling km-scale climate model data on a regional domain.

References:

Brunner L. and Sippel S. (2023): Identifying climate models based on their daily output using machine learning, Environmental Data Science, <https://doi.org/10.1017/eds.2023.23>