universität wien
Department of Meteorology and Geophysics

Separation of climate models and observations based on daily output using machine learning

Lukas Brunner[1], Sebastian Sippel[2], and Aiko Voigt[1]

1) Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria

2) Institute for Meteorology, University of Leipzig, Leipzig, Germany

1. Motivation and research question

- Climate model performance is typically evaluated on climatological timescales to minimise the effects of internal variability.
- However, in some cases multi-decadal data are not (yet) available (decadal predictions, km-scale simulations, e.g. from NextGEMS).
- Machine and statistical learning methods are opening up new ways to identify patterns even in the presence of high internal variability.

Can climate models and observations be separated based only on their temperature output from a single day?

2. Data

- 43 CMIP6 models using historical forcing
- 4 observational datasets (ERA5, MERRA2, 20CR, DOISST)
- 1 km-scale model (ICON-Sapphire; Hohenegger et al. 2023)
- Daily global temperatures at 2.5°x2.5° in the period 1982 to 2014
- Land grid cells masked and daily global mean removed

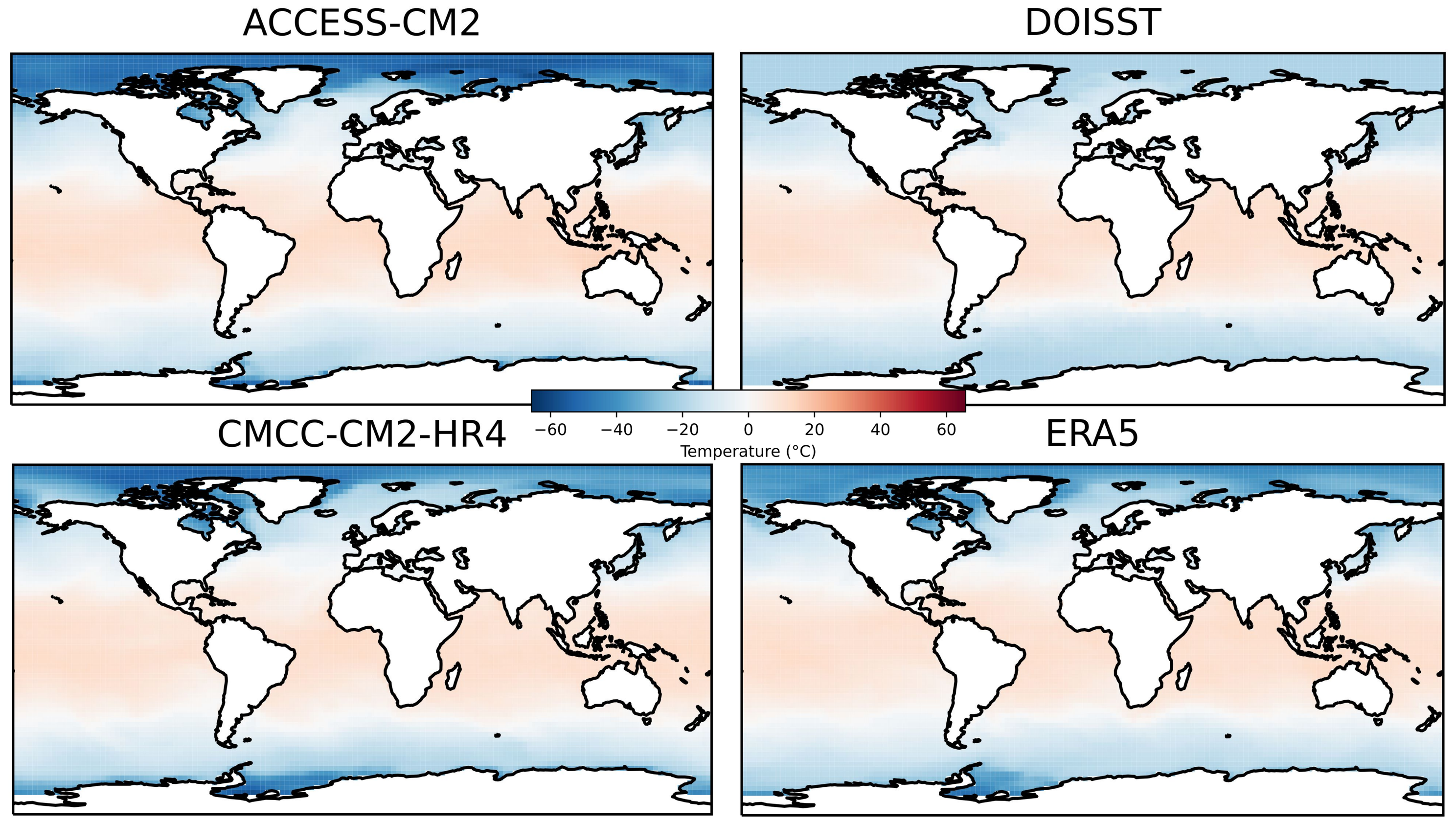

***Figure**: Example temperature maps from four datasets on March 21st 2010. These maps are used as input for the classifiers.*

3. Logistic regression and out-of-sample framework

The logistic regression classifier estimates coefficients to minimise

$$\min_w \left\{ -C \sum_{n=1}^{N} \left[y_n \log(\hat{p}_n) + (1 - y_n) \log(1 - \hat{p}_n) \right] + \sum_{m=0}^{M} w_m^2 \right\}$$

The predicted probability is defined as $\hat{p}_n = \dfrac{1}{1 + \exp\left(-(w_0 + \sum_{m=1}^{M} w_m x_{n,m})\right)}$

A separate classifier is trained for each test dataset, excluding all samples from that dataset from training (dataset out-of-sample).

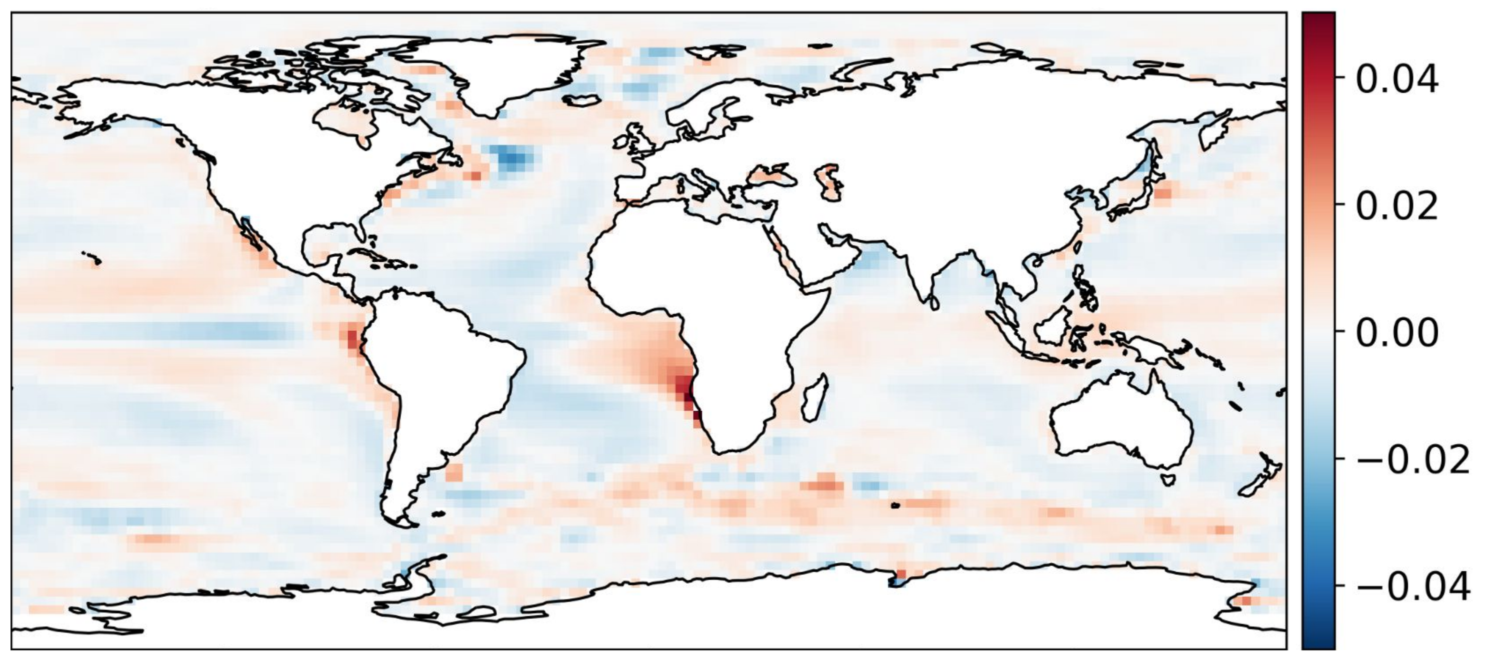

Figure: Spatial representation of the regression coefficients used to separate daily temperature maps from models and observations.

Climate models can be identified from observations based on temperature maps from individual days.

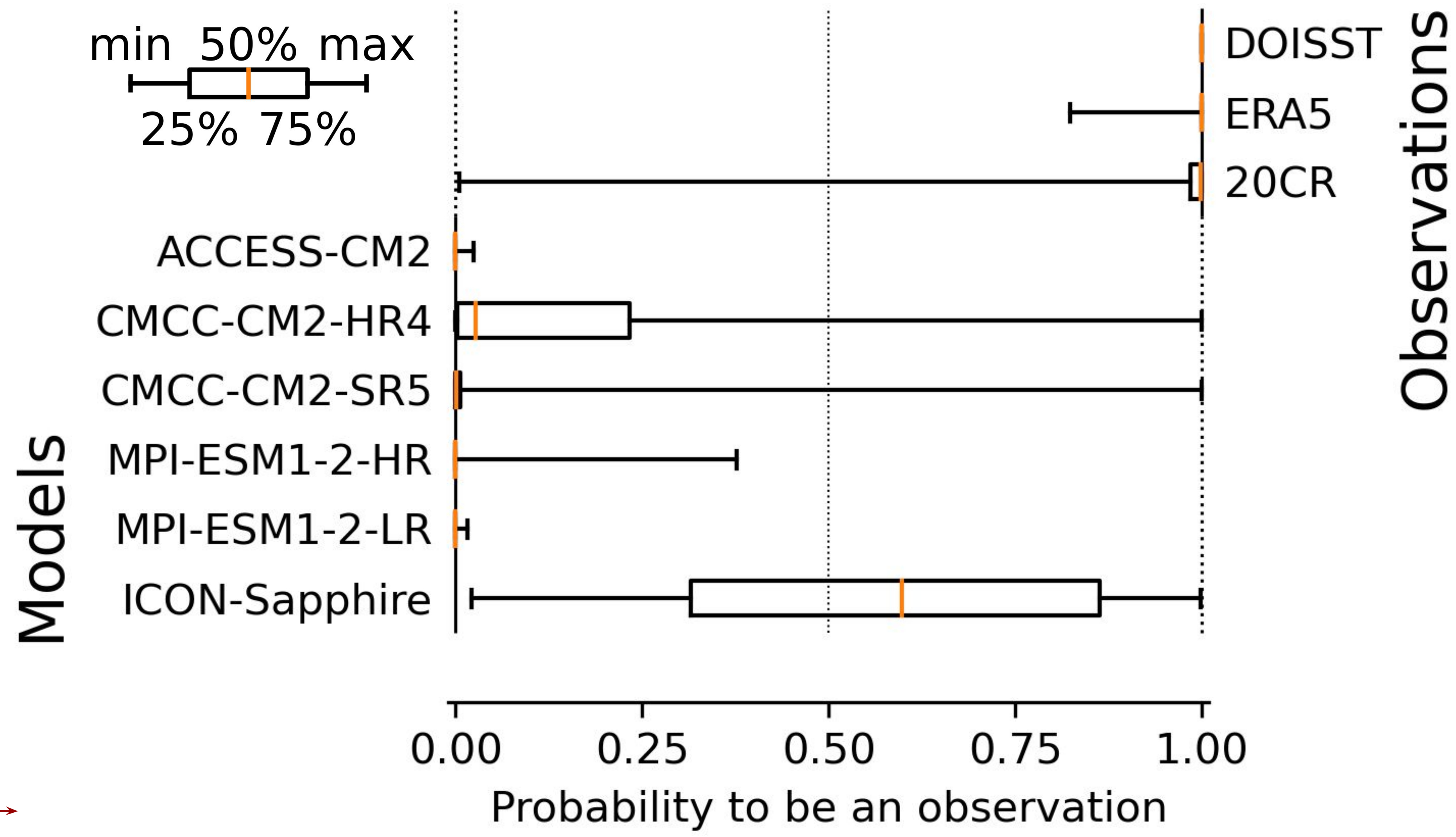

***Figure**: Distribution of predicted probabilities for test days from selected out-of-sample datasets in the period 2005-2014 using logistic regression classifiers.*

4. Main results logistic regression

- Most test samples are classified correctly via logistic regression
- Higher resolved models tend to be misclassified more often
- ICON (resolution 5km) cannot be clearly assigned to either class

5. Additional results

- A more complex convolutional neural network (CNN) is also able to correctly classify 75% of samples from ICON-Sapphire.
- CNN classifiers are able to correctly classify samples even after bias correction using the mean seasonal cycle (not shown).

See pre-print for a full description of the methods and results.

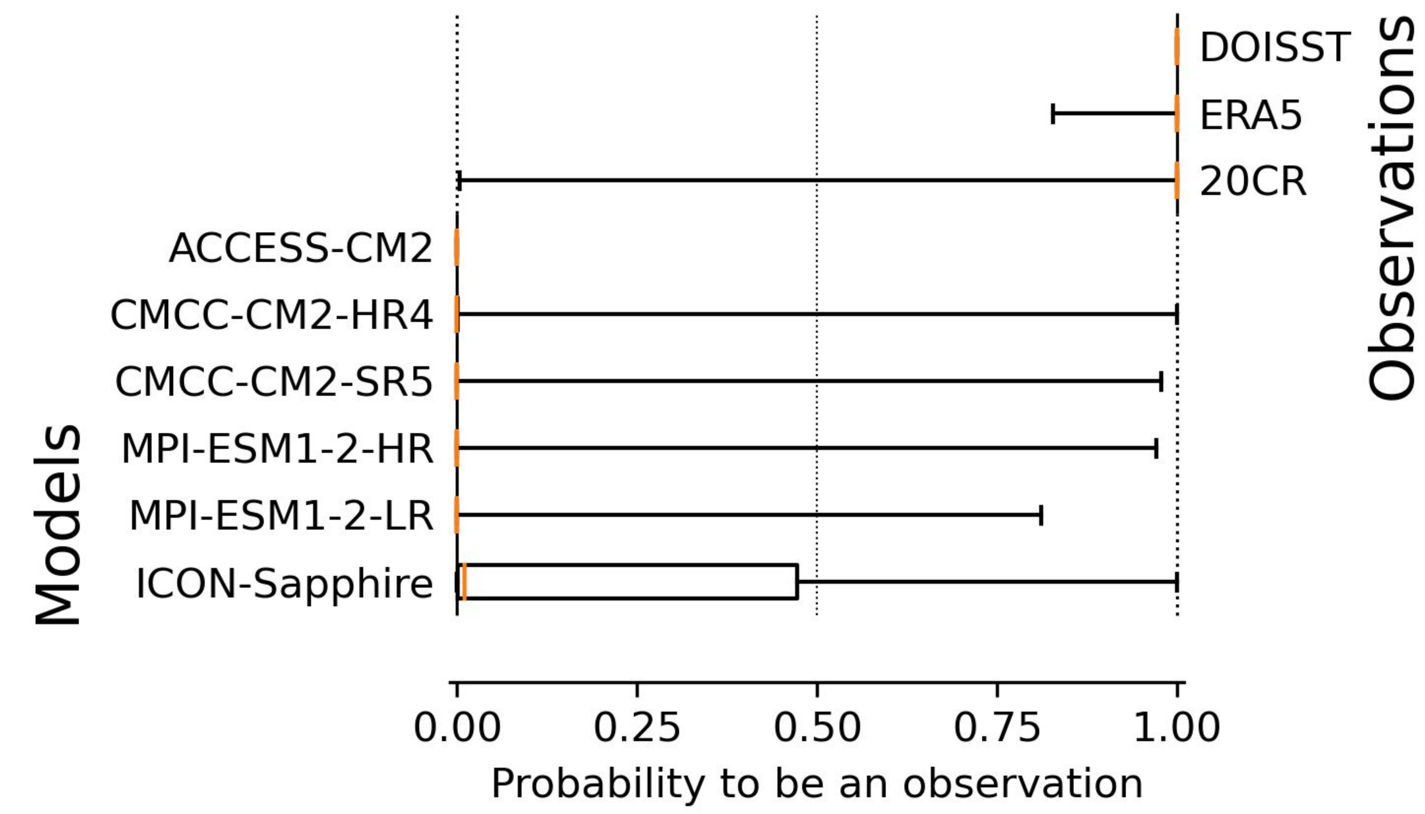

***Figure**: Distribution of predicted probabilities for test days from selected out-of-sample datasets in the period 2005-2014 using CNNs.*

5. Conclusions and outlook

- **Individual days are enough to robustly identify temperature maps as coming from climate models or from observations.**
- Output from the latest, km-scale models can't be identified as belonging to either category by a logistic regression classifier.
- More complex classifiers based on convolutional neural networks are able to correctly identify also km-scale model output.
- Future applications of this framework will contribute to the model evaluation toolbox and, for example,
  - provide assessments of model performance from daily data,
  - pinpoint regions important for model-observation separation,
  - separate other dimensions such as model generations, and
  - build adversarial networks as novel ways for bias correction.

6. References

Brunner and Sippel (under review): Identifying climate models based on their daily output using machine learning. Pre-print: https://doi.org/10.31223/X53M0J

Hohenegger et al. (2023). ICON-Sapphire: Simulating the Components of the Earth System and Their Interactions at Kilometer and Subkilometer Scales. https://doi.org/10.5194/gmd-16-779-2023.

This research has been supported by H2020 (EUCP; grant no. 776613 and XAIDA; grant no. 101003469), SNF (grant no. 167215), SDSC (DASH; C17-01), and ETH Zurich.