Project: **1179**
Project title: **Machine learning-based parametrisations and analysis for the ICON model (ICON-ML)**
Project leader: **Prof. Dr. Veronika Eyring**
Report period: **2023-05-01 to 2024-04-30**

Parametrisations of subgrid-scale effects in climate models lead to biases and uncertainties in their future projections which could be improved with machine learning (ML)-based methods. This report summarizes the progress made within the project 1179 in the last 12 months on developing ML-based parametrisations for ICON-A (Giorgetta et al., 2018) and observational products for process-oriented model evaluation. While we prioritized the development of ML microphysics parametrization through high-resolution simulations, ongoing efforts involve conducting regional large-eddy simulations to generate additional training data, as initially outlined in the proposal.

Task 2.1: Development of ML-based parametrisations for ICON-A

The development of ML-based parametrisations for ICON-A is on-going for several subgrid-scale parametrisations. This development involves the integration of individual ML-based parametrizations with ICON-A one at a time, resulting in a hybrid model, namely ICON-A-ML, reflecting their implementation separately. **(a) Cloud cover**: we have now extended our ML-based cloud cover approach to discover interpretable underlying equations describing a cloud cover parametrisation by combining symbolic regression, sequential feature selection, and physical constraints (Grundner et al., 2024; Grundner, 2024), see Fig. 1. **(b) Convection:** We have explored different methods for developing ML-based stochastic parametrisations (Behrens et al., Under review, 2024; Behrens, 2024). Our results indicate that compared to individual neural networks, multi-network ensembles improve the prediction of convective processes in the planetary boundary layer and enhance the representation of extreme precipitation and its associated diurnal cycle (not shown). Moreover, we have developed several ML-based convection parametrisations trained on coarse-grained high-resolution subgrid fluxes, assessing their performance against different benchmarks (Heuer et al., 2023). The U-Net demonstrates high accuracy in predicting convection compared to the original simulation (offline), and work is progress to couple it to ICON-A, resulting in ICON-A-ML. **c) Radiation**: We have successfully developed and coupled an ML radiation parametrisation based on coarse-resolution data (80 km) to ICON-A. The resulting hybrid model, ICON-A-ML, accurately represents the climate of the original model, and a publication detailing this work is currently being prepared (Hafner et al., 2024). **d) Gravity Waves**: Work is on-going to develop ML-based gravity wave drag parametrisation at JSC since that's where the underlying high-resolution training data is generated - only the eventual coupling will be performed at DKRZ. **e) Microphysics**: We have developed an ML-based microphysics parametrisation, including feature engineering and physics-constraints (Sarauer et al., 2024) and have performed high-resolution simulations using ICON with Hamlite to include the influence of aerosols. **f) Turbulence:** We are currently developing an ML-based parametrisation to replace ICON's vertical diffusion scheme due to atmospheric turbulence using idealized simulations. We have also designed a regional large-eddy simulation (LES) with ICON to generate real-world training data. Although setting up a suitable ICON-LES is complex (e.g., boundary conditions, external parameters, region of interest and convective regime), we currently have a near-complete configuration that addresses these intricacies and is poised for detailed simulations of atmospheric processes at high resolution. **g) Causally-informed NNs**: We have combined causal discovery and ML methodologies to enhance ML-based parametrisations (Iglesias-Suarez et al., 2024). Our demonstration showcases how integrating causality with deep learning effectively eliminates spurious correlations and optimizes the neural network algorithm.

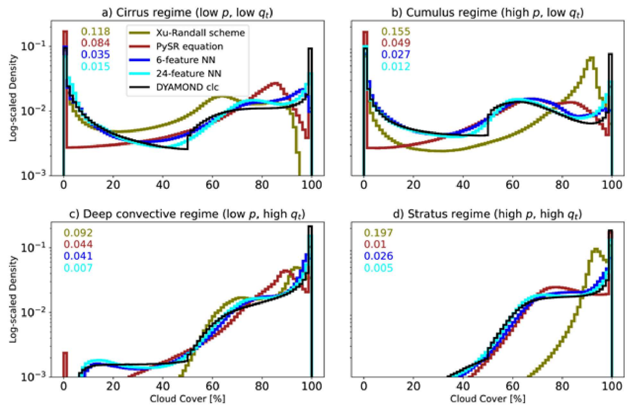

Figure 1: Predicted cloud cover distributions of selected Pareto-optimal models evaluated on the DYAMOND data, divided into four different cloud regimes. The numbers in the upper left indicate the Hellinger distance between the predicted and the actual cloud cover distributions for each model and cloud regime. Fig. 3 in Grundner et al., 2024.

Task 2.2: Development, tuning and simulations with ICON-A-ML

Once the newly developed ML-based parametrisations are coupled to ICON-A, the resulting model (ICON-A-ML) needs to be re-tuned to allow for a fair evaluation in comparison with the previously tuned ICON-A version.

We have developed an automatic tuning method following a Bayesian approach, which involves fitting Gaussian process emulators to the parameter-to-output map. These emulators not only facilitate the tuning of a state-of-the-art atmospheric model but also enable process understanding at low computational costs, as they extensively describe the parameter-to-output maps. A manuscript detailing this work is currently in preparation (Pastori et al., 2024).

In order to facilitate the evaluation of ICON-A model output using the ESMValTool with a light wrapper called ICONEval (newly developed), which further simplifies the evaluation of ICON simulations to one single command line call. For this work, the ESMValTool was run using resources of Project 1179 to make possible the evaluation of ICON-A-ML model runs. This work was published by Schlund et al., 2023 and presented at the AGU Fall Meeting.

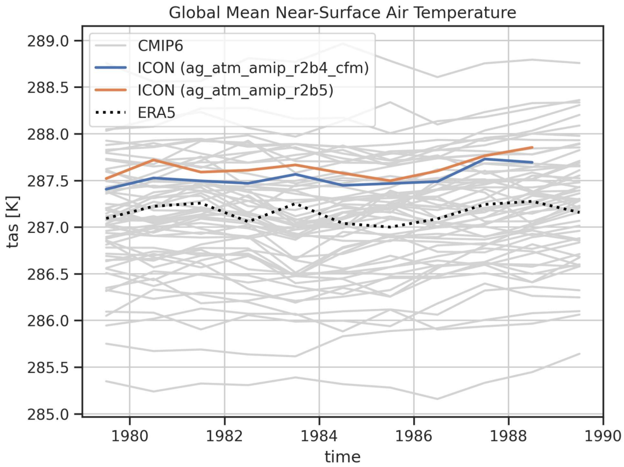

Figure 2: Time series of global and annual mean near-surface air temperature for two ICON simulations (blue and orange line), reanalysis data from ERA5 (black dotted line), and CMIP6 models (gray lines).

Task 2.3: ML-based observational products for the evaluation of ICON-ML

In order to improve understanding of the representation of clouds and their relevant processes in climate models, we had developed a new ML-based framework relying on satellite observations that assigns distributions of established cloud types to coarse data (Kaps, Lauer, Camps-Valls, et al., 2023). Expanding upon this work, we have created CCClim, a cloud-type climatology derived from available observational datasets (Kaps, Lauer, Kazeroni, et al., 2023a). This work aims to streamline the evaluation of climate models, enhancing their efficacy by enabling a more direct and objective examination of clouds according to their various types, as illustrated in the accompanying paper currently under review (Kaps, Lauer, Kazeroni, et al., 2023b). With this work and the accompanying thesis Kaps, 2024, we have completed the workpackage on ML-based observational products.

References

Behrens, G. (2024). "Machine learning-based stochasticity". PhD thesis. University of Bremen.

Behrens, G. et al. (Under review, 2024). *JAMES*. arXiv: `2402.03079`.

Giorgetta, M. A. et al. (2018). *Journal of Advances in Modeling Earth Systems* 10.7, pp. 1613–1637. DOI: `10.1029/2017ms001242`.

Grundner, A. (2024). "Data-driven Cloud Cover Parameterizations for the ICON Earth System Model Using Deep Learning and Symbolic Regression". PhD thesis. University of Bremen. DOI: `10.26092/elib/2821`.

Grundner, A. et al. (2024). *Journal of Advances in Modeling Earth Systems* 16.3, e2023MS003763. DOI: `10.1029/2023MS003763`.

Hafner, K. et al. (2024). *In Preparation.*

Heuer, H. et al. (2023). arXiv: `2311.03251 [physics.ao-ph]`.

Iglesias-Suarez, F. et al. (2024). *Journal of Geophysical Research: Atmospheres* 129.4, e2023JD039202. DOI: `10.1029/2023JD039202`.

Kaps, A. (2024). "Machine Learning Based Observational Cloud Products for Process-Oriented Climate Model Evaluation". PhD thesis. University of Bremen.

Kaps, A., A. Lauer, R. Kazeroni, et al. (2023a). Version v1. DOI: `10.5281/zenodo.8369202`.

— (2023b). *Earth System Science Data Discussions* 2023, pp. 1–26. DOI: `10.5194/essd-2023-424`.

Kaps, A., A. Lauer, G. Camps-Valls, et al. (2023). *IEEE Transactions on Geoscience and Remote Sensing* 61, pp. 1–15. DOI: `10.1109/tgrs.2023.3237008`.

Pastori, L. et al. (2024). *In Preparation.*

Sarauer, E. et al. (2024). *Accepted to ICLR.*

Schlund, M. et al. (2023). *Geoscientific Model Development* 16.1, pp. 315–333. DOI: `10.5194/gmd-16-315-2023`.