Project: **1179**
Project title: **Machine learning-based parametrisations and analysis for the ICON model (ICON-ML)**
Project leader: **Prof. Dr. Veronika Eyring**
Report period: **2022-05-01 to 2023-04-30**

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. We had also planned to perform regional simulations to generate further training data, but we had to prioritize the work on tuning the ICON model since this forms the foundation of the evaluation needed to assess the quality of the developed products.

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. **(a) Cloud cover**: During the allocation period, we published one paper on using neural nets (NNs) for parametrisation of cloud cover based on high-resolution simulations (Grundner, Beucler, Gentine, Iglesias-Suarez, et al., 2022). This work was extended during the allocation period to use ML-based methods to discover underlying equations describing a cloud cover parametrisation, see Fig. 1. A publication on this work is under preparation (Grundner, Beucler, Gentine, and Eyring, 2023). **(b) Convection:** During the allocation period, we published a paper on using ML-based methods to predict convection: Behrens et al., 2022 demonstrated that a variational autoencoder can predict subgrid-scale thermodynamics from a coarse-scale superparametrisation climate state. Building upon this study, we are designing an **ensemble approach in stochastic and deterministic** fashion to improve Deep Learning predictions of convective processes (on-going work). Ensemble approaches have the potential to overcome long-standing biases of the deep learning reproduction with respect to convection in an ESM-like configuration. Additionally, they provide uncertainty estimates for a set of variables related to convective processes. In parallel, we are also using high-resolution simulations stored on DKRZ to test various ML-based approaches to developing a parametrisation for convection, and showed that U-Nets have the best performance of the various architectures we tested. A publication on this work in under preparation. **c) Radiation and Gravity Waves**: Work is on-going to generate training data for developing ML-based parametrisations for radiation and gravity wave drag. In case of the latter, we are working on using the MODES package (Žagar et al., 2015) to identify regions with gravity waves to use as input for ML-algorithms. **d) Causally-informed NNs**: During the allocation period, we developed causally-informed neural nets in which only key direct physical drivers are used to learn subgrid-scale processes. Direct physical drivers and subgrid-scale processes are causally-linked (holding a causal relationship), in which the former determines the latter. This approach significantly improved the capabilities of the resulting causalNN-based parametrisation. A publication on this topic is under review (Iglesias-Suarez et al., 2023).

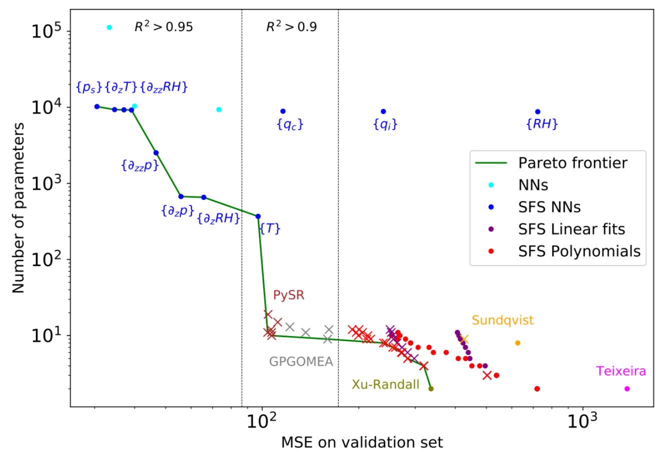

Figure 1: Developed family of data-driven cloud cover schemes in a performance x complexity – plane. Every mark corresponds to one scheme. The schemes found with the symbolic regression libraries (PySR, GPGOMEA) are equations, learned from the data, that manage to balance performance and simplicity very well. SFS is short for sequential feature selection which indicates our chosen method for data-driven feature selection.

Apart from journal publications, the results mentioned above were presented at various conferences and summer schools, e.g. at the EGU General Assembly, the AGU Fall Meeting, and the AMS Annual Meeting.

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)

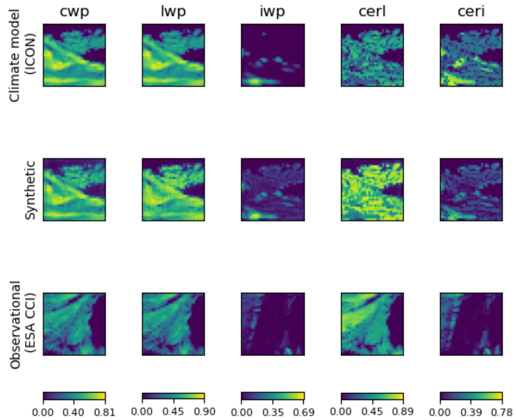

Figure 2: Domain adaptation from ICON model output to ESA CCI_Cloud, creating synthetic observations of the ICON scene. First row: 32°x32° patches of physical variables obtained from ICON simulations. Second row: The same scene after applying our generative model, obtaining synthetic observations. Third row: Reference data from ESA CCI, from the same location and similar season. The generator preserves the structure of the scene and creates characteristics that are more similar to ESA CCI (e.g. cerl). Variables: cwp/lwp/iwp : cloud/liquid/ice water path, cerl/ceri: effective radius of liquid/ice particles.

needs to be re-tuned to allow for a fair evaluation in comparison with the previously tuned ICON-A version. For this, we are working on an automatic tuning procedure of ICON-A. We performed ensemble-runs of ICON-A with a range of values of free parameters within the parametrisations to conduct automatic tuning and simulations of 10-20 years. These were evaluated with the Earth System Model Evaluation Tool (ESMValTool) for sanity checks. The work on developing the auto-tuning procedure is on-going.

In order to facilitate the evaluation of ICON-A model output using the ESMValTool without prior processing (CMORization), an on-the-fly CMORizer for ICON-A was developed. 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.

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 developed a new ML-based framework relying on satellite observations that assigns distributions of established cloud types to coarse data (Kaps et al., 2023). The inputs to the ML algorithms used here are obtained from the Cloud Product of the MODIS instrument, which operates aboard the Terra and Aqua satellites. This method facilitates a more objective evaluation of clouds in ESMs and improves the consistency of cloud process analysis. This work was published in (Kaps et al., 2023) and presented at the EGU Meeting during the allocation period. We also processed large amounts of data using this method and trained NNs for domain adaptation in order to make climate model output from ICON-A more similar and comparable to observational products, see Fig. 2.

References

Behrens, G. et al. (2022). "Non-Linear Dimensionality Reduction With a Variational Encoder Decoder to Understand Convective Processes in Climate Models". *Journal of Advances in Modeling Earth Systems* 14.8. DOI: 10.1029/2022ms003130.

Giorgetta, M. A. et al. (2018). "ICON-A, the Atmosphere Component of the ICON Earth System Model: I. Model Description". *Journal of Advances in Modeling Earth Systems* 10.7, pp. 1613–1637. DOI: 10.1029/2017ms001242.

Grundner, A., T. Beucler, P. Gentine, and V. Eyring (2023). "Data-Driven Equation Discovery of a Cloud Cover Parameterization". *Journal of Advances in Modeling Earth Systems*, In prep.

Grundner, A., T. Beucler, P. Gentine, F. Iglesias-Suarez, et al. (2022). "Deep Learning Based Cloud Cover Parameterization for ICON". *Journal of Advances in Modeling Earth Systems* 14.2, e2021MS002959. DOI: 10.1029/2021ms002959.

Iglesias-Suarez, F. et al. (2023). "Causally-informed deep learning to improve climate models and projections". *In Preparation.*

Kaps, A. et al. (2023). "Machine-Learned Cloud Classes From Satellite Data for Process-Oriented Climate Model Evaluation". *IEEE Transactions on Geoscience and Remote Sensing* 61, pp. 1–15. DOI: 10.1109/tgrs.2023.3237008.

Schlund, M. et al. (2023). "Evaluation of native Earth system model output with ESMValTool v2.6.0". *Geoscientific Model Development* 16.1, pp. 315–333. DOI: 10.5194/gmd-16-315-2023. URL: https://gmd.copernicus.org/articles/16/315/2023/.

Žagar, N. et al. (2015). "Normal-Mode Function Representation of Global 3-D Data Sets: Open-Access Software for the Atmospheric Research Community". *Geoscientific Model Development* 8.4, pp. 1169–1195. DOI: 10.5194/gmd-8-1169-2015.