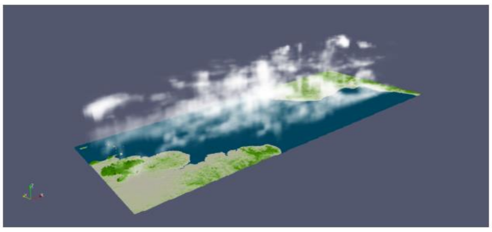
Project: **1179**

Project title: **Machine learning-based parametrisations and analysis for the ICON model (ICON-ML)**

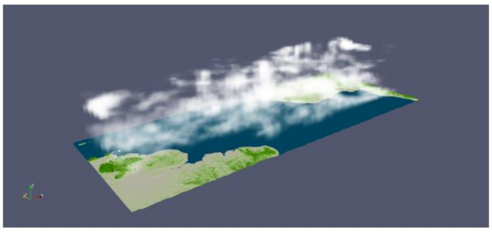
Principal investigator: **Prof. Dr. Veronika Eyring (DLR)**

Report period: **2021-05-01 to 2022-04-30**

Advancing Earth system models requires tackling key biases and uncertainties in their projections that largely stem from parametrisations such as cloud cover. This report summarizes the progress made within the project 1179 in the last 12 months regarding the development of machine learning (ML)-based parametrisations for ICON [1] in order to advance climate modelling and the generation of observational products for process-oriented model evaluation.



(a) NN cloud cover



(b) Ground Truth

Figure 1: Offline ML-based cloud cover parametrisation. *Panels a) and b) show cloud cover snapshots produced with the column-based NN trained and evaluated on the coarse-grained NARVAL R2B4 data running offline, i.e., not coupled to the ICON simulation. (a) cloud scene estimated by the NN, b) reference cloud scene from coarse-grained NARVAL data)* [2]*. Note that some columns over land could not be vertically interpolated due to overlapping topography and are therefore missing in a)*

Task (2.1): ML-based parametrisations for the ICON-A model

We used high-resolution (2.5 - 5 km) ICON simulations from the NARVAL and QUBICC projects to develop neural nets (NNs) for ML-based cloud parametrisations, and to subsequently implement them into ICON-A replacing the current parametrisation: the first version of ICON-A-ML. First, we evaluated the “offline” performance of the newly developed ML-based cloud cover parametrisation (i.e., predictions compared to the “truth”, namely the coarse-grained high-resolution simulation). As an example, Figure 1 shows the ability of the ML-based parametrisation to accurately predict vertically resolved cloud cover [2]. We have successfully coupled the ML-parametrisation to ICON to run ICON-A-ML online (see Task 2.2).

In related work, we have demonstrated that a variational-autoencoder (VAE) can predict subgrid-scale thermodynamics from a coarse-scale superparametrisation (SP) climate state [3], see Figure 2. The VAE’s latent space can distinguish convective regimes, including shallow/deep/no convection and reveals the main sources of convective predictability at different latitudes. Additional insight into physical mechanisms of different parametrisations is gained via causal discovery. Preliminary results show that causality can successfully help optimize the inputs of the NN (i.e., using only physical-drivers), the NN’s architecture (i.e., simpler models with increased interpretability), and achieve stable prognostic simulations (i.e., NN coupled to the GCM) [4].

Work is on-going to replace further parametrisations such as convection, gravity drag and radiation with dedicated ML-based parametrisations. We are also in the process to automate the retuning process of ICON-A-ML after a new parametrisation has been implemented, which will significantly speed up the development process.

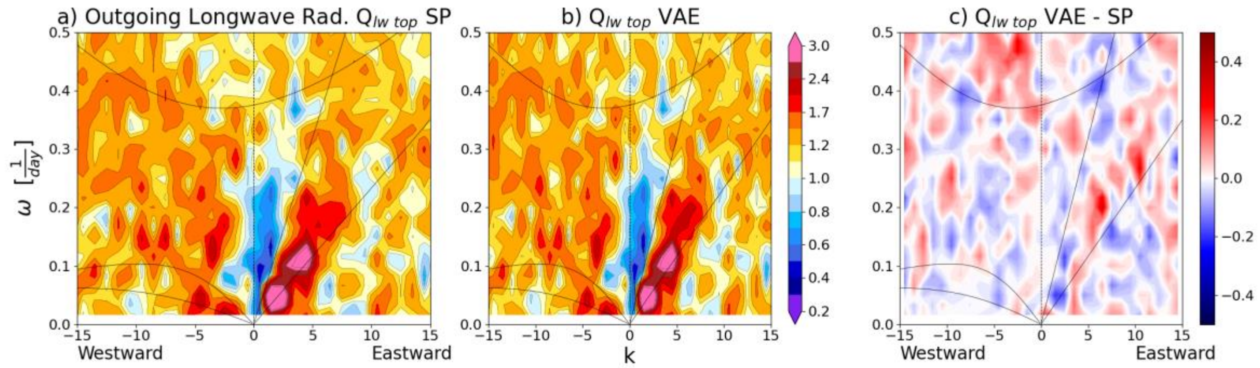


Figure 2: Wheeler Kiladis diagram based on tropical outgoing longwave radiation *[15°N-15°S] of SP (a), of VAE predictions (b) and the absolute difference of spatio-temporal wave spectra VAE-SP (c) for 1 year of SP simulations [3].*

Task (2.2): ICON-A and ICON-A-ML simulations

We have successfully implemented the above ML-based cloud cover parametrisation (see Task 2.1) into ICON-A via the Fortran-Keras Bridge (FKB; [https://github.com/scientific-computing/ FKB](https://github.com/scientific-computing/FKB)), resulting in a hybrid ICON-A-ML model. Figure 3 shows first results of running the parametrisation online.

The evaluation of the newly developed ICON-A-ML model is accomplished via the ESMValTool [5], which has been successfully extended to handle native ICON output (icosahedral grid) [6]. The ESMValTool provides common pre-processing operations and a large collection of diagnostics that entails climate mean state, trends, and variability. Furthermore, the ESMValTool also allows benchmarking different model versions. Therefore, we have also performed a number of reference runs with the ICON-A model to compare with the ICON-A-ML model. Although this is still work in progress, first tests (see Figure 3) show reasonable cloud cover results of the ICON-A-ML model (80 km resolution; R2B5 grid).

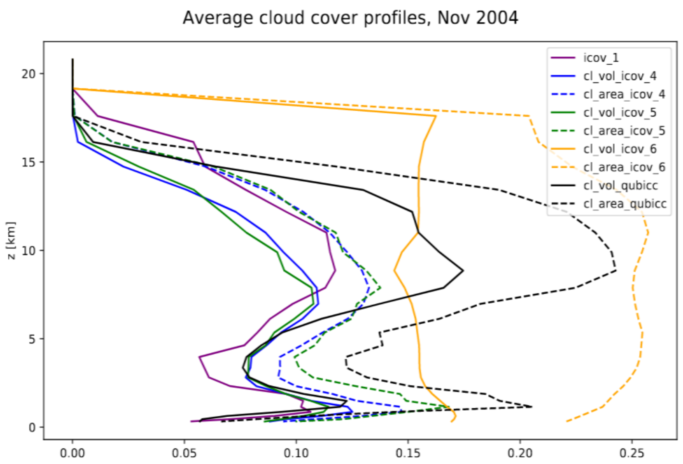


Figure 3: First preliminary results of running the ML-based cloud cover parametrisation online: cloud cover vs. height. *icov_1: Conventional ICON-A simulation with the Sundqvist scheme as cloud cover scheme; cl_area/vol_icov_4/5/6: ICON simulations with different types of NNs instead of the Sundqvist scheme; cl_vol/area_qubicc: Coarse-grained cloud volume and area fraction of a QUBICC simulation (of Nov, 2004).*

Task (2.3): ML-based cloud observational products

Process-oriented model evaluation enables a more rigorous and physics-related understanding of the model performance and its biases. Evaluating cloud properties by cloud type is a promising way to identify weaknesses in the representation of clouds in models.

We have developed a satellite-data-driven ML-based evaluation framework for clouds that is applicable to climate model output of varying horizontal and temporal resolution. As an example, Figure 4 shows the predicted fraction of cirrus clouds in coarse data that is similar to model output [7]. Cloud types predicted this way show characteristics that are in agreement with expectations from observations.

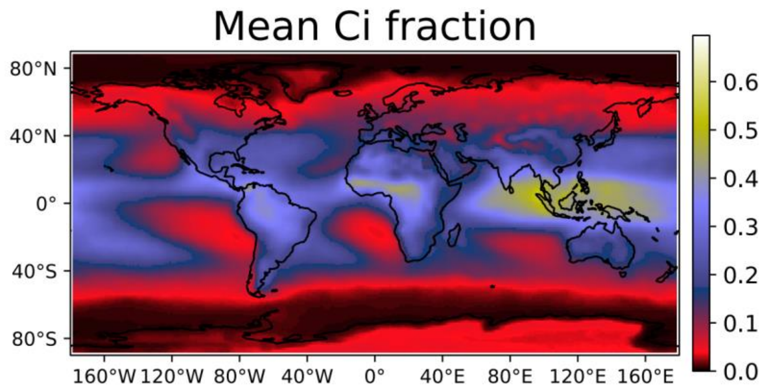


Figure 4 Mean predicted fraction of the Ci (cirrus/cirrostratus clouds) type per grid cell, averaged over a 2.5-year timespan. *The predictions where performed on coarse resolution (5°) ESA Cloud_cci data.*

References

[1] Giorgetta, M. A. et al., “ICON-A, the atmosphere component of the ICON earth system model: I. Model description,” *JAMES*, vol. 10, p. 1613, 2018.
[2] Grundner, A. et al. "Deep Learning Based Cloud Cover Parameterization for ICON," *arXiv*, 2112.11317 2021.
[3] Behrens, G. et al. “Non-Linear Dimensionality Reduction with a Variational Autoencoder Decoder to Understand Convective Processes in Climate Models,” *arXiv*, 2204.08708, 2022.
[4] Iglesias-Suarez, F. et al., “Causal discovery and deep learning to improve convection in climate models,” *in preparation*, 2022.
[5] Righi, M. et al., “Earth System Model Evaluation Tool (ESMValTool) v2.0 – technical overview,” *Geoscientific Model Development*, vol. 13, p. 1179–1199, 3 2020.
[6] Schlund, M. et al., “Evaluation of native Earth system model output with the ESMValTool,” *in preparation*, 2022.
[7] Kaps, A. et al., “Machine-learned cloud classes from satellite data for process-oriented climate model evaluation,” *IEEE Trans. Geosci. Rem. Sens.*, 2022 (submitted).