**Seminar on Physics and Chemistry
of the Atmosphere
10.11.2023, WiSe 2023/24, IUP Bremen**

Data-Driven Cloud Cover Parameterizations for ICON using Deep Learning and Symbolic Regression

Arthur Grundner

(DLR)

Abstract

A promising approach to improving cloud parameterizations in climate models, and thus climate projections, is to train machine learning algorithms on the coarse-grained output of high-resolution storm-resolving model (SRM) simulations. The ICOsahedral Non-hydrostatic (ICON) modeling framework enables simulations ranging from numerical weather prediction to climate projections, making it an ideal target for developing machine learning based parameterizations. Here, we focus on the improvement of the semi-empirical cloud cover parameterization used in the ICON Earth System Model. It diagnoses subgrid-scale fractional cloud cover from large-scale variables in every grid cell based on very simple assumptions. To instead parameterize cloud cover with more detailed complexity, we first develop three different types of neural networks (NNs) that differ in the degree of vertical locality they assume for diagnosing cloud cover. The NNs accurately estimate cloud cover in their training domain and globally-trained NNs can even estimate it for a distinct regional SRM. Using the game theory based interpretability library SHapley Additive exPlanations, we analyze our most non-local NN and identify an overemphasis on specific humidity and cloud ice as the reason why it cannot perfectly generalize from global to regional coarse-grained SRM data. While our NNs already achieve excellent predictive performance ($R^2 > 0.9$) with as few as three features, they are climate model specific and require additional tools for post-hoc interpretation. To avoid these limitations, we also add symbolic regression, sequential feature selection, and physical constraints to a combined hierarchical modeling framework. Analytical equations derived from this framework are interpretable by construction and easily transferable to other grids or climate models. Our best equation balances performance and complexity, achieving a performance comparable to that of NNs ($R^2 = 0.94$) while remaining simple (with only 11 trainable parameters) and physically consistent. It learns to utilize the vertical relative humidity gradient to detect elusive marine stratocumulus clouds. Our results show the potential of deep learning to derive accurate cloud cover parameterizations from global SRMs. They also demonstrate the effectiveness of symbolic regression to discover interpretable, physically consistent, and nonlinear equations for cloud cover.