Seminar on Physics and Chemistry of the Atmosphere
26.01.2024, WiSe 2023/24, IUP Bremen

Understanding and Modelling Convection with Machine Learning

Gunnar Behrens[1,2]

[1]University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany
[2]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Abstract

Earth System Models (ESMs) are essential tools projecting changes in the Earth system due to forcings. Despite advances in the quality of ESMs, they display longstanding biases (e.g. in cloud radiative effects) compared to observations. These biases are related to the use of traditional parameterization schemes that approximate the effect of subgrid processes such as convection on the large-scale environment. In recent years, training machine learning (ML) models on high-resolution data and using them as ML-based parameterizations proved a way to overcome the reliance on traditional convective schemes. However, these ML algorithms usually lack interpretability due to their large internal dimensionality, resulting in reduced trustworthiness. We use Variational Encoder Decoder structures (VEDs), a non-linear dimensionality reduction technique with a latent space, to learn and understand convective processes in a climate model. The VED distinguishes convective regimes and large-scale drivers in its latent space and thus helps improve our understanding of convective processes and their connection to the large-scale environment. Our results demonstrate that VEDs can be applied as powerful postprocessing technique to give new insights into non-linear processes in the Earth system. Apart from its limited interpretability, current ML has deficiencies in reproducing convective processes in the planetary boundary layer due to the underlying stochasticity. To improve the representation of these convective processes we propose both, novel stochastic, and deterministic ML ensembles. We show that both types of ensembles reproduce convective processes in the planetary boundary layer with enhanced skill compared to single state-of-the-art ML algorithms. Moreover, using ensembles enables us to quantify the uncertainty of the parameterizations in a multi-variate test case. This enhances the interpretability and trustworthiness of the developed schemes. We use the two best-performing schemes as machine-learned subgrid parametrizations in an ESMs and conduct a hybrid model simulation of more than five months. The ensemble stochastic parameterizations have a stabilizing effect compared to individual deterministic ML algorithms. The results of the hybrid simulations with our developed ensemble and stochastic schemes indicate an improved reproduction of precipitation extremes and more skilful representation of the diurnal cycle of precipitation compared to a traditional convection scheme. Overall, our research paves a way forward with interpretable and uncertainty-quantifying ML for non-linear processes in the Earth system. These are steps towards the operational use of machine learning-based parameterizations in ESMs.