

EGU22-7011

<https://doi.org/10.5194/egusphere-egu22-7011>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Unlocking the potential of ML for Earth and Environment researchers

Tobias Weigel^{1,2}, Frauke Albrecht^{1,2}, Caroline Arnold^{1,2}, Danu Caus^{1,2}, Harsh Grover^{1,2}, and Andrey Vlasenko^{1,2}

¹German Climate Computing Center, Department of Data Management, Hamburg, Germany (weigel@dkrz.de)

²Helmholtz AI

This presentation reports on support done under the aegis of Helmholtz AI for a wide range of machine learning based solutions for research questions related to Earth and Environmental sciences. We will give insight into typical problem statements from Earth observation and Earth system modeling that are good candidates for experimentation with ML methods and report on our accumulated experience tackling such challenges with individual support projects. We address these projects in an agile, iterative manner and during the definition phase, we direct special attention towards assembling practically meaningful demonstrators within a couple of months. A recent focus of our work lies on tackling software engineering concerns for building ML-ESM hybrids.

Our implementation workflow covers stages from data exploration to model tuning. A project may often start with evaluating available data and deciding on basic feasibility, apparent limitations such as biases or a lack of labels, and splitting into training and test data. Setting up a data processing workflow to subselect and compile training data is often the next step, followed by setting up a model architecture. We have made good experience with automatic tooling to tune hyperparameters and test and optimize network architectures. In typical implementation projects, these stages may repeat many times to improve results and cover aspects such as errors due to confusing samples, incorporating domain model knowledge, testing alternative architectures and ML approaches, and dealing with memory limitations and performance optimization.

Over the past two years, we have supported Helmholtz-based researchers from many subdisciplines on making the best use of ML methods along with these steps. Example projects include wind speed regression on GNSS-R data, emulation of atmospheric chemistry modeling, Earth System model parameterizations with ML, marine litter detection, and rogue waves prediction. The poster presentation will highlight selected best practices across these projects. We are happy to share our experience as it may prove useful to applications in wider Earth System modeling. If you are interested in discussing your challenge with us, please feel free to chat with us.