

Virtual Event: ECMWF-ESA Workshop on Machine Learning for Earth System Observation and Prediction



Contribution ID: 61

Type: **Oral presentation**

Accelerating and Explaining Earth System Process Models with Machine Learning

Monday, 5 October 2020 11:30 (45 minutes)

Earth scientists have developed complex models to represent the interactions among small particles in the atmosphere, such as liquid water droplets and volatile organic compounds. These complex models produce significantly different results from their simplified counterparts but are too computationally expensive to run directly within weather and climate models. Machine learning emulators trained on a limited set of runs from the complex models have the potential to approximate the results of these complex models at a far smaller computational cost. However, there are many open questions about the best approaches to generating emulation data, training ML emulators, evaluating them both offline and within Earth System Models, and explaining the sensitivities of the emulator.

In this talk, we will discuss the development of machine learning emulators for warm rain bin and superdroplet microphysics as well as the Generator of Explicit Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A) model. For microphysics, we ran CAM6 for 2 years with the cloud-to-rain processes handled by either the bin or superdroplet scheme and saved the inputs and outputs of the scheme globally at selected time steps. We used this information to train a set of neural network emulators. The neural network emulators are able to approximate the behavior of the bin and superdroplet schemes while running within CAM 6 but at a computational cost close to the bulk MG2 scheme. Machine learning interpretation techniques also reveal the relative contributions and sensitivities of the different inputs to the emulator. We will discuss lessons learned about both the training process and the resulting model climate.

For GECKO-A, we ran the model forward in time with multiple sets of fixed atmospheric conditions and different precursor compounds. Then we trained fully connected and recurrent neural networks to emulate GECKO-A's. We tested the different machine learning methods by running them forward in time with both fixed and varying atmospheric conditions. We plan to incorporate the GECKO-A emulator into a full 3D atmospheric model to evaluate how the transitions between precursors, gases, and aerosols evolve spatio-temporally. We will also discuss lessons learned from working with this dataset and the challenges of problem formulation and evaluation. Finally, we are releasing data from both of these domains as machine learning challenge problems to encourage further innovation in this area by the broader Earth Science and Machine Learning communities.

Thematic area

1. Machine Learning for Model Identification and Development - Including Model identification, Fast Emulation of Parameterisations, Data driven Parameterisations

Primary author: GAGNE, David (National Center for Atmospheric Research)

Co-authors: Dr HODZIC, Alma (NCAR); GETTELMAN, Andrew (NCAR); Mr BECKER, Charlie (NCAR); Dr CHEN, Chih-Chieh (Jack) (NCAR); Ms GANTOS, Gabrielle (NCAR); Ms LAWRENCE, Keely (NCAR); Dr WANG, Siyuan (NCAR)

Presenter: GAGNE, David (National Center for Atmospheric Research)

Session Classification: Session 1: ML for the Earth System - Setting the scene

Track Classification: ECMWF-ESA Workshop on Machine Learning for Earth System Observation and Prediction