

EGU2020-13794

<https://doi.org/10.5194/egusphere-egu2020-13794>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



Data-driven parametrizations in numerical models using data assimilation and machine learning.

Julien Brajard^{1,2}, Alberto Carrassi^{3,4}, Marc Bocquet⁵, and Laurent Bertino¹

¹NERSC, Bergen, Norway (julien.brajard@nersc.no)

²Sorbonne Université, Paris, France

³Dept of Meteorology, University of Reading, UK

⁴Mathematical Institute, University of Utrecht, Netherlands

⁵CEREA, joint laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

Can we build a machine learning parametrization in a numerical model using sparse and noisy observations?

In recent years, machine learning (ML) has been proposed to devise data-driven parametrizations of unresolved processes in dynamical numerical models. In most of the cases, ML is trained by coarse-graining high-resolution simulations to provide a dense, unnoisy target state (or even the tendency of the model).

Our goal is to go beyond the use of high-resolution simulations and train ML-based parametrization using direct data. Furthermore, we intentionally place ourselves in the realistic scenario of noisy and sparse observations.

The algorithm proposed in this work derives from the algorithm presented by the same authors in <https://arxiv.org/abs/2001.01520>. The principle is to first apply data assimilation (DA) techniques to estimate the full state of the system from a non-parametrized model, referred hereafter as the physical model. The parametrization term to be estimated is viewed as a model error in the DA system. In a second step, ML is used to define the parametrization, e.g., a predictor of the model error given the state of the system. Finally, the ML system is incorporated within the physical model to produce a hybrid model, combining a physical core with a ML-based parametrization.

The approach is applied to dynamical systems from low to intermediate complexity. The DA component of the proposed approach relies on an ensemble Kalman filter/smoothener while the parametrization is represented by a convolutional neural network.

We show that the hybrid model yields better performance than the physical model in terms of both short-term (forecast skill) and long-term (power spectrum, Lyapunov exponents) properties. Sensitivity to the noise and density of observation is also assessed.