Geophysical Research Abstracts
Vol. 21, EGU2019-6843, 2019
EGU General Assembly 2019

Combining Data Assimilation and Machine Learning to emulate a numerical model from noisy and sparse observations.

Julien Brajard (1,2), Marc Bocquet (3), Alberto Carrassi (1,4), and Laurent Bertino (1)

(1) Nansen Environment and Remote Sensing, Bergen, Norway., (2) Sorbonne University, CNRS-IRD-MNHN, LOCEAN, Paris, France., (3) CEREA joint laboratory Ecole des Pont ParisTech and EdF R&D, Université Paris-Est, France., (4) Geophysical Institute, University of Bergen, Norway.

Is it possible to emulate a numerical model from noisy and sparse observations? How realistic and skillful can it be?

A popular way to produce data-driven models from a set of observations is to apply machine-learning methods. These approaches proved efficient for chaotic systems using various machine learning algorithms such as, for instance, deep-learning or reservoir computing. In most of previous works, these approaches were used with a noisy-free, completely observed, state vector, and their performance was evaluated on short-term predictability skills. When it comes to noisy and sparse observations, data assimilation techniques provide the natural tools to produce an optimal estimation of the state of a system, provided a numerical model, is available (though imperfect).

In this work, we consider the case in which noisy and partial observations of a given phenomena are available but the evolution model is unknown. The idea is to apply machine-learning and data assimilation algorithms alternatively to *learn* the underlying dynamics of the system. First, the data assimilation procedure completes the state estimate, which is then used as a training set for machine learning. Reciprocally, the machine learning-based model is used as the forward dynamical model in the data assimilation framework. The sequence is reiterated with increasing complexity of the machine learning model.

This combined approach is tested numerically in a twin experiments setup with chaotic models. The machine learning algorithm used in this work is a convolutional neural network architecture (CNN) whereas the data assimilation scheme is an ensemble Kalman filter (using the finite-size variant to avoid the inflation: EnKF-N). The resulting CNN showed both forecast skills and abilities to reproduce the "climate" (*i.e.* spectral properties and statistical moments) of the underlying dynamical model on long-term simulations. Perspectives towards large-scale systems will be discussed on the basis of the benchmarks presented.