Contribution ID: **68** Type: **not specified**

Using machine learning in geophysical data assimilation (some of the issues and some ideas)

Tuesday, 20 December 2022 10:00 (30 minutes)

In recent years, data assimilation, and more generally the climate science modelling enterprise have been influenced by the rapid advent of artificial intelligence, in particular machine learning (ML), opening the path to various form of ML-based methodology.
In this talk we will schematically show how ML can be included in the prediction and DA workflow in three different ways. First, in a so-called "non-intrusive"ML, we will show the use of supervised learning to estimate the local Lyapunov exponents (LLEs) based exclusively on the system's state [1]. In this approach, ML is used as a supplementary tool, added to the given physical model. Our results prove ML is successful in retrieving the correct LLEs, although the skill is itself dependent on the degree of local homogeneity of the LLEs on the system's attractor.
In the second and third approach, ML is used to substitute fully [4] or partly [5]a physical model with a surrogate one reconstructed from data. Nevertheless, for high-dimensional chaotic dynamics such as geophysical flows this reconstruction is hampered by (i) the partial and noisy observations that can realistically be gathered, (ii) the need to learn from long time series of data, and (iii) the unstable nature of the dynamics. To achieve such inference successfully we have suggested to combine DA and ML in several ways. We will show how to unify these approaches from a Bayesian perspective, together with a description of the numerous similarities between them [2,3]. We will show that the use of DA in the combined approach is pivotal to extract much information from the sparse, noisy, data. The full surrogate model achieves prediction skill up to 4 to 5 Lyapunov time, and its power spectra density is almost identical to that of the original data, except for the high-frequency modes which are not well captured [4]. The ML-based parametrization of the unresolved scales in the third approach [5] is also extremely skilful. This has been studied using a coupled atmosphere-ocean model and again the use of coupled DA [6] in the combined DA-ML method makes possible to exploit the data information from one model compartment (e.g., the ocean) to the other (e.g., the atmosphere).

[1] Ayers D, J Amezcua, A Carrassi and V Ohija, 2022. Supervised machine learning to estimate instabilities in chaotic systems: estimation of local Lyapunov exponents, Under Review, Available at https://arxiv.org/abs/2202.04944
[2] Bocquet M, Brajard J, A Carrassi, and L Bertino, 2019. Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models, Nonlin. Proc. Geophys.,26, 175–193
[3] Bocquet M, Brajard J, A Carrassi, and L Bertino, 2020. Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization, Found. Data Sci.,2, 55–80
[4] Brajard J, A Carrassi, M Bocquet and L Bertino, 2020. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model., J. Comp. Sci.,44, 101171
[5] Brajard J, A Carrassi, M Bocquet and L Bertino. 2021. Combining data assimilation and machine learning to infer unresolved scale parametrisation., Phil. Trans A of the Roy Soc., 379 (2194). 20200086
[6] Tondeur M, A. Carrassi, S. Vannitsem and M. Bocquet: On temporal scale separation in coupled data assimilation with the ensemble Kalman filter. J. Stat. Phys., 44. 101171, 2020

Primary author: CARRASSI, Alberto (University of Bologna)

Presenter: CARRASSI, Alberto (University of Bologna)

Session Classification: Session 5