Bayesian online algorithms for learning data-driven models of chaotic dynamics

Marc Bocquet
CEREA, École des Ponts and EdF R&D, Île-de-France, France
marc.bocquet@enpc.fr

The advent of machine and more specifically deep learning techniques significantly boosts the capabilities of data assimilation and inverse problem techniques used in the geosciences. It also spurs new, more ambitious goals for data assimilation in high dimensions. One of the key, currently very popular, area of research consists in learning data-driven models of dynamical systems. With the natural constraints of geoscience, i.e., sparse and noisy observations, this typically requires the joint use of data assimilation and neural networks. However, the vast majority of algorithms are offline; they rely on a set of observations from the physical system, which must be available before the start of the training.

We propose new algorithms that update the knowledge of the surrogate (i.e., data-driven) model when new observations become available. We carry out this objective with both variational (weak-constraint 4D-Var like) and ensemble (EnKF and IEnKS like) techniques. We test these algorithms on low-order Lorenz models, on quasi-geostrophic models, the ERA5 dataset, and sea-ice dynamics. Remarkably, in several cases, the online algorithms significantly outperform the offline ones. This opens the way to adaptive surrogate models that progressively learn trends and conform to real-time constraints of operational weather forecasting.