Hybrid Data Assimilation and Deep Learning: Improving Earth-system representation and prediction

Phd Thesis proposal for a SCAI funding

February 2022

Keywords: *Data Assimilation, Deep Learning, Satellite Data, Numerical modeling of the environment*

Context

Monitoring and modeling the ocean and atmosphere have been a constant scientific preoccupation since the earliest days of scientific thought. The high-dimensional and chaotic nature of earth's systems enforce the use of data from different sensors. But geophysical observations are imperfect by nature either they are incomplete, noisy or indirect. Estimating the state of a physical system from such data has then led to a variety of inverse problems. When involving knowledge about the system evolution in the form of a dynamical model, these problems are solved with data assimilation [3], a set of techniques producing state-of-the-art results in various numerical weather forecasting application. When it comes to operational forecasting, meteorological centers tend to use ensemble methods, such as the ensemble Kalman filter [5], and variational methods such as 4D-Var [12]. Even though data assimilation is now a very mature field, challenges still remains and particularly on model error representation. Indeed, dynamical models are often incomplete, ill-posed, and model errors statistics poorly known due to the high dimension of the system.

Over the last decade, deep learning has revolutionized data science, offering methods to discover and model complex relationships from data. Such methods can learn spatio-temporal relations which makes them appealing for numerical weather prediction. However, deep learning approaches may have some drawbacks: as these algorithms are only data-based, they can failed to achieve the task when the data set is too incomplete, and their explainability is still being explored.Today, great efforts are engaged by the machine learning community to embed prior information in neural networks [14, 13]. For instance a physical law represented by a partial differential equation may appear in the loss function [4].

Objectives

In this thesis, we are interested in injecting deep learning techniques and architectures into data assimilation schemes in order to adress physics-based model error representation by learning either dynamical systems or model error statistics.

Parallels between data assimilation and neural networks architectures have already been drawn as both can be seen from the Bayesian point of view [10]. Moreover, both training a neural network and estimating a system state with variational data assimilation are performed in similar manner. By leveraging automatic differentiation, gradient backpropagation through hidden layers or through physical models behaves in a similar manner. Variational assimilation methods can then be easily implemented using tools from the deep learning community, as we show in [8].

To tackle the model-error representation challenge, we are interested in hybrid architecture combining deep learning and differentiable physcis layers [4, 13, 14, 6]. In a previous research, we have show that is possible to learn a missing dynamics from the error model [9] alternating data assimilation optimization and deep learning optimization to complete physics-based model. Using the same tools, we also proposed an End-to-End architecture to retrieve the initial condition of a Lorentz system highly noisy (work currently submitted). This thesis is then the continuation of these works on several aspects:

- Proposing new methods to learn model error representation from imperfect data;
- Using a well-suited deep neural architecture to generate a solution of the variational problem that can act as a handcrafted regularization. This means that the control parameters are shifted from the system state space to the neural network parameters space (a preliminary work has been currently submitted);
- Exploiting the generative properties of neural networks in order to infer plausible ensembles (Variational Auto-encoders, Completions by Deep Prior with various initialisations...). These ensembles then be used to estimate covariance matrices for Data Assimilation purposes.

The developed methods will then be applied on dynamical systems related to fluid motion and particularly advection-based motion which intervenes a lot in the atmosphere and the ocean. First test will be twin experiments involving 1 and 2 dimensional toy model (e.g. Lorenz96, Shallow water) before expanding them to datasets such as the Meteonet database [11] or the Mercator ocean reanalysis [7]. The aim of these is to be able to expand on the group's previous results on rain nowcasts [2] and ocean circulation [submitted work under revision].

Institutional Context

The student will be hosted in LIP6 (Pequan team), located in the center of Paris, under the supervision of Dominique Béréziat (MCF, HDR at LIP6/Sorbonne Université) and Anastase Charantonis (MCF at LOCEAN/École Nationale Supérieure d'informatique pour l'Industrie et l'Entreprise). The group has an internal dynamic focused on Artificial Intelligence for Climate, which includes other Phd students and researchers working on similar interdisciplinary problems, with a strong dynamic, including seminars and a journal club (more can be found on ai4climate.lip6.fr).

Contacts: dominique.bereziat@lip6.fr, anastase.charantonis@ensiie.fr.

References

[1] M. Bonavita, Y. Trémolet, E. Hólm, STK Lang, M. Chrust, M. Janiskova, P. Lopez, P. Laloyaux, P. de Rosnay, M. Fisher, M. Hamrud, and S. English. A strategy for data assimilation. Technical Report 800, ECMWF, 2017.

[2] Vincent Bouget, Dominique Béréziat, Julien Brajard, Anastase Charantonis, and Arthur Filoche. Fusion of rain radar images and wind forecasts in a deep learning model applied to rain nowcasting. *Remote Sensing*, 13(2):246, 2021.

[3] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e535, 2018.

[4] E. de Bézenac, A. Pajot, and P. Gallinari. Deep learning for physical processes: Incorporating prior scientific knowledge. *CoRR*, abs/1711.07970, 2017.

[5] G. Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.

[6] A. Farchi, M. Bocquet, P. Laloyaux, M. Bonavita, and Q. Malartic. A comparison of combined data assimilation and machine learning methods for offline and online model error correction, 2021.

[7] Nicolas Ferry, Laurent Parent, Gilles Garric, Bernard Barnier, Nicolas C Jourdain, et al. Mercator global eddy permitting ocean reanalysis glorys1v1: Description and results. *Mercator-Ocean Quarterly Newsletter*, 36:15–27, 2010.

[8] A. Filoche, D. Béréziat, J. Brajard, and AA Charantonis. Variational assimilation of Geophysical images leveraging deep learning tools. In *ORASIS 2021*, September 2021.

[9] A. Filoche, J. Brajard, A. Charantonis, and D. Béréziat. Completing physics-based models by learning hidden dynamics through data assimilation. In *NeurIPS 2020, AI for Earth Sciences Workshop*, pages 1–4, 2020. https://hal.sorbonne-universite.fr/hal-03004938.

[10] A. Geer. Learning earth system models from observations: machine learning or data assimilation? 863, May 2020.

[11] G. Larvor, L. Berthomier, V. Chabot, B. Le Pape, B. Pradel, and L. Perez. Meteonet, an open reference weather dataset by meteo-france, 2020. https://meteonet.umr-cnrm.fr/.

[12] F-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, 38A(10):97, 1986.

[13] L. Mosser, O. Dubrule, and M. Blunt. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior, 2018.

[14] J. Tompson, K Schlachter, P. Sprechmann, and K. Perlin. Accelerating eulerian fluid simulation with convolutional networks. *CoRR*, abs/1607.03597, 2016.