Offre de Stage IPSL 2022

(soutenu par le programme EUR IPSL-*Climate Graduate School)*

Titre du sujet de stage : Learning variational data assimilation directly from observations. Application to geophysical motion estimation.

Description du sujet (1 page maximum) : voir document joint

Résumé en anglais (5 lignes) : Adapt an existing End to End 4Var algorithm to jointly retrieve the sea surface circulation and its model of dynamics.

Responsable du stage (Nom/prénom/statut) : Béréziat Dominique MCF

Laboratoire concerné : LIP6 / SU

Adresse à laquelle a lieu le stage : 4 place Jussieu, Paris

Equipe de recherche concernée (si pertinent) ou autre participant à l'encadrement du stage:

Niveau du stage (Licence, M1, M2, internship) : M1/M2

Licence ou Master(s) où sera proposé le sujet :

Thème scientifique de l'IPSL concerné : 4Dvar / Ocean

Durée du stage : 5 mois

Période : 1/3/22 → 31/8/22

Rémunération de l'ordre de 580 euros par mois

Est-il prévu une thèse dans le prolongement du stage ? non.

Learning Variational Data Assimilation directly from Observations Application to Geophysical Motion Estimation

A. Filoche, D. Béréziat (LIP6, Sorbonne University)

November 21, 2022

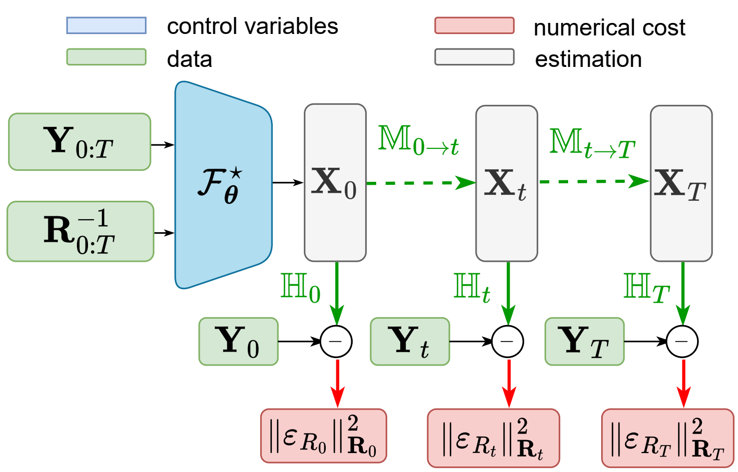
Context

Variational data assimilation and deep learning share many algorithmic aspects in common. While the former focuses on system state estimation, the latter provides great inductive biases to learn complex relationships. It has already been argued that both methods can benefit from each other [6, 5]. Data assimilation provides a proper Bayesian framework to combine sparse and noisy data with physics-based knowledge while deep learning can leverage a collection of data extracting complex relationships from it. Hybrid methods have already been developed either to correct model error [3, 2], to jointly estimate parameters and system state [1], or to fasten the assimilation process [7]. Most of these algorithms rely on **iterative** optimization schemes alternating data assimilation and machine learning steps.

Following a previous work [4], we propose a hybrid architecture learning the assimilation task **directly** from partial and noisy observations, using the mechanistic constraint of the 4DVAR algorithm. Preliminary experiments on the Lorenz96 system show that the proposed method was able to learn the desired inversion with interesting regularizing and computational properties. The aim of this internship is to adapt the algorithm to higher-dimensional systems involving motion fields.

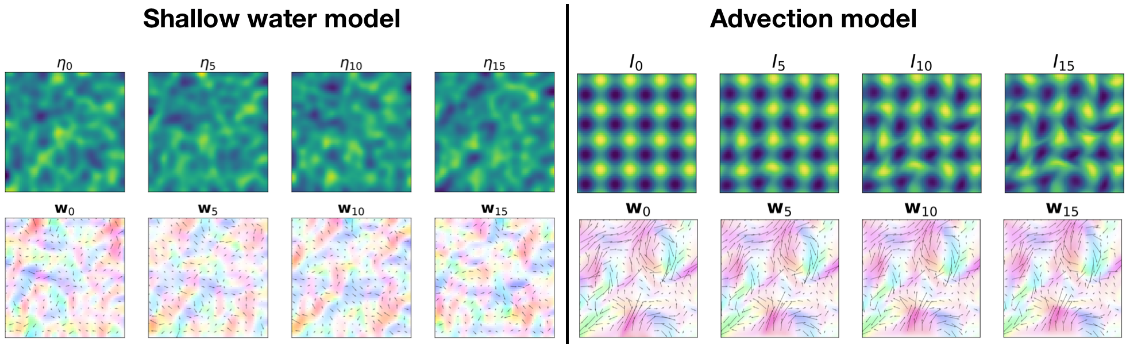
Algorithm

The proposed architecture bridges a neural network and a mechanistic model to directly learn system state estimation from a collection of partial and noisy observations. The optimization is done only in one step using the variational assimilation loss function. Here is a schematic view of the hybrid architecture learning the 4DVAR inversion.



Dataset – Geophysical Motion

The internship will focus on applying this algorithm using shallow water and advection models. Numerical schemes are already coded in differentiable software (PyTorch). The assimilation goal is to recover the motion field.



Potential experiments

- Baseline: 4DVAR, iterative approach (learning over 4DVAR)
- Sensitivity to noise/sparsity
- Accounting for uncertainty quantification with generative models

Administrative stuffs

The internship is funded by SCAI for a duration of 5-6 months. It will be held in LIP6 laboratory (Sorbonne University) located in the center of Paris.

References

[1] M. Bocquet et al. "Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization". In: *Foundations of Data Science* 2.1 (2020), pp. 55–80. ISSN: 2639-8001. DOI: 10.3934/fods.2020004.

[2] P. Düben et al. "Machine learning at ECMWF: A roadmap for the next 10 years". In: *ECMWF Technical Memoranda* 878 (2021).

[3] A. Farchi et al. "Using machine learning to correct model error in data assimilation and forecast applications". In: *Quarterly Journal of the Royal Meteorological Society* 147.739 (2021), pp. 3067–3084. DOI: doi.org/10.1002/qj.4116.

[4] A. Filoche et al. *Learning 4DVAR inversion directly from observations*. arXiv. http://arxiv.org/abs/2211.09741. Oct. 2022.

[5] A. Geer. "Learning earth system models from observations: machine learning or data assimilation?" In: *Philosophical Transactions of the Royal Society A* 379 (Feb. 2021).

[6] M. Reichstein et al. "Deep learning and process understanding for data-driven Earth system science". In: *Nature* 566.7743 (2019), pp. 195–204. DOI: 10.1038/s41586-019-0912-1.

[7] P. Wu et al. "Fast data assimilation (FDA): Data assimilation by machine learning for faster optimize model state". In: *Journal of Computational Science* 51 (2021), p. 101323. ISSN: 1877-7503. DOI: 10.1016/j.jocs.2021.101323.