Deep learning enhanced ultra-fast nudging data assimilation

Maksym Veremchuk[1], K. Andrea Scott[1], Zhao Pan[1]
1 Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON, Canada

Data assimilation (DA) is a key technique in fluid dynamics that combines observational data with physical models to estimate the system state. A widely used DA method is the Ensemble Kalman Filter (EnKF) [1], which operates by evolving an ensemble of model states and updating them using observations. Given the initial ensemble of m states: $\boldsymbol{x}_0^1, \ldots, \boldsymbol{x}_0^m$, the EnKF performs a forecast step to predict the next ensemble sequence using the formula:

$$\tilde{\boldsymbol{x}}_t^i = F(\boldsymbol{x}_{t-1}^i) + \boldsymbol{\eta}_t^i, \quad \boldsymbol{\eta}_t^i \sim \mathcal{N}_m(0, \mathcal{Q}_t), \tag{1}$$

$$\boldsymbol{x}_t^i = \tilde{\boldsymbol{x}}_t^i + \mathcal{K}_t\big(\boldsymbol{y}_t + \boldsymbol{\epsilon}_t^i - \mathcal{H}_t(\tilde{\boldsymbol{x}}_t^i)\big), \quad \boldsymbol{\epsilon}_t^i \sim \mathcal{N}_m(0, \mathcal{R}_t), \tag{2}$$

where $\tilde{\boldsymbol{x}}_t^i$ is the background state for the i-th ensemble member, F is a forward model which is typically governed by physics-based equations, $\mathcal{Q}_t$ and $\mathcal{R}_t$ are model and observational error covariance matrices respectively, $\boldsymbol{y}_t$ is the available observations, $\mathcal{H}_t$ is an observational operator and $\mathcal{K}_t$ is a Kalman gain.

Despite its high accuracy, EnKF is computationally expensive. The computation of the Kalman gain requires matrix inversions, which are computationally expensive, especially for large systems. In addition to accurate reconstruction of fine-scale dynamics, the number of ensemble members may need to be large (e.g., $O(100)$), further increasing the computational cost.

A faster alternative to EnKF is Azouani-Olson-Titi (AOT)-nudging DA [2]. This approach avoids ensembles and replaces the Kalman gain with a scalar term:

$$\boldsymbol{x}_t = F(\boldsymbol{x}_{t-1}) - \kappa\left[I(\boldsymbol{x}_{t-1}) - I(\boldsymbol{y}_{t-1})\right], \tag{3}$$

where κ is a relaxation parameter and I is an interpolation operator.

In addition to faster inference than EnKF, AOT-nudging has been shown to achieve comparable accuracy to EnKF for Navier-Stokes-type flows when the parameter κ is properly chosen. However, it still faces two key limitations: the physical model F remains computationally expensive on large domains, and classical interpolation operators I (e.g., radial basis function or bicubic) often fail to accurately capture complex dynamics from sparse observations.

In our research, we propose a new AOT-nudging framework that addresses both limitations. We replace the physical model F with a neural operator [3] that approximates the system dynamics while enabling significantly faster inference. In addition, we replace the classical interpolation operator with a learned interpolation operator, which provides higher accuracy while maintaining comparable or even lower computational cost on large domains.

Our framework was evaluated on the quasi-geostrophic flow [4], a simplified model of the Navier-Stokes equations that captures mesoscale ocean dynamics. The proposed approach demonstrates superior performance compared to EnKF, achieving more than 3000× speedup while maintaining comparable accuracy, even when only 5% of sensor observations are available.

Email address for correspondence: *mrveremc@uwaterloo.ca*

References

[1] Matthias Katzfuss, Jonathan R. Stroud, and Christopher K. Wikle. "Understanding the Ensemble Kalman Filter". In: *The American Statistician* 70.4 (Oct. 1, 2016), pp. 350–357. ISSN: 0003-1305, 1537-2731. DOI: 10.1080/00031305.2016.1141709. URL: https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1141709.

[2] Abderrahim Azouani, Eric Olson, and Edriss S. Titi. "Continuous Data Assimilation Using General Interpolant Observables". In: *Journal of Nonlinear Science* 24.2 (Apr. 2014), pp. 277–304. ISSN: 0938-8974, 1432-1467. arXiv: 1304.0997[nlin,physics:physics].

[3] Gege Wen et al. "U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow". In: *Advances in Water Resources* 163 (2022), p. 104180.

[4] Jeffrey Covington, Nan Chen, and Monica M. Wilhelmus. "Bridging Gaps in the Climate Observation Network: A Physics-Based Nonlinear Dynamical Interpolation of Lagrangian Ice Floe Measurements via Data-Driven Stochastic Models". In: *Journal of Advances in Modeling Earth Systems* (2022).

Email address for correspondence: *mrveremc@uwaterloo.ca*