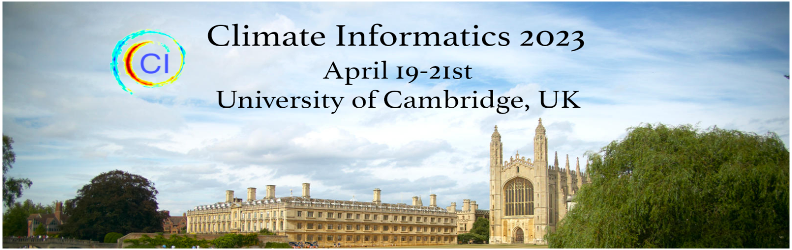


Climate Informatics 2023

Ensemble-based 4DVarNet uncertainty quantification for the reconstruction of Sea Surface Height dynamics

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Maxime Beauchamp, Quentin Febvre and Ronan Fablet
maxime.beauchamp@imt-atlantique.fr April 19, 2023

Table of contents

Problem statement

ML frameworks and data assimilation (DA) schemes: growing interest to address challenges in Earth system modeling.
This includes both:

- the integration of learning-based components in DA schemes
- the design of DA-inspired learning-based schemes to address inverse problems and uncertainty quantification (UQ) for dynamical processes

Problem

Using a data assimilation (DA) state space formulation, we aim at estimating the hidden space

$$\mathbf{x} = \{\mathbf{x}_k(\mathcal{D})\}$$

- $\mathbf{y}(\Omega) = \{\mathbf{y}_k(\Omega_k)\}$: the partial and potentially noisy observational dataset
- $\Omega = \{\Omega_k\} \subset \mathcal{D}$, the subdomain with observations and index k refers to time t_k.

Current solutions

1. Covariance-based **Kriging** [Chilès and Delfiner, 2012], **BLUE**, **OI** [Traon et al., 1998]
and **SPDE**-based version [Lindgren et al., 2011]

$$\mathbf{x}^\star = \mathbf{\Sigma_{xy}}\mathbf{\Sigma_{yy}}^{-1}\mathbf{y} = -\mathbf{Q_{xx}}^{-1}\mathbf{Q_{xy}}\mathbf{y}$$

2. Model-based **sequential** data assimilation (DA), (En)KF [Evensen, 2009] or **variational** assimilation, (3DVar, 4DVar) [Asch et al., 2016] with a state-space formulation:

$$\begin{cases} \mathbf{x}_{k+1} & = \mathcal{M}_{k+1}(\mathbf{x}_k) + \eta_k \\ \mathbf{y}_k & = \mathcal{H}_k(\mathbf{x}_k) + \varepsilon_k \end{cases}$$

Hybrid methods to combine flow-dependent covariance matrix from EnKF into variational schemes

3. Data-driven DA: **Analog forecasting operator** embedded in EnKF [Tandeo et al., 2015], **hybrid ML/DA** synergy for the inference of unresolved scale parametrizations [Brajard et al., 2021, Bocquet et al., 2019, O'Gorman and Dwyer, 2018, Rasp et al., 2018]

4DVarNet

4DVarNet

Joint learning of both prior φ and solver Γ backboned on variational data assimilation scheme:

$$\mathbf{x}^{\star} = \text{argmin}_{\mathbf{x}} \mathcal{J}(\mathbf{x}, \mathbf{y}, \Omega) = \text{argmin}_{\mathbf{x}} ||\mathbf{x} - \mathbf{y}||^2_{\Omega} + \lambda ||\mathbf{x} - \varphi(\mathbf{x})||^2$$

For inverse problems with time-related processes, the minimization of functional $\mathcal{J}_{\Phi}$ usually involves an iterative gradient-based approach, denoted here as the solver Γ:

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - \alpha \nabla_{\mathbf{x}} \mathcal{J}_{\Phi}(\mathbf{x}^{(i)}, \mathbf{y}, \Omega)$$

$\mathbf{x}$ spatio-temporal state: $\mathbf{x_0}, \cdots, \mathbf{x_T}$, $\mathbf{y}$ the observations on $\Omega \subset \mathcal{D}$,
φ a trainable prior, Γ a trainable solver to speed up the gradient descent.

Let denote by $\Psi_{\Phi,\Gamma}(\mathbf{x}^{(0)}, \mathbf{y}, \Omega)$ the output of the 4DVarNet learning scheme, then the joint learning of operators $\{\Phi, \Gamma\}$ is stated as the minimization of a reconstruction cost:

$$\arg\min_{\Phi,\Gamma} \mathcal{L}(\mathbf{x}, \mathbf{x}^{\star}) \text{ s.t. } \mathbf{x}^{\star} = \Psi_{\Phi,\Gamma}(\mathbf{x}^{(0)}, \mathbf{y}, \Omega)$$

with $\mathcal{L}$ the MSE w.r.t Ground Truth.

Observation System Simulation Experiment

- Ground truth dataset **x**: high-resolution 1/60° NATL60 configuration of the NEMO (Nucleus for European Modeling of the Ocean) model
- A $10^\circ \times 10^\circ$ GULFSTREAM region is used with downgraded resolution to $1/20^\circ$, principally led by mesoscale processes

Figure: GULFSTREAM domain

- OSSE : pseudo-altimetric nadir and SWOT observational datasets $\mathbf{y} = \{\mathbf{y}_k\}$ at time t_k are generated by a realistic sub-sampling satellite constellations on subdomain $\Omega = \{\Omega_k\}$ of the grid.

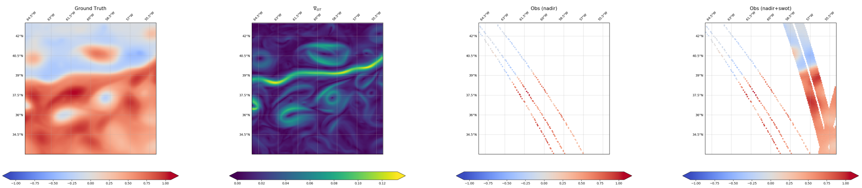


Figure: From left to right: Ground Truth (SSH & ∇_{SSH}) and pseudo-observations (nadir & nadir+swot) on August 4, 2013

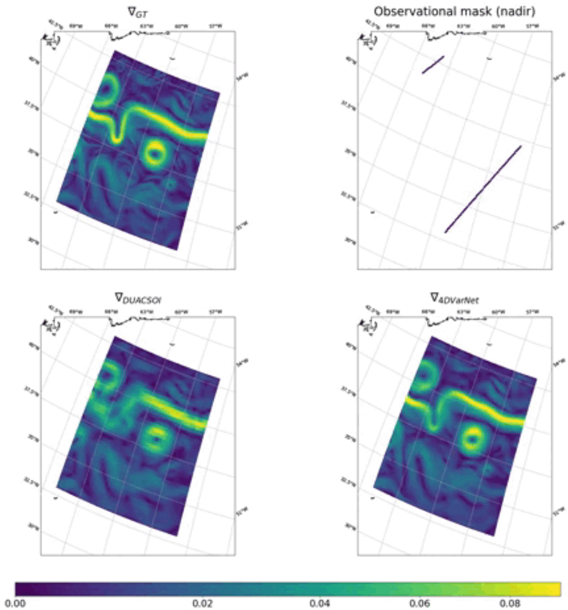


Figure: GT, Obs, OI and 4DVarNet

4DVarNet: on-going and related works

1. Sparse sampling operator
   $||\mathcal{H}(\mathbf{z}) * (\mathbf{x} - \mathbf{y})||$ with $||H(\mathbf{z})||_1 < \epsilon$
2. Multimodal observation:
   $||\mathbf{x} - \mathbf{y}||^2 + \alpha||\mathcal{G} * \mathbf{x} - \mathcal{F} * \mathbf{z}||^2$
3. Calibration operator
   $||\mathcal{H}(\mathbf{y}) - \mathbf{x}||^2$
4. Link with Gaussian Processes:

$$\mathbf{x}^\star = \mathrm{argmin}_{\mathbf{x}} \mathcal{J}(\mathbf{x}, \mathbf{y}, \Omega)$$
$$= \mathrm{argmin}_{\mathbf{x}} ||\mathbf{x} - \mathbf{y}||^2_\Omega + \mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x}$$

where $\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x} = ||\mathbf{x} - \Phi(\mathbf{x})||^2$ and $\Phi = (1 - \mathbf{S})$, and $\mathbf{S}$ is the square root of the precision matrix.

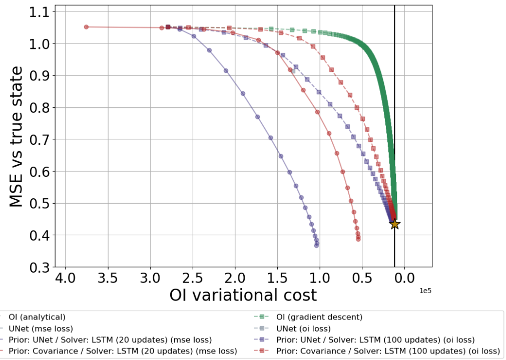


Figure: Asymptotic convergence of 4DVarNet to OI (Gaussian case)

Towards a stochastic 4DVarNet: the prior distribution

1. Key idea: sample a simulation $\mathbf{x}^{s,i}$ with same properties than the state

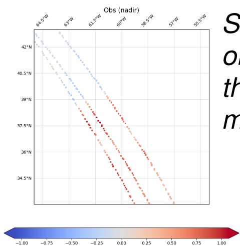
$$i.E\left[\mathbf{x}^{s,i}(\mathbf{u})\right] = \mathbf{m}(\mathbf{u}) \tag{1a}$$

$$ii.var\left[\mathbf{x}^{s,i}(\mathbf{u})\right] = \sigma_x^2(\mathbf{u}) \tag{1b}$$

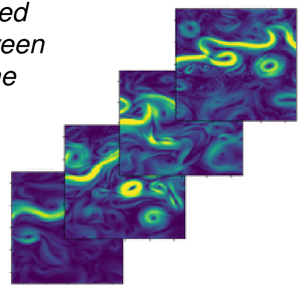
$$iii.C\left[\mathbf{x}^{s,i}(\mathbf{u}), \mathbf{x}^{s,i}(\mathbf{u}')\right] = C\left[\mathbf{x}(\mathbf{u}), \mathbf{x}(\mathbf{u}')\right] \tag{1c}$$

   How to do it? Draw (from catalog): easy, Generate (more difficult)

2. In real-world application, use simulation-based training on real datasets works fine (see *Learning from simulations for real data*, Febvre et al. this afternoon), then use the simulation catalog and analog strategy [Tandeo et al., 2015] to draw samples from the prior distribution.



Sample in the prior based on analog strategy between the observations and the model-based catalog



Towards a stochastic 4DVarNet: estimating the posterior distribution

1. Posterior pdf: geostatistical-based conditioning of the simulation

$$\mathbf{x}^{\star,i} = \mathbf{x}^{\star} + \{\mathbf{x}^{s,i} - \mathbf{x}^{\star,s,i}\} \tag{2}$$

2. Properties (gaussian case):

$$i. E\left[\mathbf{x}^{\star,i}(\mathbf{u})\right] \rightarrow E\left[\mathbf{x}^{\star}(\mathbf{u})\right] \tag{3a}$$

$$ii. var\left[\mathbf{x}^{\star,i}(\mathbf{u})\right] \rightarrow var\left[\mathbf{x}^{\star}(\mathbf{u})\right] \tag{3b}$$

$$iii. C\left[\mathbf{x}^{\star,i}(\mathbf{u}), \mathbf{x}^{\star,i}(\mathbf{u}')\right] \rightarrow C(\mathbf{u},\mathbf{u}') - \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha(\mathbf{u})\lambda_\beta(\mathbf{u}')C(\mathbf{u}_\alpha,\mathbf{u}_\beta) \tag{3c}$$

where $\alpha, \beta = 1, \cdots, n$ denote the observation index and λ_α are the optimal weights

3. If $\mathbf{x}$ not Gaussian/linear, 4DVarNet may outperform OI and traditional DA [Beauchamp et al., 2022, Fablet et al., 2021].
   - unbiased
   - lower MSE w.r.t the ground truth, i.e. lower variance of its error
   - Then, running N simulations conditioned by 4DVarNet $\rightarrow p_{\mathbf{x}|\mathbf{y}}$ with both improvements on the two first moments $\mathbf{x}^\star$ and $\mathbf{P}^\star$

Towards a stochastic 4DVarNet: estimating the posterior distribution

We apply our ensemble-based 4DVarNet approach with 60 members on a 6 nadir constellation in January 2017.

- Focus on a small bottom-left 50×50 pixels subdomain: the members are able to reproduce different realistic patterns and small eddy structures
- SSH spread from 2016-12-31 to 2017-01-25 (every 5 days): observations help to reduce UQ. Away from the observations, the uncertainty grows quickly, not only based on the distance from the nadir altimeters but also influenced by the SSH spatio-temporal dynamics.

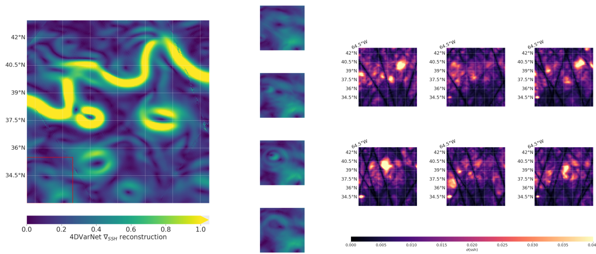


Figure: Ensemble-based 4DVarNet (Left: variability amongst members; Right: spread variations every 5 days)

On-going and future works

1. Generate samples from the prior distribution: one idea, use a GP approximation of φ in 4DVarNet driven by stochastic PDEs:

$$\mathcal{L}\mathbf{x} = \mathbf{z}$$

   1. $\mathcal{L}$ a fractional differential operator (embedding advection, diffusion, etc.) and $\mathbf{z}$ a white/colored noise
   2. 4DVarNet is trained with an augmented state formulation $\{\mathbf{x}, \boldsymbol{\Theta}\}$ with $\boldsymbol{\Theta}$ the (non-stationary) SPDE parameters
   3. Joint learning of SPDE parametrization $\boldsymbol{\Theta}$ and solvers Γ:

$$\arg\min_{\boldsymbol{\Theta},\Gamma} \mathcal{L}(\mathbf{x}, \boldsymbol{\Theta}^\star, \mathbf{x}^\star) \text{ s.t. } \mathbf{x}^\star = \Psi_{\boldsymbol{\Theta},\Gamma}(\mathbf{x}^{(0)}, \mathbf{y}, \Omega)$$

   with $\mathcal{L} = \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}_2$
   $\mathcal{L}_1$ the MSE w.r.t Ground Truth (**reconstruction cost**)
   $\mathcal{L}_2$ the negative log-likelihood $-\mathcal{L}(\boldsymbol{\Theta}^\star|\mathbf{x})$ (**prior regularization cost**).

2. Go to a full neural formulation with GAN or diffusion-based models (the SPDE-based GP approximation is in fact a "linear" diffusion model) for the prior distribution

Take-home messages

4DVarNet: current state

- We can bridge DNN and variational models to solve inverse problem. **Key tools**, unrolling of the gradient-step descent with a recurrent neural network:

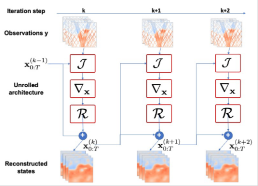


Figure: Unrolling emulation of the gradient-descent with $\mathcal{R}$ a recurrent neural network, typically a ConvLSTM

- Learning **jointly** neural (but also physical prior), observation models and solvers
- Considerably ease the use of **multimodal observations** (computational cost and trainable feature extraction operator)
- **Stochastic implementation** for now based on a post-processing of mean state estimation, but training scheme embedding the UQ is on its way

References I

M. Asch, M. Bocquet, and M. Nodet. *Data Assimilation*. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, Dec. 2016. ISBN 978-1-61197-453-9. doi: 10.1137/1.9781611974546. URL https://doi.org/10.1137/1.9781611974546.

M. Beauchamp, J. Thompson, H. Georgenthum, Q. Febvre, and R. Fablet. Learning neural optimal interpolation models and solvers, 2022. URL https://arxiv.org/abs/2211.07209.

M. Bocquet, J. Brajard, A. Carrassi, and L. Bertino. Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models. *Nonlinear Processes in Geophysics*, 26(3):143–162, 2019. doi: 10.5194/npg-26-143-2019. URL https://npg.copernicus.org/articles/26/143/2019/.

J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino. Combining data assimilation and machine learning to infer unresolved scale parametrization, 2021. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2020.0086.

J. Chilès and P. Delfiner. *Geostatistics : modeling spatial uncertainty*. Wiley, New-York, second edition, 2012.

G. Evensen. *Data Assimilation*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 9783642037108 9783642037115. URL http://link.springer.com/10.1007/978-3-642-03711-5.

References II

R. Fablet, M. Beauchamp, L. Drumetz, and F. Rousseau. Joint interpolation and representation learning for irregularly sampled satellite-derived geophysical fields. *Frontiers in Applied Mathematics and Statistics*, 7:25, 2021. ISSN 2297-4687. doi: 10.3389/fams.2021.655224. URL https://www.frontiersin.org/article/10.3389/fams.2021.655224.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73 (4):423–498, 2011. doi: 10.1111/j.1467-9868.2011.00777.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x.

P. A. O'Gorman and J. G. Dwyer. Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, 10(10):2548–2563, 2018. doi: https://doi.org/10.1029/2018MS001351. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001351.

S. Rasp, M. S. Pritchard, and P. Gentine. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115 (39):9684–9689, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1810286115. URL https://www.pnas.org/content/115/39/9684.

References III

P. Tandeo, P. Ailliot, J. Ruiz, A. Hannart, B. Chapron, A. Cuzol, V. Monbet, R. Easton, and R. Fablet. Combining Analog Method and Ensemble Data Assimilation: Application to the Lorenz-63 Chaotic System. In V. Lakshmanan, E. Gilleland, A. McGovern, and M. Tingley, editors, *Machine Learning and Data Mining Approaches to Climate Science*, pages 3–12. Springer, 2015.

P.-Y. Traon, F. Nadal, and N. Ducet. An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technology - J ATMOS OCEAN TECHNOL*, 15:522–534, 04 1998. doi: 10.1175/1520-0426(1998)015<0522:AIMMOM>2.0.CO;2.