



Bridging Data assimilation & SPDE with a generative extension of 4DVarNet schemes

A step towards physically-sounded generation of large ensembles in Geosciences



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

Maxime Beauchamp, Ronan Fablet and many others...

maxime.beauchamp@imt-atlantique.fr

maxime.beauchamp76@gmail.com

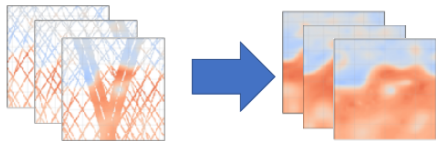
May 22, 2024

Table of contents

- 1 Problem statement**
- 2 Introduction
- 3 Current solutions
 - OI
 - state space DA
 - ML DA
- 4 4DVarNet: a Neural scheme backboneed on variational DA
- 5 Ensemble-based neural variational scheme
- 6 Conclusion
- 7 References

Challenges in spatial oceanography

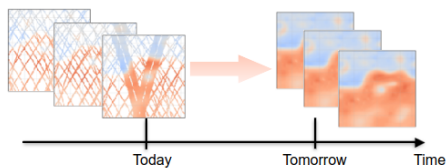
How to reconstruct from observations ?



Partial observations y

Gap-free states x

How to simulate/forecast from observations ?

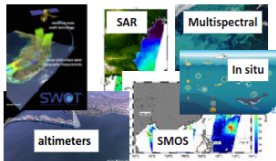


Today

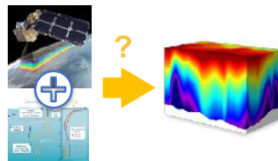
Tomorrow

Time

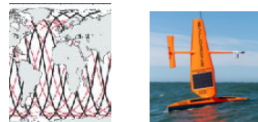
Using multi-tracer observations ?



From surface to interior ?



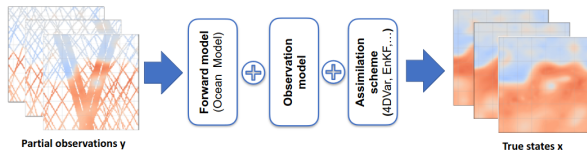
Where and what to sample ?



DA or?, and? ML

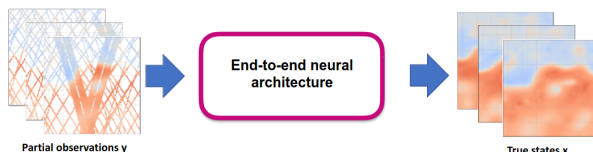
Data assimilation

Each component designed using model-driven principles and mostly independently. . . . But somewhat difficult to fully exploit observational datasets.



Machine learning

How to exploit prior knowledge for DA ? Can we calibrate all the components of a DA scheme at once?



Problem statement

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 .

Optimal Interpolation

Current solutions

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

$$\mathbf{x}^* = \boldsymbol{\Sigma}_{\mathbf{xy}} \boldsymbol{\Sigma}_{\mathbf{yy}}^{-1} \mathbf{y} = -\mathbf{Q}_{\mathbf{xx}}^{-1} \mathbf{Q}_{\mathbf{xy}} \mathbf{y}$$

where:

$\boldsymbol{\Sigma}_{\mathbf{xy}} = \text{Cov}\{\mathbf{x}_k, \mathbf{y}\} \in \mathbb{R}^{m \times p}$ denotes the (cross-)covariance between \mathbf{x}_k and \mathbf{y}

$\boldsymbol{\Sigma}_{\mathbf{yy}} = \text{Cov}\{\mathbf{y}, \mathbf{y}\} \in \mathbb{R}^{p \times p}$ denotes the covariance between the observations

State space data assimilation

Current solutions

2 Model-based data assimilation (DA) with a state-space formulation:

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

- **sequential** DA: (En)KF [Evensen, 2009]
- **variational** assimilation, (3DVar, 4DVar) [Asch et al., 2016]
- **Hybrid** methods to combine flow-dependent covariance matrix from EnKF into variational schemes

Machine learning and data assimilation

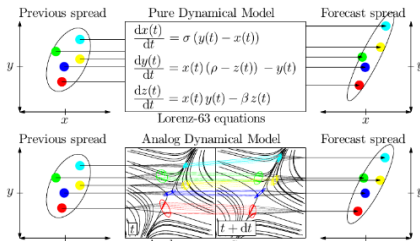
ML frameworks and data assimilation (DA) schemes: growing interest to address challenges in Earth system modeling.

This includes both:

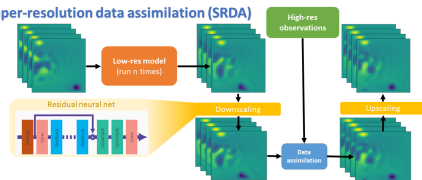
- the integration of learning-based components in DA schemes
- the design of DA-inspired learning-based schemes to address inverse problems and uncertainty quantification (UQ) for dynamical processes

Current solutions

- 3 Data-driven DA based on **analog forecasting operator** embedded in EnKF [Tandeo et al., 2015]
- 4 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]



Super-resolution data assimilation (SRDA)



4DVarNet: a Neural scheme backboneed on variational DA

Neural variational scheme

Joint learning of both prior Φ and solver Γ backboneed on variational data assimilation scheme:

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \mathcal{J}(\mathbf{x}, \mathbf{y}, \Omega) = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_{\Omega}^2 + \lambda \|\mathbf{x} - \Phi(\mathbf{x})\|^2$$

For inverse problems with time-related processes, the minimization of functional \mathcal{J}_{Φ} usually involves an iterative gradient-based approach, denoted here as the solver Γ_S , with a neural operator \mathcal{S} that learns how to speed up the gradient descent:

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

\mathbf{x} spatio-temporal state: $\mathbf{x}_0, \dots, \mathbf{x}_T$, \mathbf{y} the observations on $\Omega \subset \mathcal{D}$,
 Φ a trainable prior, Γ a trainable solver to speed up the gradient descent.

Both Φ and \mathcal{S} are neural operators: we benefit from deep learning automatic differentiation tools to run the gradient descent of the variational scheme

Neural variational scheme

Interpretation

- Fully data-driven generic interpolation/forecasting tool
- Φ must be seen as a projection operator of state \mathbf{x} . It is not here the emulator of any deterministic *prior* model...
- But it can be (future works) or something else (Gaussian Process for instance, see e.g. Beauchamp et al. [2023])

In the current version:

- Φ involves a UNet architecture
- \mathcal{S} has a 2D convolutional LSTM architecture

Bi-level optimization for end-to-end learning

Let denote by $\Psi_{\Phi, \Gamma}(\mathbf{x}^{(0)}, \mathbf{y}, \Omega)$ the output of the neural learning scheme, then the joint learning of operators $\{\Phi, \Gamma\}$ to optimize the gradient-based descent on the **inner variational cost**:

$$\mathcal{J}(\mathbf{x}, \mathbf{y}, \Omega) = \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_{\Omega}^2 + \lambda \|\mathbf{x} - \Phi(\mathbf{x})\|^2$$

is stated as the minimization of the **outer reconstruction loss**:

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

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

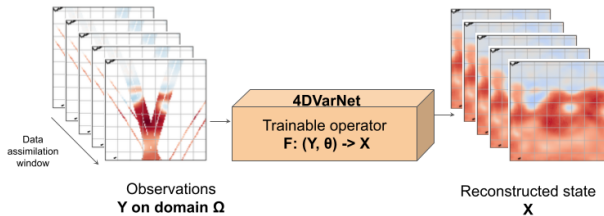


Figure 1: Generic scheme of the DL approach: the so-called neural network 4DVarNet is fed with a sequence of gappy observations as inputs, and aims at providing a reconstruction of the entire state sequence

Which training loss

- **Supervised loss:** The GT is known everywhere (OSSE setup with model-based or reanalysis GT).

$$\mathcal{L}(\mathbf{x}, \mathbf{x}^*) = \|\mathbf{x} - \mathbf{x}^*\|^2$$

- **Semi-supervised loss:** Only the observations are available on subdomain Ω of \mathcal{D} . Then we remove some additional observations on subdomain $\tilde{\Omega} \subset \Omega$ and the loss becomes:

$$\mathcal{L}(\mathbf{y}, \mathbf{x}^*) = \|\mathbf{y} - \mathbf{x}^*\|_{\tilde{\Omega}}^2$$

Observation System Simulation Experiment

- Ground truth dataset \mathbf{x} : high-resolution $1/60^\circ$ 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
- 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.

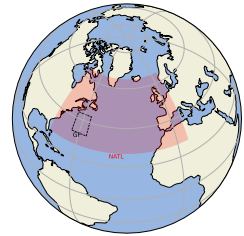


Figure 2: GULFSTREAM domain

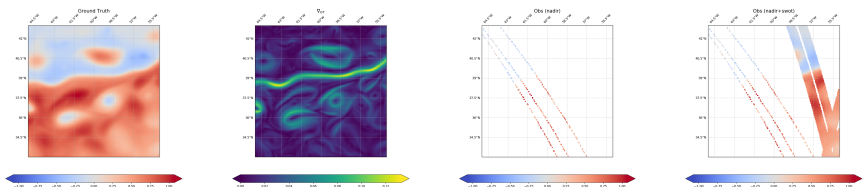


Figure 3: From left to right: Ground Truth (SSH & ∇_{SSH}) and pseudo-observations (nadir & nadir+swot) on August 4, 2013

Figure 4: GT, Obs, OI and the neural variational scheme

4DVarNet: Transfer learning

- 5 Transfer learning to real data at regional/global scale: **Learn** on simulations and **inference** on real altimetric data

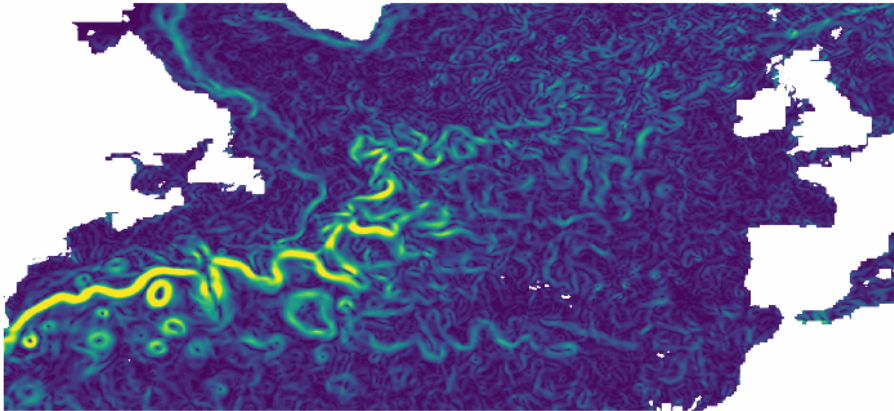


Figure 5: Transfer learning to real nadir data on the whole North Atlantic basin scale

4DVarNet: a link to GP

- Link with Gaussian Processes: let consider a Gaussian prior $\mathbf{x}^b \sim \mathcal{N}(\mu^b, \mathbf{Q}_\Theta^{-1})$ instead of the UNet prior formulation
- The inner variational cost in the 4DVarNet scheme becomes:

$$\begin{aligned}\mathbf{x}^* &= \operatorname{argmin}_{\mathbf{x}} \mathcal{J}(\mathbf{x}, \mathbf{y}, \Omega) \\ &= \operatorname{argmin}_{\mathbf{x}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_{\Omega}^2 + \left[\mathbf{x} - \mu^b \right]^T \mathbf{Q}_\Theta \left[\mathbf{x} - \mu^b \right]\end{aligned}$$

- If the true state \mathbf{x} is also Gaussian, then OI/kriging/Blue estimator is optimal.
- 4DVarNet should asymptotically reach the OI

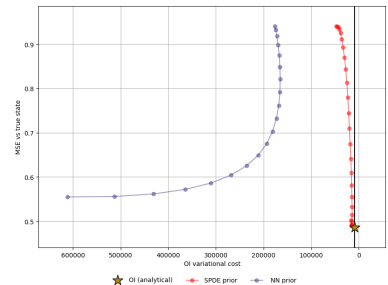


Figure 6: Asymptotic convergence of 4DVarNet to OI (Gaussian case)

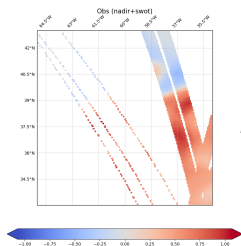
Towards a stochastic neural variational scheme: playing with the prior distribution

Stochastic neural variational scheme

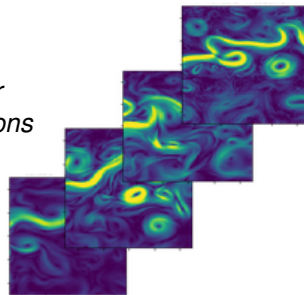
- 1 Key idea: sample members $\mathbf{x}^{S,i}$ in the prior distribution
- 2 Posterior pdf: use the neural solver for conditioning the prior:

$$\mathbf{x}^{*,i} = \mathbf{x}^* + \{\mathbf{x}^{S,i} - \mathbf{x}^{*,S,i}\}$$

How to generate ensemble members from "a" prior distribution?



*How to sample in the prior
based on partial observations*



Towards a stochastic 4DVarNet: SPDE-based generative modelling (I)

Could we try to generate ensemble members from a prior distribution?

- 1 Geostatistical-based idea, use a GP approximation of prior operator Φ in 4DVarNet, driven by stochastic PDEs (SPDE):

$$\mathcal{F}_{\Theta} \mathbf{x}^b = \mathbf{z}$$

- 2 \mathcal{F}_{Θ} a fractional differential operator (embedding advection, diffusion, etc. - all SPDE parameters being grouped in Θ) and \mathbf{z} a white/colored noise
- 3 Why doing that: get back to interpretable DA scheme using SPDE as a surrogate prior model along the DAW:

$$\frac{\partial \mathbf{x}^b}{\partial t} + \left\{ \kappa^2(\mathbf{s}, t) + \mathbf{m}(\mathbf{s}, t) \cdot \nabla - \nabla \cdot \mathbf{H}(\mathbf{s}, t) \nabla \right\}^{\alpha/2} \mathbf{x}^b(\mathbf{s}, t) = \tau \mathbf{z}(\mathbf{s}, t)$$

Using the FDM:

$$\mathbf{x}_{t+dt} = \mathbf{M}_{t+dt} \mathbf{x}_t + \mathbf{T}_{t+dt} \mathbf{z}_t$$

$\mathbf{M}_{t+dt} = (\mathbf{I} + dt \mathbf{B}_{t+dt})^{-1}$ denotes the matrix operator that emulates the SPDE dynamical evolution of prior state \mathbf{x}^b from time t to $t + dt$.

Towards a stochastic 4DVarNet: SPDE-based generative modelling (II)

We know how to compute the precision matrix \mathbf{Q} of prior state sequences $\{\mathbf{x}_0^b, \dots, \mathbf{x}_{Ldt}^b\}$, where $\mathbf{x}_0^b \sim \mathcal{N}(\mu_0, \mathbf{P}_0)$ denotes the initial state. By denoting $\mathbf{S}_k = \mathbf{T}_k \mathbf{T}_k^T$, we have

$$\mathbf{Q} = \begin{bmatrix} \mathbf{P}_0^{-1} + \mathbf{M}_1^T \mathbf{S}_1^{-1} \mathbf{M}_1 & -\mathbf{M}_1^T \mathbf{S}_1^{-1} & 0 & 0 & 0 & \dots & 0 \\ -\mathbf{S}_1^{-1} \mathbf{M}_1 & \mathbf{S}_1^{-1} + \mathbf{M}_2^T \mathbf{S}_2^{-1} \mathbf{M}_2 & -\mathbf{M}_2^T \mathbf{S}_2^{-1} & 0 & 0 & \dots & 0 \\ 0 & -\mathbf{S}_2^{-1} \mathbf{M}_2 & \mathbf{S}_2^{-1} + \mathbf{M}_3^T \mathbf{S}_3^{-1} \mathbf{M}_3 & -\mathbf{M}_3^T \mathbf{S}_3^{-1} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & -\mathbf{S}_{L-1}^{-1} \mathbf{M}_{L-1} & \mathbf{S}_{L-1}^{-1} + \mathbf{M}_L^T \mathbf{S}_L^{-1} \mathbf{M}_L - \mathbf{M}_L^T \mathbf{S}_L^{-1} \\ 0 & \ddots & \ddots & \ddots & 0 & -\mathbf{S}_L^{-1} \mathbf{M}_L & \mathbf{S}_L^{-1} \end{bmatrix}$$

Towards a stochastic 4DVarNet: SPDE-based generative modelling (III)

- 4DVarNet is trained with an augmented state formulation $\{\mathbf{x}, \Theta\}$ with Θ the (non-stationary) SPDE parameters of prior \mathbf{x}^b
- Joint learning of SPDE parametrization Θ and solver Γ :

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

- The (supervised) training loss function is:

$$\mathcal{L}(\mathbf{x}, \Theta^*, \mathbf{x}^*) = \underbrace{\lambda_1 \mathcal{L}_1(\mathbf{x}, \mathbf{x}^*)}_{\text{reconstruction cost}} + \underbrace{\lambda_2 \mathcal{L}_2(\mathbf{x}, \Theta^*)}_{\text{prior regularization cost}}$$

where:

- 1 $\mathcal{L}_1(\mathbf{x}, \mathbf{x}^*) = \|\mathbf{x} - \mathbf{x}^*\|^2$ the MSE w.r.t Ground Truth (**reconstruction cost**)
- 2 $\mathcal{L}_2(\Theta^*|\mathbf{x})$ the negative log-likelihood $-\mathcal{L}(\Theta^*|\mathbf{x})$ (**prior regularization cost**):

$$\mathcal{L}(\Theta^*|\mathbf{x}) = -|\mathbf{Q}(\Theta^*)| + \left[\mathbf{x} - \mu^b \right]^T \mathbf{Q}_{\Theta^*} \left[\mathbf{x} - \mu^b \right]$$

4DVarNet-SSH: the data assimilation window

- The application used to focus on **interpolation/reconstruction**: we aggregate data assimilation windows of 9 days (a sample) to reconstruct the day at the center of the window
- Let precise that the exact same scheme can be trained also for **near-real-time (NRT)** or **short-term forecasting** tasks

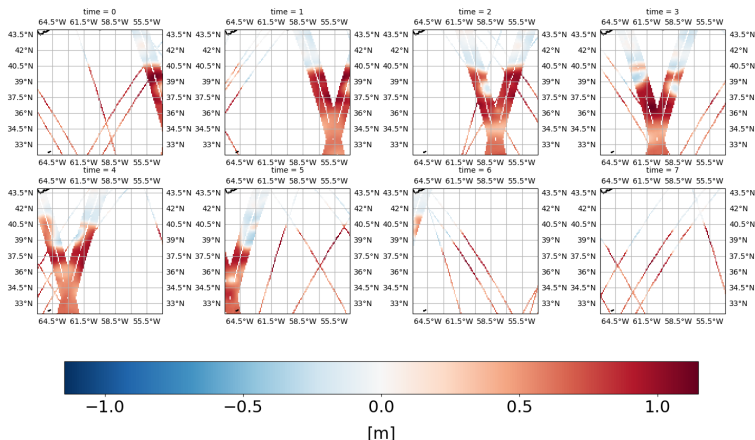


Figure 7: An example of the 9-day data assimilation window (DAW) used in the current setup. The target reconstruction relies to interpolation product, then the day at the center of the window.

4DVarNet-SSH: illustrating the iterative solver

The LSTM solver: how does it work?

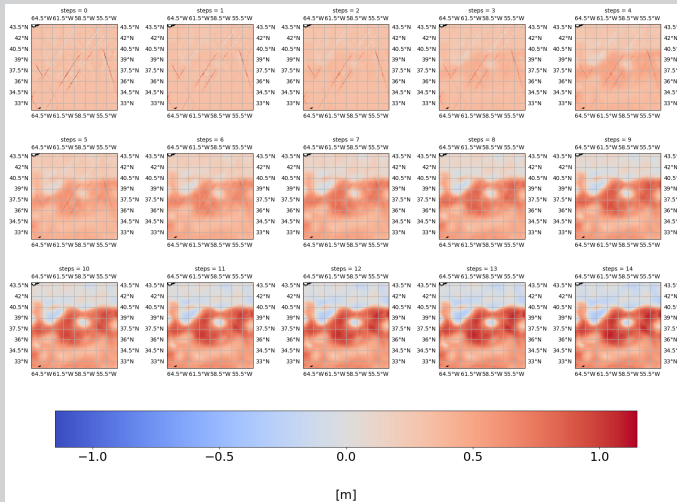


Figure 8: An example of the 10-step LSTM-based iterative solver: at step 0, the state x is only filled with observations (when available) and 0 elsewhere. From step 1 to 9, the solver has learnt how to fill in the original gaps based on the previous observations to comply with training loss \mathcal{L}

4DVarNet-SSH: Prior estimation and simulation (I)

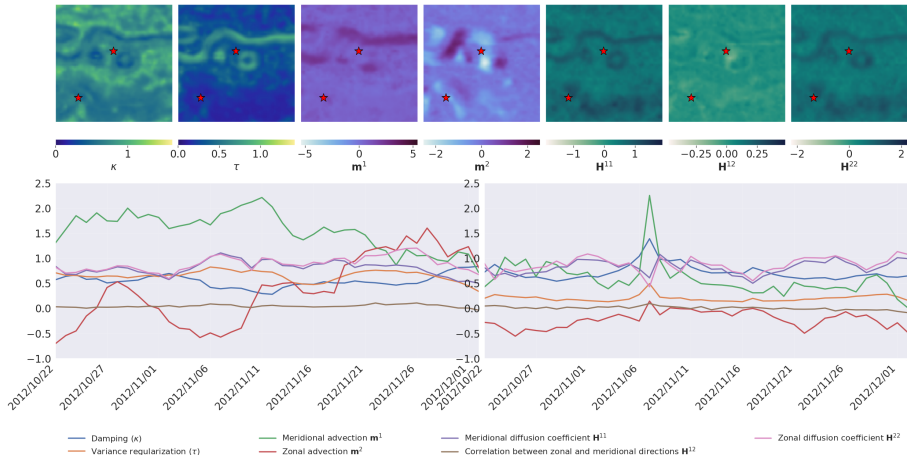


Figure 9: SPDE parameter estimations: [top] first day of the test period; [bottom] time series along the GF meandrum (left) and in the less energetic left-lower part of the domain (right).

4DVarNet-SSH: Prior estimation and simulation (II)

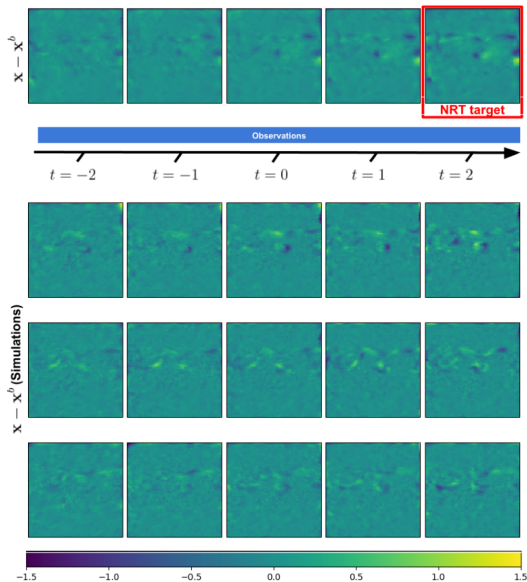


Figure 10: An example of the anomaly $\mathbf{x} - \mathbf{x}^b$ for near-real-time (NRT) application along the data assimilation window and three simulations after training of the SPDE parameters

4DVarNet-SSH: Posterior estimation and uncertainty quantification

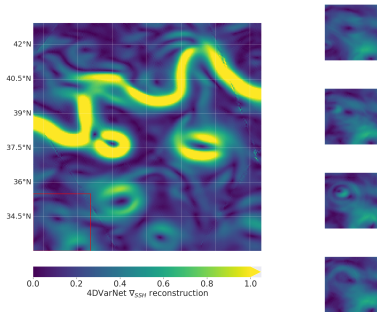


Figure 11: SSH gradient of the 4DVarNet daily reconstruction on the GF domain and the related focus on the bottom-left 50x50 pixels subdomain (red box) where the difference amongst 4 members are shown

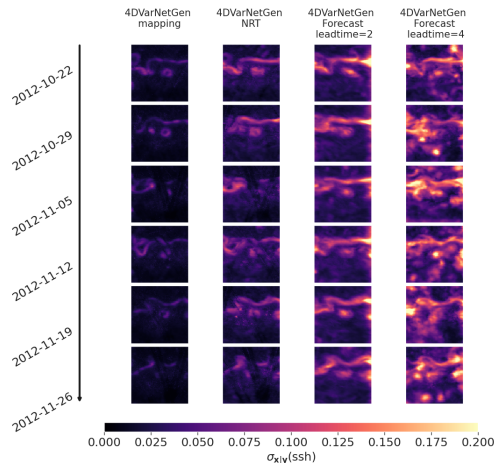


Figure 12: Standard deviations of 200 4DVarNet-Gen conditional simulations for mapping, NRT, and forecast (lead times 2 and 4) configurations along the test period.

Take-home messages

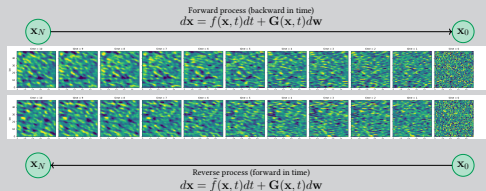
Neural variational scheme: current state

We can bridge DNN and variational models to solve inverse problems.

- **Key tools:** unrolling of the gradient-step descent with 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 (I):** UQ of 4DVarNet schemes based on the conditioning of large SPDE-based prior ensemble
- **Stochastic implementation (II):** Link to diffusion models for more general SDEs:

$$dx = \mathbf{f}(\mathbf{x}, t)dt + \mathbf{G}(\mathbf{x}, t)d\mathbf{w}$$

with $\mathbf{f}(\cdot, t) : \mathbb{R}^m \rightarrow \mathbb{R}^m$, $\mathbf{g}(\cdot, t) : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times m}$ and $\mathbf{w}(\mathbf{s}, t)$ is a Wiener process.



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, Q. Febvre, J. Thompson, H. Georgenthum, and R. Fablet. Learning neural optimal interpolation models and solvers. In J. Mikyška, C. de Mulatier, M. Paszynski, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. Sloot, editors, *Computational Science – ICCS 2023*, pages 367–381, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-36027-5.
- 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

- 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>.
- 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.