



HAL
open science

Ensemble-based 4DVarNet uncertainty quantification for the reconstruction of sea surface height dynamics

Maxime Beauchamp, Quentin Febvre, Ronan Fablet

► **To cite this version:**

Maxime Beauchamp, Quentin Febvre, Ronan Fablet. Ensemble-based 4DVarNet uncertainty quantification for the reconstruction of sea surface height dynamics. *Environmental Data Science*, 2023, 2, pp.e18. 10.1017/eds.2023.19 . hal-04147805

HAL Id: hal-04147805

<https://imt-atlantique.hal.science/hal-04147805>

Submitted on 1 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.


L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

METHODS PAPER

Ensemble-based 4DVarNet uncertainty quantification for the reconstruction of sea surface height dynamics

Maxime Beauchamp , Quentin Febvre and Ronan Fablet

Lab-STICC, IMT Atlantique, Brest, France

Corresponding author: Maxime Beauchamp; Email: maxime.beauchamp@imt-atlantique.fr

Received: 08 May 2023; **Accepted:** 31 May 2023

Keywords: data assimilation; end-to-end learning; ensemble-based method; simulation; uncertainty quantification

Abstract

Uncertainty quantification (UQ) plays a crucial role in data assimilation (DA) since it impacts both the quality of the reconstruction and near-future forecast. However, traditional UQ approaches are often limited in their ability to handle complex datasets and may have a large computational cost. In this paper, we present a new ensemble-based approach to extend the 4DVarNet framework, an end-to-end deep learning scheme backbone on variational DA used to estimate the mean of the state along a given DA window. We use conditional 4DVarNet simulations compliant with the available observations to estimate the 4DVarNet probability density function. Our approach enables to combine both the efficiency of 4DVarNet in terms of computational cost and validation performance with a fast and memory-saving Monte-Carlo based post-processing of the reconstruction, leading to the so-called En4DVarNet estimation of the state pdf. We demonstrate our approach in a case study involving the sea surface height: 4DVarNet is pretrained on an idealized Observation System Simulation Experiment (OSSE), then used on real-world dataset (OSE). The sampling of independent realizations of the state is made among the catalogue of model-based data used during training. To illustrate our approach, we use a nadir altimeter constellation in January 2017 and show how the uncertainties retrieved by combining 4DVarNet with the statistical properties of the training dataset lead to a relevant information providing in most cases a confidence interval compliant with the Cryosat-2 nadir alongtrack dataset kept for validation.

Impact Statement

This research paper discusses the extension of the 4DVarNet framework to ensemble-based approaches. It paves the way to uncertainty quantification and enables to initiate the discussion of how to retrieve posterior pdf from such deep learning framework. This paper also involves pieces of previous work initiated inside the 4DVarNet team of developers in the framework of AI Chair Oceanix. It is meant to help the readers in understanding this framework from a broader point of view while diving into some new and ongoing developments.

1. Introduction

In state-of-the-art data assimilation (DA) systems, quantifying the estimation uncertainty is a key aspect that must be addressed all together with the mean of the state space. One major issue is that uncertainty quantification (UQ) is computationally costly when it comes with variational or ensemble-based approaches and their combination (Evensen et al., 2022). Recently, we proposed an end-to-end deep learning (DL) scheme called 4DVarNet drawing from variational DA to learn jointly priors and solvers.

This approach was successfully applied on the reconstruction of sea surface state based on altimetric datasets (Beauchamp et al., 2022a) in the context of both traditional nadir altimeters and the upcoming SWOT mission. Though, 4DVarNet currently only provides a reconstruction of the state space without quantifying the related uncertainties.

In this work, we draw from geostatistical tools to easily generate ensemble of 4DVarNet reconstructions. The main advantage of our solution is that the ensemble generation is provided as a post-processing of 4DVarNet, which enables to simulate the pdf of the state space based on a lot of members (typically a hundred). It does not require any additional training that might come with a high and potentially untractable computational cost if the optimization scheme involves a large number of ensemble members.

2. Materials

2.1. 4DVarNet: Basic concepts

Recently, new neural-network-based architectures have emerged to solve inverse problems based on variational formulation. More precisely, restarting from the variational DA scheme (Asch et al., 2016), the state analysis \mathbf{x}^* is obtained by solving the minimization problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{arg\,min}} \mathcal{J}(\mathbf{x}),$$

where the variational cost function $\mathcal{J}(\mathbf{x}) = \mathcal{J}_\Phi(\mathbf{x}, \mathbf{y}, \Omega)$ is the sum of an observation term and a regularization term involving a dynamical prior operator Φ :

$$\begin{aligned} \mathcal{J}_\Phi(\mathbf{x}, \mathbf{y}, \Omega) &= \mathcal{J}^o(\mathbf{x}, \mathbf{y}, \Omega) + \mathcal{J}_\Phi^b(\mathbf{x}) \\ &= \lambda_1 \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|_\Omega^2 + \lambda_2 \|\mathbf{x} - \Phi(\mathbf{x})\|^2 \end{aligned} \tag{1}$$

and \mathbf{y} denotes the observations available on subdomain $\Omega \subset D$.

Here, the parameterization of operator Φ is based on convolutional neural networks (CNN) embedding Markovian priors, see, for example, Fablet et al. (2020). DL automatic differentiation tools are involved for the computation of the gradient operator $\nabla_{\mathbf{x}} \mathcal{J}_\Phi$ given the architecture of operator Φ and is also a composition of operators such as tensors, convolutions, and activation functions. The proposed end-to-end architecture consists in embedding an iterative gradient-based solver Γ based on a specific variational representation. Following meta-learning schemes (Andrychowicz et al., 2016), a residual LSTM-based representation (Fablet et al., 2020) of Γ is considered.

Formally, we define $\Theta = \{\Theta_\Phi, \Theta_\Gamma\}$ the set of parameters to train and we state this training of the considered neural scheme as a bilevel optimization problem:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{arg\,min}} L(\mathbf{x}, \mathbf{y}, \Omega, \mathbf{x}^*) \quad \text{s.t. } \mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{arg\,min}} \mathcal{J}_\Phi(\mathbf{x}, \mathbf{y}, \Omega), \tag{2}$$

where $\mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{x}^*)$ defines a training loss, typically the mean squared error (MSE) w.r.t the Ground Truth with additional regularization terms.

Overall, let denote by $\Psi_{\Phi, \Gamma}(\mathbf{x}^{(0)}, \mathbf{y}, \Omega)$ the output of the so-called 4DVarNet end-to-end learning scheme, see Figure 1 and Algorithm 1, given architectures for both NN-based operators Φ and Γ , the initialization $\mathbf{x}^{(0)}$ of state \mathbf{x} and the observations \mathbf{y} on domain Ω , subdomain of \mathcal{D} where observations are available.

Algorithm 1. 4DVarNet algorithm Fablet et al. (2020)

Data:

- $\mathbf{x} \in \mathbb{R}^{T \times m} = \{\mathbf{x}_k\}, k = 1, \dots, T$
- $\mathbf{y}_\Omega = \{\mathbf{y}_{k, \Omega_k}\}, k = 1, \dots, T$: observations on domains $\Omega_k \subset \mathcal{D}$
- N_I : number of iterations
- η : gradient step

List of procedures:

– Train $\Psi_{\Phi, \Gamma}$: end-to-end learning procedure with:
 * Φ : NN-based representation of the dynamical system
 * GradLSTM: residual NN-based representation of $\nabla_{\mathbf{x}} \mathcal{J}(\mathbf{x})$
 * Γ : iterative gradient-based update operator
 $\mathbf{x} = \Phi(\mathbf{y})$
while $i < N_I$ **do**
 $\mathbf{x}^{(i+1)} \leftarrow \mathbf{x}^{(i)} - \eta \times \text{GradLSTM}(\mathbf{x}^{(i)})$
 $N_I \nearrow; \eta \searrow; i \leftarrow i + 1$
end
 Train $\Psi_{\Phi, \Gamma}$
Result: $\mathbf{x}^* \leftarrow \Psi_{\Phi, \Gamma}(\mathbf{x}^{(0)}, \mathbf{y}, \Omega)$

Whereas classic gradient-based iterative methods may suffer from a relatively low convergence rate, which dramatically worsens when dealing with high-dimensional states, the proposed trainable solver Γ considerably speed up the convergence and reach good interpolation performance with only 10–100 gradient steps, depending on the complexity of underlying dynamical process. Regarding the convolutional

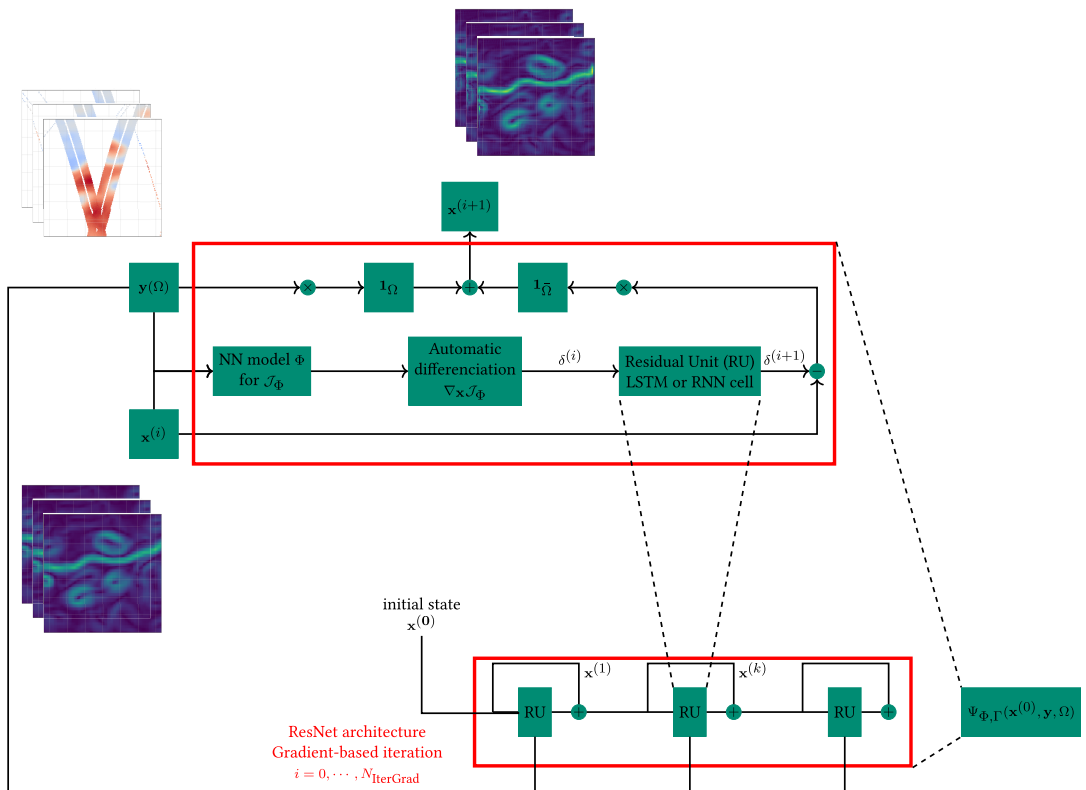


Figure 1. Sketch of the gradient-based algorithm: the upper-left stack of images corresponds to an example of SSH observations temporal sequence with missing data used as inputs. The upper-right stack of images is an example of intermediate reconstruction of the SSH gradient at iteration i while the bottom-left stack of images identifies the updated reconstruction fields used as new inputs after each iteration of the algorithm.

architecture of both trainable prior and solver, it also guarantees their scalability and linear complexity w.r.t. the size of the spatial domain as well as the number of observations. Such a framework also allows to explore different configurations for the inner variational cost \mathcal{J}_ϕ and the outer training loss \mathcal{L} .

2.2. Toward a stochastic formulation

Because 4DVarNet draws from a DA state space formalism, it is easy to consider this DL model under a stochastic formulation. Indeed, most operational DA methods are based on Gaussian assumptions and simply estimate the two first moments of the posterior distribution, mean \mathbf{x}^* and covariance \mathbf{P}^* . We suppose that for a given time t and related DA window $t-L:t+L$ of length $2L+1$, we know how to simulate N nonconditional simulations $\mathbf{x}_{t-L:t+L}^{s,i}$, $i = 1, \dots, N$ whose compatibility of their statistical properties, typically the mean and covariance for a Gaussian process (GP), with the true state $\mathbf{x}_{t-L:t+L}$ is verified. By denoting $\mathbf{u} \in \mathcal{D}$ a spatiotemporal point in the state vector $\mathbf{x} = [x(\mathbf{u})] \in \mathbb{R}^m$, we have the following set of (nonstationary) point-wise properties:

$$1. E[\mathbf{x}_{t-L:t+L}^{s,i}(\mathbf{u})] = \mathbf{m}(\mathbf{u}), \tag{3a}$$

$$2. \text{var}[\mathbf{x}_{t-L:t+L}^{s,i}(\mathbf{u})] = \sigma_x^2(\mathbf{u}), \tag{3b}$$

$$3. C[\mathbf{x}_{t-L:t+L}^{s,i}(\mathbf{u}), \mathbf{x}_{t-L:t+L}^{s,i}(\mathbf{u}')] = C[\mathbf{x}_{t-L:t+L}(\mathbf{u}), \mathbf{x}_{t-L:t+L}(\mathbf{u}')]. \tag{3c}$$

In what follows, we simplify the notations and $\mathbf{x}_{t-L:t+L}$ writes \mathbf{x} and $C[\mathbf{x}(\mathbf{u}), \mathbf{x}(\mathbf{u}')] = C(\mathbf{u}, \mathbf{u}')$. Then, the previous set of properties translates into matrix notation with state-space DA formalism as $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{P})$ with $\mathbf{P} = [C(\mathbf{u}, \mathbf{u}')]$ which is nothing else than the prior distribution involved in the Bayesian estimation making the link between all DA approaches.

By “nonconditional,” we mean that these simulations do not comply with observations $\mathbf{y}_{t-L:t+L}$. To make these simulations compliant with the observations, we draw from traditional geostatistics and kriging-based conditioning (Wackernagel, 2003) and replace the prior model and kriging solver, also known as Optimal Interpolation (OI) or BLUE (Best Linear Unbiased Estimator) in the DA community, see, for example, Asch et al. (2016), by our 4DVarNet estimation. A conditional simulation $\mathbf{x}^{*,i}$ writes:

$$\mathbf{x}^{*,i} = \mathbf{x}^* + \{\mathbf{x}^{s,i} - \mathbf{x}^{*,s,i}\}, \tag{4}$$

where \mathbf{x}^* is the 4DVarNet reconstruction, $\mathbf{x}^{s,i}$ is a nonconditional simulation of the underlying process with similar statistical properties than \mathbf{x} and $\mathbf{x}^{*,s,i}$ is the 4DVarNet reconstruction of $\mathbf{x}^{s,i}$ using as sparse observations a subsampling of the simulation with same mask Ω than true observations.

In absence of observations errors, and because it is backbone on variational DA, we proved, see Beauchamp et al. (2022b), that 4DVarNet applied to the GP Optimal Interpolation with:

- inner variational cost being the OI variational cost, and
- outer 4DVarNet training loss function being the MSE loss w.r.t the true state when the latter is available in a supervised setting, or the same OI variational cost for unsupervised configurations,

complies in the asymptotic limit with optimal kriging state $\mathbf{x}^* \rightarrow \mathbf{x}^{\text{OI}}$ and its related properties: nonbias and optimal variance.

In this GP formalism, it implies that conditioning a simulation by the observations based on equation (4) have the following set of point-wise properties, drawn from conditional Gaussian distributions theory, see, for example, Chilès and Delfiner (2012):

$$1. E[\mathbf{x}^{*,i}(\mathbf{u})] \rightarrow E[\mathbf{x}^*(\mathbf{u})], \tag{5a}$$

$$2. \text{var}[\mathbf{x}^{*,i}(\mathbf{u})] \rightarrow \text{var}[\mathbf{x}^*(\mathbf{u})], \tag{5b}$$

$$3.C[\mathbf{x}^{*,i}(\mathbf{u}), \mathbf{x}^{*,i}(\mathbf{u}')] \rightarrow C(\mathbf{u}, \mathbf{u}') - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}(\mathbf{u}) \lambda_{\beta}(\mathbf{u}') C(\mathbf{u}_{\alpha}, \mathbf{u}_{\beta}), \tag{5c}$$

where $\alpha, \beta = 1, \dots, n$ denotes the observation index and λ_{α} are the optimal kriging weights in the estimation $\mathbf{x}^{OI}(\mathbf{u})$. Making the link with state-space DA formalism, equation (5a) becomes:

$$\mathbf{P}^* = \Sigma[\mathbf{x}^i|\mathbf{y}] = (\mathbf{I} - \mathbf{KH})\mathbf{P}, \tag{6}$$

where we recognize the posterior covariance formulation with $\mathbf{K} = [\lambda_{\alpha}(\mathbf{u})]_{\alpha=1, \dots, n} \in \mathbf{R}^{n \times m}$ and \mathbf{H} the observation operator mapping the state to the observations. $\mathbf{u} \in \mathcal{D}$

Then, the covariance between two conditionally simulated locations \mathbf{u} and \mathbf{u}' is less than the prior covariance which makes sense because conditioning a process with observations shall constrain the uncertainty.

The latter set of properties is verified when a GP approximation of state \mathbf{x} is valid. When the underlying covariance model is not known and/or a GP approximation with linear dynamics does not apply, it was shown that 4DVarNet estimation (Fablet et al., 2021; Beauchamp et al., 2022a) enables to improve the OI interpolation, that is, it is still unbiased but with lower MSE w.r.t the ground truth, that is, lower variance of its error $\text{var}[\mathbf{x}^*(\mathbf{u})] < \text{var}[\mathbf{x}^{OI}(\mathbf{u})]$. Consequently, running an ensemble of N simulations conditioned by 4DVarNet will provide an approximation of the probability distribution function $p_{\mathbf{x}|\mathbf{y}}$ of state $\mathbf{x}|\mathbf{y}$ with both improvements on the two first moments \mathbf{x}^* and \mathbf{P}^* compared to simulations conditioned by kriging. The ensemble mean $\overline{\mathbf{x}^{*,i}}$ will be the 4DVarNet reconstruction in the limits of $N \rightarrow +\infty$:

$$\overline{\mathbf{x}^{*,i}} = \frac{1}{N} \sum_i \mathbf{x}^{*,i} \xrightarrow{N \rightarrow +\infty} \mathbf{x}^*.$$

We stress that such a formulation is consistent with traditional DA: indeed when looking at the spatial sequential scheme of the Ensemble Kalman Filter (Evensen, 2009), the forecast step can be seen as nonconditional simulations at time $t - 1$ generating N realizations for time t . The average forecast and the covariance matrix are then computed on this set of realizations. and the analysis step corresponds to the conditioning of these realizations. Here, our approach follows the same idea but we do not address the problem sequentially and produce spatio-temporal nonconditional simulations over the assimilation window, before conditioning them with the observations. Such a formulation is more suitable to produce ensemble of 4DVarNet realizations. In the following, we refer to our approach as En4DVarNet. We believe this formulation is the more appropriate amongst the set of hybrid methods which aims to incorporate the statistical flow-dependent strength of EnKF into 4DVar. Indeed, when comparing this approach with En4DVar (Zhang et al., 2009), our method share some ideas for both prior and solvers:

- the minimization procedure for both En4DVar and En4DVarNet are still backboneed on their original 4DVar and 4DVarNet optimization scheme, respectively based on the adjoint model and the neural prior encoder Φ . It is not always the case for other hybrid methods, see, for example, the class of 4DVar methods (Qiu et al., 2007) which avoid the adjoint model by approximating the linearized tangent operator with the simulated observation perturbations;
- En4DVar retrieves the statistical ensemble background covariance from a parallel EnKF implementation, which allows for flow-dependency of the latter. In En4DVarNet, because we sampled spatiotemporal simulation with similar properties than the true state along the DA window, their associated prior mean and covariance matrix may differ from time to time. We precise that our approach, though combining ensemble-based formulation with OI-based conditioning should not be confused with the so-called EnOI approach in DA, see, for example, Counillon and Bertino (2009)

and Asch et al. (2016), which is just a EnKF in which the flow-dependent error covariance matrix is replaced by a stationary matrix calculated from an historical ensemble.

Also, as mentioned above, the generic presentation of En4DVarNet supposes to be able to sample in the prior distribution. In practice, this is not an easy task when no model is available, see Section 2.3 for a practical solution. When not possible, future works may consider the DL problem of joint learning for both the reconstruction and prior emulation based on neural or differentiable formulations of the latter, see, for example, differentiable physical quasi-geostrophic model (Frezat et al., 2022) or SPDE-based linear approximations (Beauchamp et al., 2022b). Such additional developments would enable to incorporate the ensemble statistics into the 4DVarNet outer training loss, in order to comply with maximum likelihood criteria w.r.t to the Ground Truth (supervised setting) or the observations (unsupervised setting).

In the end, such a name for the proposed approach leaves space for additional flavors of ensemble-based 4DVarNets in the future, as it is done with a combination of ensemble and variational approaches in DA.

2.3. Practical aspects

The key aspect of our approach lies in the capability of providing a prior model from which we are able to sample realizations of the underlying process. In the current 4DVarNet formulation, the operator Φ plays the role of autoencoding the state and is learnt jointly with the solver to satisfy at most the loss function used in the 4DVarNet training process. Such an operator should not be seen as an emulator able to generate simulations of a geophysical variable, like sea surface height (SSH), as used in Section 3. For now, because the main 4DVarNet applications were developed in a supervised setting, we propose to sample the realizations in the catalogue of the model-based experiments used during training to estimate the pdf of the 4DVarNet reconstructions applied on real-world datasets. Such a sampling will require to retrieve the analogs (Lguensat et al., 2017) of the state for a given DA window, which is done by comparing the low-resolution properties of the available observations to the model-based catalogue. This is why the En4DVarNet explored in this paper shall be seen as a post-processing as the traditional 4DVarNet method.

Algorithm 2. Ensemble-based 4DVarNet (En4DVarNet) algorithm

Data:

$$\mathbf{y}(\Omega)_{t-L \rightarrow t+L}$$

Ω , subdomain of space \mathcal{D} with observations

\mathbf{x}_T , model-based 4DVarNet training dataset, also use as catalogue for post-processing

List of procedures:

$\Psi_{\Phi, \Gamma}(\mathbf{y}(\Omega))$: 4DVarNet reconstruction

Φ : CNN-based representation of the dynamical system

Γ : iterative gradient-based update operator with LSTM-based representation of $\nabla_{\mathbf{x}} \mathcal{J}(\mathbf{x})$

knn($\mathbf{y}(\Omega), \mathbf{x}_T, k$): Get the k th nearest neighbor (analog) of \mathbf{y} in the training dataset

Init:

$$\mathbf{x}^{*,(0)} = \begin{cases} \mathbf{y} \text{ over } \Omega \\ 0 \text{ elsewhere} \end{cases}$$

for $i \in 1, \dots, N$ **do**

$$\mathbf{x}^* = \Psi_{\Phi, \Gamma}(\mathbf{y}(\Omega))$$

$$\mathbf{x}^{s,i} = \text{knn}(\mathbf{y}(\Omega), \mathbf{x}_T, i)$$

$$\mathbf{x}^{*,s,i} = \Psi_{\Phi, \Gamma}(\mathbf{x}^{s,i}(\Omega))$$

end

Result: $\mathbf{x}^{*,i} = \mathbf{x}^* + (\mathbf{x}^{s,i} - \mathbf{x}^{*,s,i})$

Algorithm 2 details the entire procedure to derive the ensemble of 4DVarNet reconstructions based on the analog sampling of independent realizations in the training dataset.

3. Real-World SSH Application

To evaluate our approach, we use a real-world nadir altimetric dataset in January 2017. The SSH observations include a constellation of seven nadir altimeters: SARAL/Altika, Jason 2, Jason 3, Sentinel 3A, Haiyang-2A, and Cryosat-2. Note that the Cryosat-2 altimeter dataset is not included in the mapping reconstruction, it is kept to perform an independent validation of both reconstruction and UQ.

In the previous 4DVarNet SSH-related publications, we use the NATL60 OSSE (Observation System Simulation experiment) experiment ran with the state-of-the-art NEMO model (Nucleus for European Modeling of the Ocean) at high resolution ($1/60^\circ$) over the whole North Atlantic (NATL) basin to train 4DVarNet (Beauchamp et al., 2022a). We used a supervised configuration where the target variable is defined as the anomaly between the Ground Truth and DUACS (Data Unification and Altimeter Combination System) OI (Taburet et al., 2019). Recent applications of 4DVarNet on the raw SSH, without providing the DUACS OI baseline as a coarse version of the reconstruction, showed that this flavor improves again the performance when using as length for the DA window ~ 15 days. For our experiment, we use this 4DVarNet pretrained model to reconstruct SSH over the same $10^\circ \times 10^\circ$ Gulf-Stream domain used during the OSSE-based training.

3.1. Results

In the following experiment, we apply our ensemble-based 4DVarNet approach with 60 members on the 6 nadir constellation on January 2017. Figure 2 displays an example of the 4DVarNet reconstruction on January 10, 2017 as a mean state and we extract the time series of three spatial positions P1, P2, and P3,

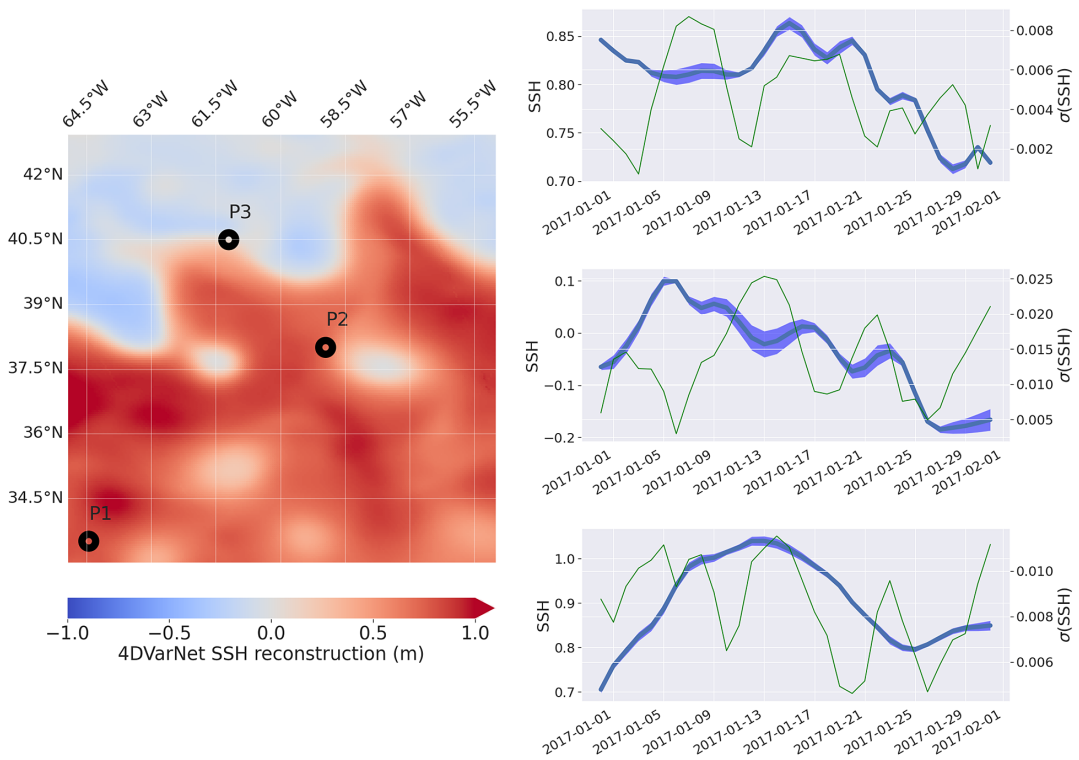


Figure 2. Example of 4DVarNet daily reconstruction on the GulfStream domain (January 10, 2017) based on the six nadirs aggregation dataset. Black circles indicate spatial positions P1, P2, and P3 where ensemble-based 4DVarNet mean and spreads are extracted.

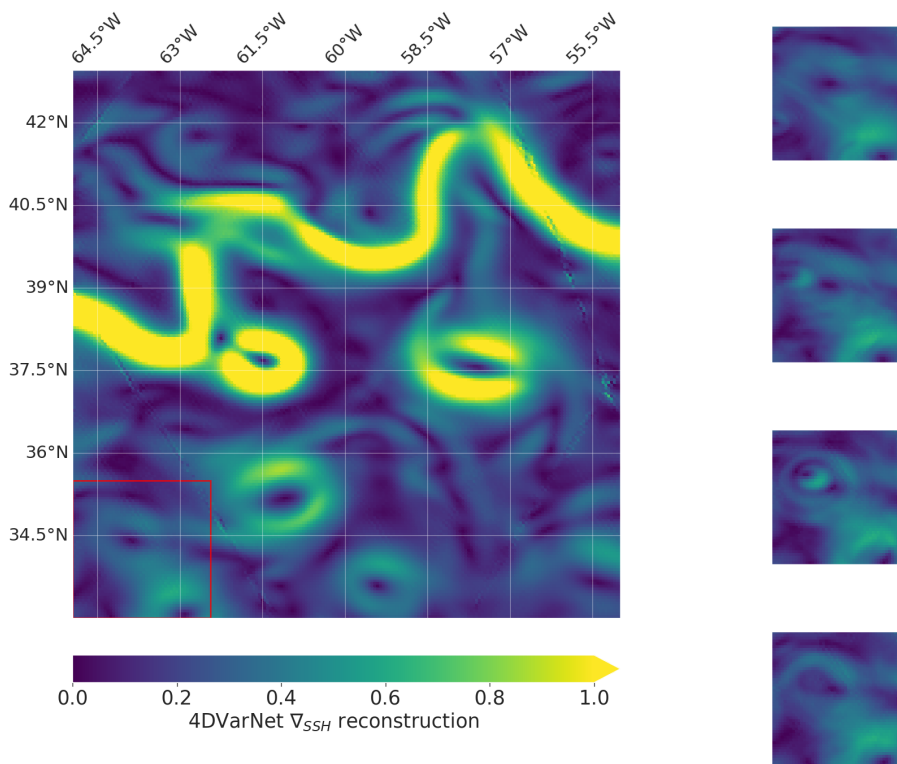


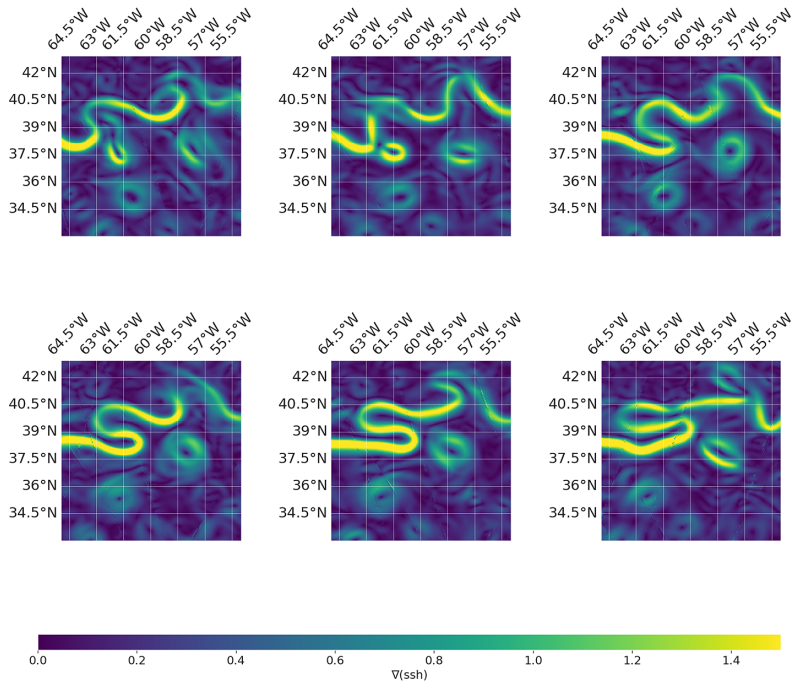
Figure 3. SSH gradient of the 4DVarNet daily reconstruction on the GulfStream domain (January 4, 2017) and the related focus on the bottom-left 50×50 pixels subdomain (red box) where the difference amongst four members are shown.

marked by black circles. P1 is located in a region less energetic than the rest of the GulfStream domain, P2 lies along the main meander and P3 is away from the latter, but it stays close to energetic eddies that may appear from time to time. The corresponding time series of ensemble mean and ensemble spread are provided on the right-hand side of Figure 2: we can see how depending on the position, the variability among the members may differ. Interestingly, a periodic pattern seems to reproduce for all the positions selected which indicates the strong weight of the observation nadir sampling on the 4DVarNet uncertainty.

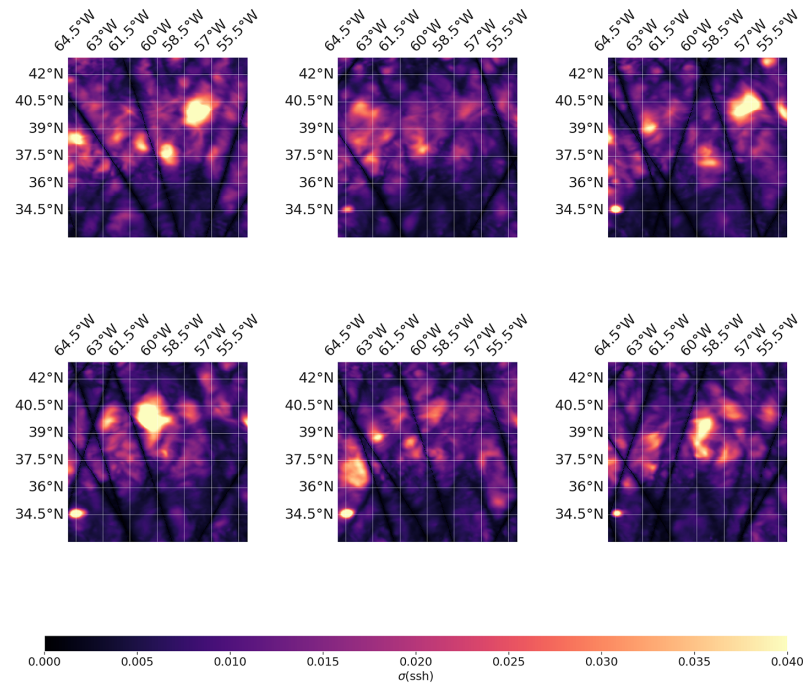
As an illustration of the En4DVarNet variability, Figure 3 displays the SSH gradient field of the 4DVarNet reconstruction (January 4, 2021): we focus on a small bottom-left 50×50 pixels subdomain to show how members may be able to reproduce different realistic patterns and small eddy structures while maintaining the members compliant with the observations along the DA window.

In the same vein, Figure 4a,b respectively displays the SSH gradient and corresponding spreads from December 31, 2016 to January 25, 2017 (every 5 days). We can see how the main GulfStream meander and smaller eddies evolve over time. It is clear in Figure 4b how the observations help in decaying the reconstruction uncertainty. Then, when moving away from the observations at the center of the assimilation window, the uncertainty grows quickly, not only based on the distance from the nadir altimeters but also influenced by the SSH spatiotemporal dynamics.

Figure 5 enables to understand the evolution of the 4DVarNet uncertainty over time. We provide the full standard deviation maps for $t = 0$ (December 31, 2016) and only displays the highest uncertainties levels along z -axis for the next 30 dates over the test period. While the observation sampling evolves every day, it is clear that the ensemble approach identifies areas of the GulfStream domain with persistently high



(a) 4DVarNet SSH gradient



(b) 4DVarNet SSH standard deviations

Figure 4. 4DVarNet SSH gradient and the corresponding ensemble-based standard deviations from December 31, 2016 to January 25, 2017 (every 5 days).

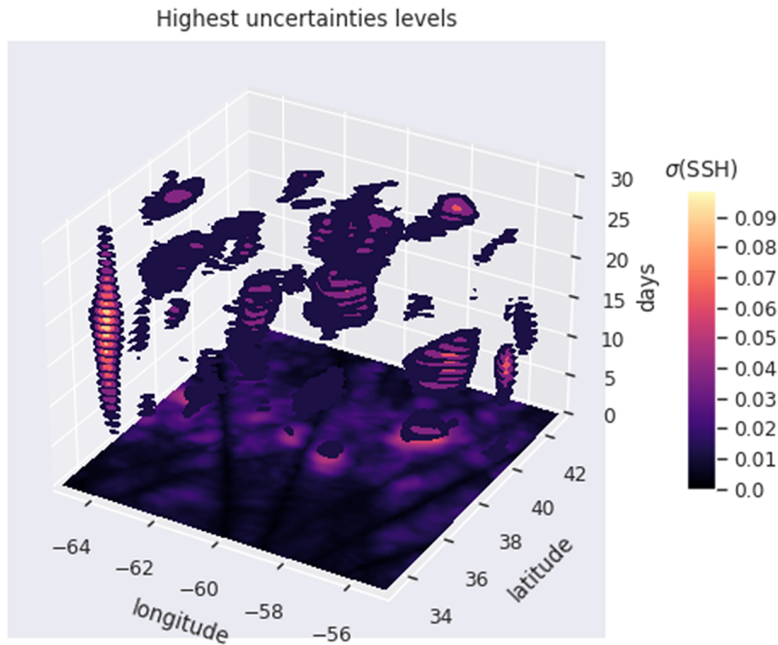


Figure 5. Ensemble-based 4DVarNet standard deviations over the test period (January 2017): the full standard deviation map is given for $t = 0$ and only the highest uncertainties levels are given along z-axis for the other dates.

uncertainty levels. This is the case on some positions along the main meandrum where no observations are ever available. Let us note the specific case of the bottom-left region (see again Figure 3) where a small area exhibits a high and recurrent ensemble spread. Such behaviors are typical for smoothers like 4DVarNet which cannot retrieve smaller energy areas without any additional observations. On this point, the future SWOT mission will surely help to solve for such problems (Gaultier et al., 2015; Metref et al., 2020; Beauchamp et al., 2022a; Febvre et al., 2022).

Because we are using an OSE experiment, we kept the Cryosat-2 alongtrack dataset for validation. Our computations take into account the observational noise of the Cryosat-2 nadir altimeter above the GulfStream domain (<https://doi.org/10.48670/moi-00144>) whose standard deviation varies between 0.01 and 0.02 m. Figure 6a,b displays the two alongtrack Cryosat-2 nadir crossings on January 10, 2021, with corresponding DUACS OI, 4DVarNet reconstruction and both En4DVarNet mean and interval of confidence. We also built a $1^\circ \times 1^\circ$ map over the GulfStream domain to accumulate all the occurrences of Cryosat-2 appearing in a pixel. This allows for computing the empirical probability that observations fall within the estimated 4DVarNet confidence interval (between the 5 and 95 percentile of the ensemble members) during the test period. Figure 6c shows this probability map on January 2017, with the number of Cryosat-2 occurrences given on the z-axis. The UQ provided by En4DVarNet seems often consistent with Cryosat-2 observations and probability values are greater than 0.7. Still, smaller areas display lower values (about 0.5) which may indicate the NATL60 simulation does not fully retrieve the statistical properties of true SSH dynamics over the GulfStream domain and there is a need for a larger catalogue or additional tools to emulate a stochastic version of such a variable in our framework.

4. Conclusion

In this work, we propose an ensemble-based version of the end-to-end DL 4DVarNet framework to both reconstruct and quantify uncertainty of SSH based on sparse altimetric nadir observations. Our solution

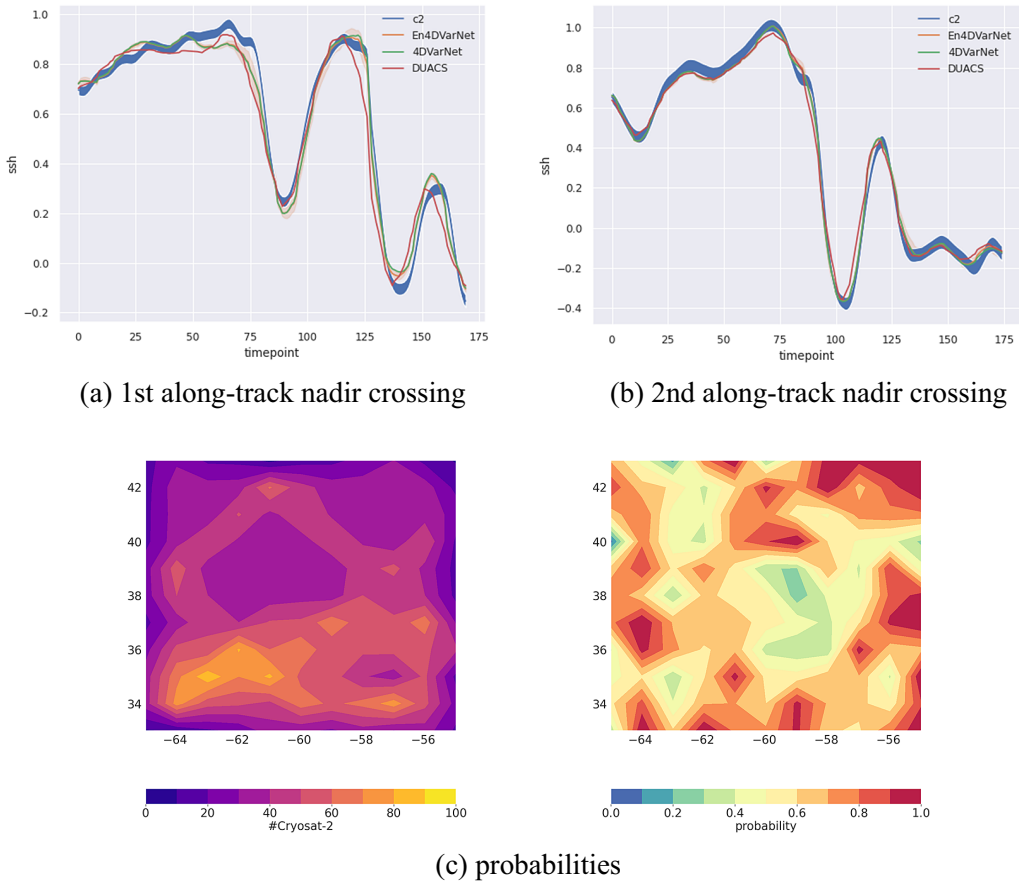


Figure 6. (a,b) SSH confidence interval for the two Cryosat-2 nadir crossings available on the GulfStream domain (January 10, 2017) and the corresponding DUACS, 4DVarNet and ensemble-based 4DVarNet reconstruction; (c) Left: number of Cryosat-2 occurrences inside the pixel during the test period, Right: Probabilities for Cryosat-2 dataset to be included in the 90% interval.

relies on the ability to select the analog situations (nearest neighbors) of the available observations from a model-based dataset. Doing so, the underlying assumption is that the samples drawn would share similar statistical properties than the true state. Then, a 4DVarNet realization compliant with the observations is obtained by adding to the classic 4DVarNet reconstruction a residual which is simply the difference between the analog and its 4DVarNet estimation using as observation a subsampling of the analog with similar sparsity mask than the actual observations. This solution and its application to real-world nadir SSH observations demonstrates that post-conditioning independent simulations is a promising way to create a stochastic formulation of 4DVarNet. Along this line, future works would consist in generalize the approach by getting free from the analog solution and being able to generate independent realizations. We also plan to explore other ways involving ensemble approaches. For instance, an alternate post-processing of 4DVarNet may be to add to the observations a random noise as it is done in stochastic EnKF formulations. An other solution would be to embed through a Gaussian or ensemble formulation a new uncertainty term to optimize during training based on likelihood loss functions.

Author contribution. Conceptualization: M.B.; Data curation: Q.F.; Methodology: M.B., R.F.; Resources: Q.F.; Supervision: R.F.; Visualization: M.B.; Writing—original draft: M.B. All authors approved the final submitted draft.

Competing interest. The authors declare no competing interests exist.

Data availability statement. The open-source 4DVarNet version of the code is available at <https://doi.org/10.5281/zenodo.7186322>. The datasets for training 4DVarNet are shared through the 2020 BOOST-SWOT ocean data challenge (ODC) available on GitHub (https://github.com/ocean-data-challenges/2020a_SSH_mapping_NATL60). The datasets for application on real-world nadir are also available on GitHub and shared by the 2021 BOOST-SWOT ODC (https://github.com/ocean-data-challenges/2021a_SSH_mapping_OSE).

Funding statement. This work was supported by LEFE program (LEFE MANU project IA-OAC), CNES (grant OSTST DUACS-HR), and ANR Projects Melody and OceaniX. It benefited from HPC and GPU resources from Azure (Microsoft Azure grant) and from GENCI-IDRIS (Grant 2020-101030).

Provenance statement. This article is part of the Climate Informatics 2023 proceedings and was accepted in *Environmental Data Science* on the basis of the Climate Informatics peer review process.

References

- Andrychowicz M, Denil M, Gomez S, Hoffman MW, Pfau D, Schaul T, Shillingford B and De Freitas N (2016) Learning to learn by gradient descent by gradient descent. In *Advances in Neural Information Processing Systems*. Red Hook, NY: Curran Associates, pp. 3981–3989.
- Asch M, Bocquet M and Nodet M (2016) *Data Assimilation. Fundamentals of Algorithms*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Beauchamp M, Febvre Q, Georgenthum H and Fablet R (2022a) 4DVarNet-SSH: End-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry. *Geoscientific Model Development Discussions* 2022, 1–37.
- Beauchamp M, Thompson J, Georgenthum H, Febvre Q and Fablet R (2022b) Learning neural optimal interpolation models and solvers.
- Chilès J and Delfiner P (2012) *Geostatistics: Modeling Spatial Uncertainty*, 2nd Edn. New York: Wiley.
- Counillon F and Bertino L (2009) Ensemble optimal interpolation: Multivariate properties in the gulf of Mexico. *Tellus A* 61(2), 296–308.
- Evensen G (2009) *Data Assimilation*. Berlin: Springer.
- Evensen G, Vossepoel FC and van Leeuwen PJ (2022) *Data Assimilation Fundamentals*. Springer Textbooks in Earth Sciences, Geography and Environment (STEGE). Cham: Springer.
- Fablet R, Beauchamp M, Drumetz L and Rousseau F (2021) Joint interpolation and representation learning for irregularly sampled satellite-derived geophysical fields. *Frontiers in Applied Mathematics and Statistics* 7, 25.
- Fablet R, Drumetz L and Rousseau F (2020) Joint learning of variational representations and solvers for inverse problems with partially-observed data.
- Febvre Q, Fablet R, Sommer JL and Ubelmann C (2022) Joint calibration and mapping of satellite altimetry data using trainable variational models. In *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Singapore: IEEE, pp. 1536–1540.
- Frezat H, Le Sommer J, Fablet R, Balarac G and Lguensat R (2022) A posteriori learning for quasi-geostrophic turbulence parametrization. *Journal of Advances in Modeling Earth Systems* 14(11), e2022MS003124.
- Gaultier L, Ubelmann C and Fu L-L (2015) The challenge of using future SWOT data for oceanic field reconstruction. *Journal of Atmospheric and Oceanic Technology* 33(1), 119–126.
- Lguensat R, Huynh Viet P, Sun M, Chen G, Fenglin T, Chapron B and Fablet R (2017) Data-driven interpolation of sea level anomalies using analog data assimilation.
- Metref S, Cosme E, Le Guillouf F, Le Sommer J, Brankart J-M and Verron J (2020) Wide-swath altimetric satellite data assimilation with correlated-error reduction. *Frontiers in Marine Science* 6, 822.
- Qiu C, Shao A, Xu Q and Wei L (2007) Fitting model fields to observations by using singular value decomposition: An ensemble-based 4dvar approach. *Journal of Geophysical Research: Atmospheres* 112(D11), D11105.
- Taburet G, Sanchez-Roman A, Ballarotta M, Pujol M-I, Legeais J-F, Fournier F, Faugere Y and Dibarboure G (2019) DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Science* 15(5), 1207–1224.
- Wackernagel H (2003) *Multivariate Geostatistics. An Introduction with Applications*. Berlin: Springer-Verlag.
- Zhang F, Zhang M and Hansen JA (2009) Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Advances in Atmospheric Sciences* 26(1), 1–8.

Cite this article: Beauchamp M, Febvre Q and Fablet R (2023). Ensemble-based 4DVarNet uncertainty quantification for the reconstruction of sea surface height dynamics. *Environmental Data Science*, 2: e18. doi:10.1017/eds.2023.19