

END-TO-END PHYSICS-INFORMED REPRESENTATION LEARNING FOR SATELLITE OCEAN REMOTE SENSING DATA: APPLICATIONS TO SATELLITE ALTIMETRY AND SEA SURFACE CURRENTS

Ronan Fablet¹, Mohamed Mahmoud Amar¹, Quentin Febvre¹, Maxime Beauchamp¹, Bertrand Chapron²

¹IMT Atlantique, UMR CNRS Lab-STICC, Brest, France.

²Ifremer, UMR CNRS LOPS, Brest, France.

KEY WORDS: Space oceanography, sea surface dynamics, satellite altimetry, SWOT mission, end-to-end learning, inverse problems, data assimilation, space-time interpolation, short-term forecasting, adaptive sampling.

ABSTRACT:

This paper addresses physics-informed deep learning schemes for satellite ocean remote sensing data. Such observation datasets are characterized by the irregular space-time sampling of the ocean surface due to sensors' characteristics and satellite orbits. With a focus on satellite altimetry, we show that end-to-end learning schemes based on variational formulations provide new means to explore and exploit such observation datasets. Through Observing System Simulation Experiments (OSSE) using numerical ocean simulations and real nadir and wide-swath altimeter sampling patterns, we demonstrate their relevance w.r.t. state-of-the-art and operational methods for space-time interpolation and short-term forecasting issues. We also stress and discuss how they could contribute to the design and calibration of ocean observing systems.

1. INTRODUCTION AND PROBLEM STATEMENT

Satellite sensors provide invaluable information for the observation, reconstruction, forecasting and simulation of upper ocean dynamics, which is of key importance for variety of scientific and societal challenges, including for instance marine pollution monitoring, offshore activities, maritime traffic, climate studies,... Ocean processes involve a variety of space-time scales along with multi-scale interactions. As such, no in situ and space observing system can inform all the scales and processes in play at once. This may result both from the properties of the sampling pattern of the observing systems (e.g., punctual observation for fixed-point systems, along-track data of polar-orbiting satellites) as well as from the sensitivity of sensors of atmospheric conditions (e.g., the cloud cover for infrared satellite sensors, heavy rain and/or strong wind conditions for radiometers and SAR sensors). Overall, ocean remote sensing data, including from satellite earth observation mission, result in irregularly-sampled datasets. The full exploitation of the resulting large-scale geospatial dataset collected over the last decades through learning-based schemes then asks for specific methods and tools as investigated in this paper.

As a typical case-study, we focus here on satellite altimetry and sea surface currents. The launch of the first operational altimetry missions in the 80's has deeply renewed the knowledge of ocean circulation, especially with the importance, much stronger than expected before, of ocean turbulence, especially ocean eddies (i.e., processes characterized by horizontal scales below a few hundreds of kilometers). Satellite altimeters currently involve narrow-swath polar-orbiting radar sensors. As illustrated in Fig.2, they only sample the ocean surface under the track of the satellite. On a daily scale, even with a 4-satellite constellations, they may only cover about 1% of the ocean surface. Space-time interpolation techniques are then critical to deliver gridded gap-free products for sea surface currents, which are

key data for a wide range of applications (e.g., plastics drift, marine pollution monitoring, maritime traffic routing, data assimilation for ocean-atmosphere models,...). State-of-the-art operational products rely either on optimal interpolation (OI) techniques (Taburet et al., 2019) or on the assimilation of ocean models (Tranchant et al., 2019). In both cases, these operational products may not retrieve sea surface processes with characteristic scales below 100km. Optimal interpolation schemes usually rely either on some expert-based calibration of covariance models, which prove very complex for finer-scale processes. Regarding model-driven data assimilation schemes, one may question the complexity of the inversion of a full ocean state given only very scarce observation data as well as the ability of ocean models to match all the features and patterns informed by observation data.

These limitations of both OI and model-driven data assimilation along with the availability of large-scale observation datasets have motivated the investigation of data-driven schemes. Broadly speaking, data-driven methods generally involve: (i) the identification of some representation of the underlying process from data, (ii) the use of this representation to reconstruct the considered space-time processes from the available irregularly-sampled observations. PCA-based and analog (i.e., nearest-neighbor) methods (Alvera-Azcarate et al., 2005, Lguensat et al., 2017) were first considered. As it has rapidly become the state-of-the-art approaches for numerous signal and image processing issues, including for instance for super-resolution or denoising problems (Dong et al., 2016, Chen et al., 2015), deep learning naturally arises as a particularly appealing class of methods for ocean remote sensing data. Especially, recent advances have been reported for the development of deep learning schemes for the resolution of inverse problems (Chen et al., 2015, Aggarwal et al., 2019). In this context, an end-to-end learning scheme based on a variational formulation has recently been introduced to address inverse problems (Fablet et al., 2020). The

underlying variational formulation provides means to embed some physics-guided knowledge on the considered observing systems and geophysical processes, while the end-to-end learning framework makes it direct to train both the different terms of the variational model and the associated solver.

In this paper, we investigate this framework for the monitoring of sea surface dynamics from satellite data. Through OSSEs (Observing System Simulation Experiment) for altimetry case-studies in a Gulf Stream region, we illustrate three different applications. We first address illustrate the reconstruction of SSH (Sea Surface Height) fields from satellite-derived altimeter data and include an evaluation w.r.t. state-of-the-art methods for (Section 3.2). A second application addresses the short-term forecasting of SSH fields (Section 3.3). The third application explores whether we may learn to predict where to sample new observations to complement satellite altimetry and best inform sea surface dynamics (Section 3.4). These different applications point out new means to learn task-specific or task-adapted representation and solvers for space-based ocean observing systems.

This paper is organized as follows. Section 2 describes the proposed end-to-end learning framework. Section 3 presents the considered OSSE framework for satellite altimetry and the three different applications we carry out in this work. Section 4 further discusses our main contributions and future work.

2. PROPOSED END-TO-END LEARNING FRAMEWORK

This section presents the end-to-end learning framework that we apply to satellite altimetry case-studies in Section 3. We first introduce the considered variational formulation to deal with inverse problems with irregularly-sampled observation data (Section 2.1). Section 2.2 introduces the associated trainable solver, while the overall end-to-end learning approach is summarized in Section 2.3 and sketched in Fig.1.

2.1 Variational model

In this work, we develop the application of the end-to-end learning framework introduced in (Fablet et al., 2020) to ocean remote sensing case-studies. This framework relies on a variational formulation of inverse problems given some partial observation y over a subdomain Ω of a space-time state of interest x . Formally, the reconstruction of state x from observation y relies on the minimization of a following variational cost

$$U_{\Phi, \mathcal{H}}(x, y, \Omega) = \lambda_1 \|\mathcal{H}_{\Omega} \cdot (x - y)\|^2 + \lambda_2 \|x - \Phi(x)\|^2 \quad (1)$$

with $\|\mathcal{H}_{\Omega} \cdot (x - y)\|^2$ the observation term and $\|x - \Phi(x)\|^2$ the prior. \mathcal{H}_{Ω} is the indicator of subdomain Ω . When considering space-time dynamics governed by differential operator, operator Φ may be stated as flow operator

$$\Phi(x)(t) = x(t - \Delta) + \int_{t-\Delta}^t \mathcal{M}(x(u)) du \quad (2)$$

where Δ is the predefined time step and dynamical model \mathcal{M} governs differential equation $dx/dt = \mathcal{M}(x(t))$.

Within this general variational setting, model-driven strategies come to minimize the above variational cost according to a

physics-driven parameterization of flow operator Φ . This parameterization generally relies on some numerical integration scheme. When considering sea surface dynamics, the derivation of approximate dynamical model, such as Surface Quasi-Geostrophy (SQG) (Klein et al., 2009), is a complex issue, which may not fully account for all dynamical regimes in play. Besides, one may question the extent to which such models are relevant to solve reconstruction problems.

Following (Fablet et al., 2020), we derive a neural network architecture backed on variational formulation (1). Interestingly, for a given parameterization of the prior $\|x - \Phi(x)\|^2$, we may jointly train operator Φ and the associated solver so that we minimize some reconstruction performance metrics. Based on interpolation experiments reported in (Beauchamp et al., 2020), we consider a two-scale parameterization for operator Φ . The latter may be regarded as a two-scale U-Net architecture initially introduced for image segmentation (Cicek et al., 2016). Formally, this architecture relates to the following decomposition

$$\Phi(x) = Up(\Phi_1(Dw(x))) + \Phi_2(x) \quad (3)$$

where $Up()$ and $Dw()$ are upsampling and downsampling operators. Operator Φ_1 applies to downsampled states and operator Φ_2 processes to the full-resolution states. In our implementation, both operators Φ_1 and Φ_2 involve bilinear blocks (Beauchamp et al., 2020). We may point out that other architectures could be considered, especially auto-encoder and neural ODE architectures. They were benchmarked in previous interpolation experiments with toy models and datasets, especially chaotic Lorenz dynamics (Fablet et al., 2020). U-Net outperformed auto-encoder and ODE-based architectures, which motivate our choice here.

2.2 Trainable solver architecture

Given variational cost (1) and the associated neural network implementation, we design a neural-network solver architecture. Within a model-driven framework, the minimisation of this variational cost generally exploits an iterative gradient descent algorithms. Here, we benefit from automatic differentiation tools embedded in deep learning frameworks to proposed a gradient-based solver architecture. More precisely, given a current estimate $x^{(k)}$ at iteration k , we consider the following iterative update rule

$$\begin{cases} g^{(k+1)} = LSTM[\alpha \cdot \nabla_x U_{\Phi, \mathcal{H}}(x^{(k)}, y), h(k), c(k)] \\ x^{(k+1)} = x^{(k)} - \mathcal{G}(g^{(k+1)}) \end{cases} \quad (4)$$

where α is a scalar parameter, $LSTM()$ a LSTM (Long Short Term Memory) model, $\{h(k), c(k)\}$ the internal LSTM states and \mathcal{G} a linear layer which maps the output of the LSTM to the space of state x . Given the ability of LSTMs to capture long-term dependencies, this LSTM-based iterative update may be regarded as a means to keep track of the full sequence of gradient updates when computing a new updates at iteration k . Similar LSTM-based descent updates have been considered optimizer learning issues (Hospedales et al., 2020).

Overall, for a given initialization, the considered gradient-based solver architecture applies a predefined number of steps of the considered gradient-based iterative update, typically from 5 to 15 in our implementation.

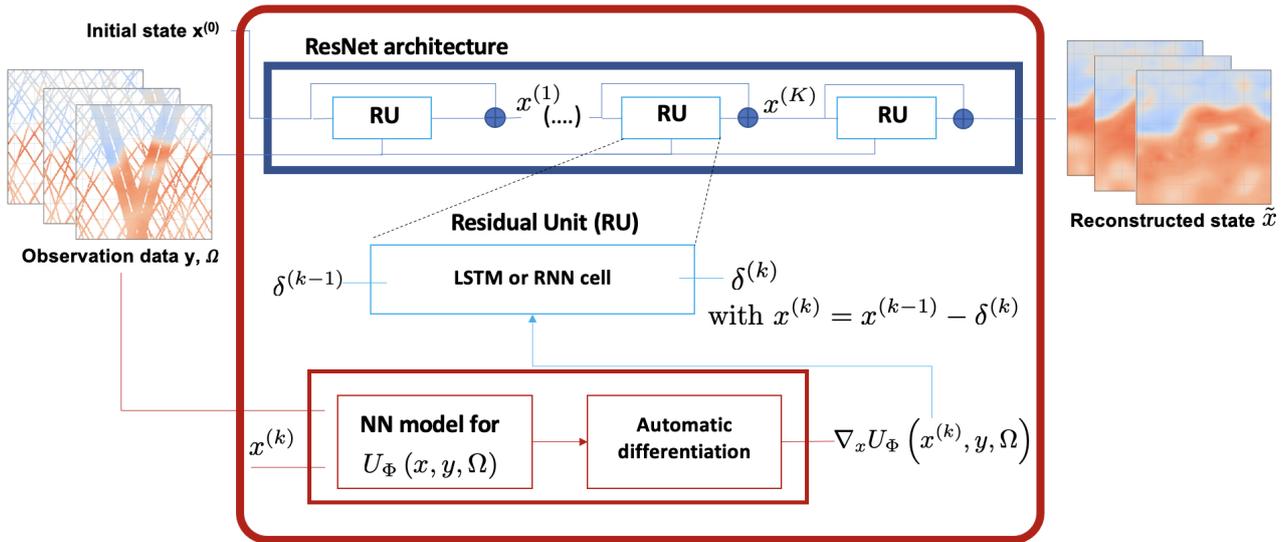


Figure 1. Sketch of the proposed end-to-end architecture for the reconstruction and forecasting of sea surface dynamics from irregularly-sampled satellite observations: the proposed framework relies on the definition of a variational cost $U_{\Phi, \mathcal{H}}$ from which we derive a gradient-based iterative solver. The latter is implemented as a residual network using a LSTM update rule. We let the reader refer to the main text for a detailed presentation of the variational setting and of the associated solver.

2.3 End-to-end learning scheme

Based on variational formulation (1) and on the associated gradient solver, we introduce a end-to-end trainable architecture as sketched in Fig.1. A key feature of this architecture is that it considers as inputs the raw observation data. As such, we may jointly train all the components of this architecture w.r.t. some performance measure. In our experiments, we typically implement training losses based on the mean square error for the reconstruction or forecasting of variable x as well as of the norm of its gradient x . Let us denote by $\Psi_{\Phi, \mathcal{H}, \Gamma}(x^{(0)}, y, \Omega)$ the output of the end-to-end architecture given some observation y over subdomain Ω and an initial estimate $x^{(0)}$. This amounts to defining the following training loss

$$\mathcal{L} = \nu_1 \sum_n \|x_n - \tilde{x}_n\|^2 + \nu_1 \sum_n \|\nabla x_n - \nabla \tilde{x}_n\|^2 \quad (5)$$

with \tilde{x}_n standing for reconstructed state

$$\Psi_{\Phi, \mathcal{H}, \Gamma}(x_n^{(0)}, y_n, \Omega_n), y_n, \Omega_n).$$

To further constrain the training of operator Φ in (3), we consider additional training losses corresponding to the auto-encoder capabilities of operator Φ for both true and reconstructed states

$$\nu_3 \sum_n \|x_n - \Phi(x_n)\|^2 + \nu_3 \sum_n \|\tilde{x}_n - \Phi(\tilde{x}_n)\|^2.$$

Weighing factors ν_{1-4} are set using a cross-validation procedure to balance the different losses in the training stage.

Regarding implementation issues, all experiments are run using a pytorch implementation using Adam solver. During the training phase, we gradually increase the number of gradient-based iterations in the considered end-to-end architecture and decrease the learning rate. We may stress that we optimise all trainable components, *i.e.* operator Φ and solver Γ . We refer

the interested reader to our pytorch implementation ¹.

3. APPLICATION TO SATELLITE-DERIVED OCEAN SURFACE TOPOGRAPHY

We report different application of the proposed end-to-end learning scheme to the analysis and reconstruction of sea surface currents from satellite-derived observations. We focus on satellite altimetry data (Chelton et al., 2001, 1). We first introduce the considered dataset within an OSSE (Observing System Simulation Experiment) framework. We then report numerical experiments for the following case-studies: (i) the space-time interpolation of sea surface height (SSH) from along-track nadir and wide-swath altimetry data, (ii) the short-term forecasting of SSH dynamics, (iii) the design of adaptive sampling strategies to improve the reconstruction of SLA fields.

3.1 Satellite altimetry OSSE

In this work, all the datasets are based on the NEMO model (Nucleus for European Modeling of the Ocean) NATL60 configuration (Molines, 2018) which is a state-of-the-art basin-scale high-resolution ($1/60^\circ$) simulation used here as Ground Truth (GT). A $10^\circ \times 10^\circ$ subpart of the GULFSTREAM domain [33° N, 43° N ; -65° W, -55° W] is used (see Fig. 2), where Sea Surface Height (SSH) is mainly driven by energetic mesoscale dynamics. The resolution of the nature run is here downgraded to $1/20^\circ$. The pseudo-altimetric nadir and SWOT observational datasets are generated through a sub-sampling of the Ground Truth with realistic satellite constellations.

Regarding the pseudo-nadir dataset, representative of the current observational capabilities, a constellation with 4 altimetry missions (TOPEX/Poseidon, Geosat, Jason-1 and Envisat) from October 1st, 2012 to September 29th, 2013 is used to simulate along-track nadir observation from NATL60 data. An instrumental acquisition Gaussian white noise with variance $\sigma^2 =$

¹ https://github.com/CIA-Oceanix/DinAE_4DVarNN_torch

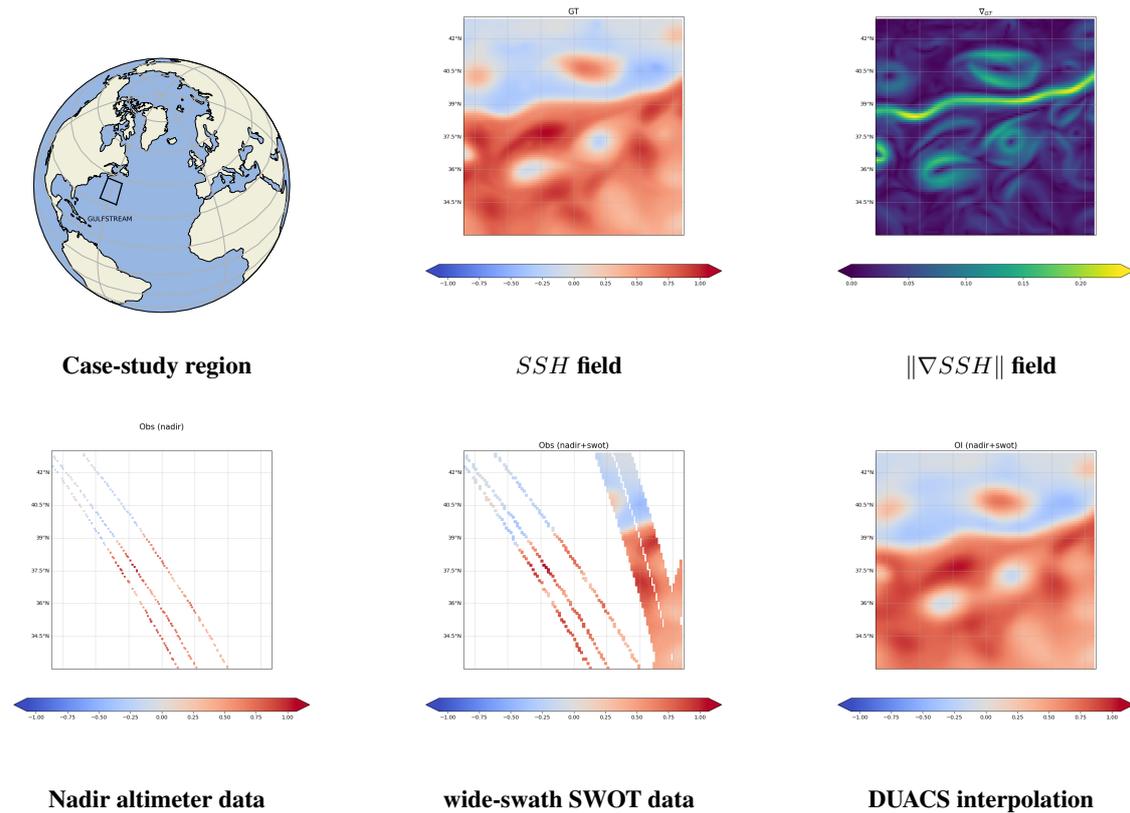


Figure 2. **Considered OSSE case-study on satellite-derived sea surface altimetry:** we implement an Observing System Simulation Experiment (OSSE) for a case-study region along the Gulf Stream (top, left). We also depict an example of SSH (Sea Surface Height) field (top center), of its gradient norm (top right), of the associated irregular-sampling observation data for nadir altimeters (bottom left) and the upcoming SWOT wide-swath altimeter (bottom right), and of the corresponding gap-free DUACS interpolation using all altimeter data (bottom right).

30cm is then added to the interpolated NATL60 simulation (Ballarotta et al., 2019). Similarly, we generate SWOT pseudo-observations using the SWOT simulator tool (Gaultier et al., 2015) are provided. Instrumental noise is also added to the NATL60 subsampling. Because the SWOT-related noise can have strong space-time correlations, they are first filtered out via an initial preprocessing. Besides, our dataset comprises the reconstruction issued from the operational DUACS OI-based (Optimal Interpolation) products (Taburet et al., 2019). All data are given as gridded fields with regular $(0.05^\circ \times 0.05^\circ)$ resolution on a scale daily.

We illustrate in Fig.2 an example of SSH field along with the corresponding nadir altimeter and SWOT wide-swath altimeter observations. This example illustrates how scarce is the spatial sampling of the sea surface as sampled points cover only about 1% of the domain.

3.2 Space-time interpolation of SSH fields

This first application addresses the space-time interpolation of SSH fields. In (1), observation operator \mathcal{H}_Ω encodes the mask of the available altimeter data for each date and variable x refers to the anomaly w.r.t. the OI fields over 10 consecutive days. Regarding the parameterization of operator Φ , for each scale we apply sequentially a convolution layer with 3×3 kernels, 10 filters and ReLu activation, a convolution later with 1×1 kernels and 5 filters, a bilinear layer with 1×1 kernels and 10 fil-

ters and a convolution later with 1×1 kernels and 5 filters. The coarse scale operator involves a subsampling by a factor of 8. We sum the output of the two scales after upsampling using a ConvTranspose layer. Regarding the architecture of solver Γ , we consider a convolutional LSTM in gradient-based update (4) with 100-dimensional hidden states. For benchmarking purposes, we evaluate the proposed scheme w.r.t. DUACS OI, DINEOF (Alvera-Azcarate et al., 2005) and AnDA (Fablet et al., 2017). To evaluate the relevance of the gradient-based solver, we also include a comparison of the proposed end-to-end learning framework where the solver is a fixed-point algorithm as presented in (Fablet et al., 2019), referred to FP-4DVarNet. We refer to the proposed end-to-end scheme using a gradient solver as Grad-4DVarNet. As test dataset, we consider a 20-day period from day 60 to day 80 of the considered time series. For training, we use the remaining data with a 10-day lag prior and after the test period.

As detailed in Tab.1, we report reconstruction performance in terms of explained variance for the SSH field (R-score) and the norm of its gradient (∇ R-score), which relates to the magnitude of the sea surface current. For the proposed scheme, we also evaluate the representation score (AE-score) of the trained operator Φ through the explained variance of projection $\Phi(x)$ w.r.t. x . The proposed framework outperforms the other data-driven approaches and results in a relative gain of about 35% (resp. 45%) in terms of mean square error for the reconstruc-

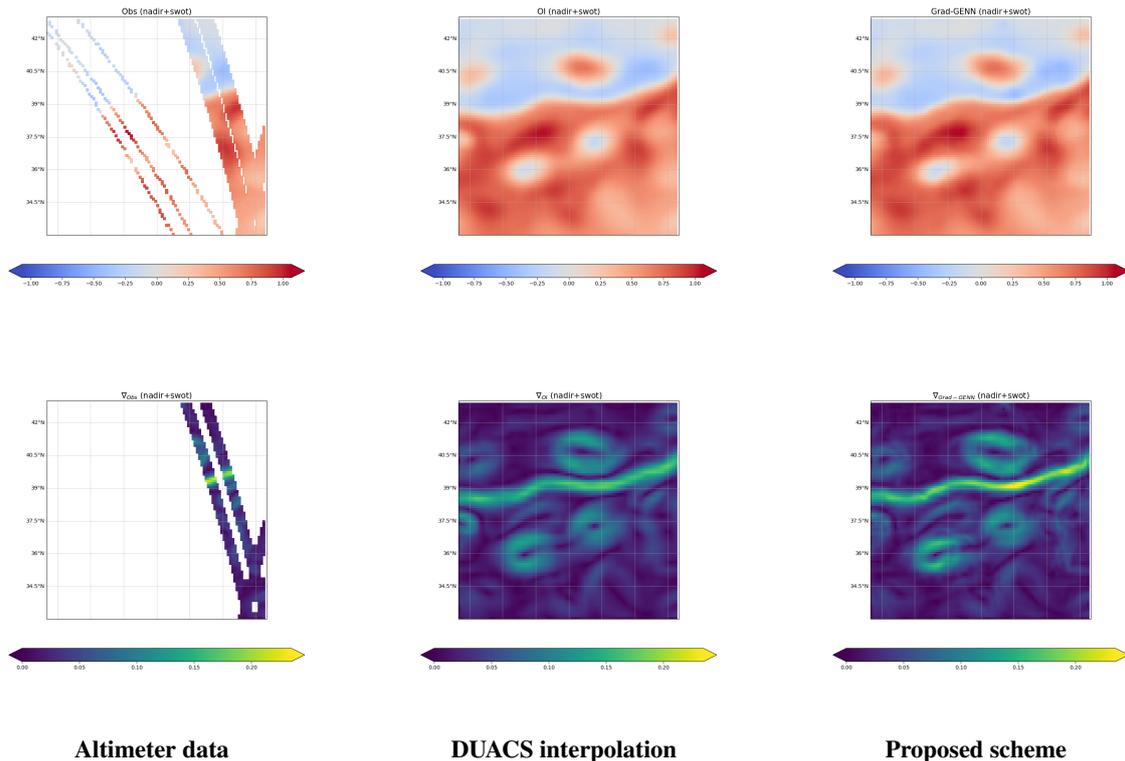


Figure 3. **Space-time interpolation of SSH fields from altimeter data:** the first row reports the interpolated SSH fields using nadir and SWOT wide-swath altimeter data (top left) for DUACS optimal interpolation (top center) and the proposed end-to-end learning scheme (top right). The second row depicts the associated gradient norm. The reader may refer to Fig.2 to compare the reported reconstructions with the true state.

tion of the SSH field (resp. its gradient norm) w.r.t. DUACS OI. These results also stress the additional benefit of the trainable gradient-based solver compared with a parameter-free fixed-point architecture. We may also note that, besides the reported reconstruction performance, the trained end-to-end architecture retrieves a relevant representation of the SSH space-time dynamics through operator Φ , which cover up to 99% of their variance. We further illustrate these results in Fig.3. Especially, the reconstruction of the more intense currents is clearly improved compared with OI reconstruction. We may remind the reader that in Fig.3, we only depict the altimeter data available for the reconstruction date, given that the end-to-end architecture is provide as inputs with 5 consecutive days of altimeter data as well as the optimally-interpolated field. This explains why, even if on the considered day no observation data are available for the western area of the domain, we can reconstructs some eddies, which are most likely partly sampled before or after the considered date.

3.3 Short-term forecasting of SSH dynamics

The second case-study addresses the short-term forecasting of SSH dynamics. It comes to apply a configuration similar to that of the space-time interpolation case-studies where observation operator \mathcal{H}_{\otimes} comes to assume that past dates are observed and future ones are not. For instance, when considering a 5-day-ahead forecasting from 3 past states, it amounts to considering a 8-day state x , where observation domain Ω states that observation y is available only for the first three days. We also adapt the training loss such that it evaluates the reconstruction per-

Model type	R-Score	∇ R-score	AE-score
OI	96.39	74.94	-
AnDA	96.23	80.25	-
DINEOF	96.49	81.62	-
FP-4DVarNet	96.69	78.90	99.27
G-4DVarNet	97.68	86.51	99.42

Table 1. **Interpolation performance of the benchmarked schemes for SSH interpolation from nadir altimeter and SWOT wide-swath altimeter data:** reconstruction performance for SSH fields and their gradient norms for the considered validation period. We refer the reader to the main text for the definition of the different scores and the details on the benchmarked schemes.

formance for the forecasting period. In this setting, for a n -day-ahead forecasting, we jointly train variational model (1), more specifically operator Φ , and solver Γ such that we best predict the last n steps of each training sequence.

We consider a configuration for operator Φ and solver Γ similar to the space-time interpolation case-study. Following (Ouala et al., 2020), we also test a parameterization where state x is higher-dimensional than the observation sequence (i.e., here, the true SSH sequence). The motivation for considering augmented states is that SSH dynamics clearly depends on unobserved processes, for instance ocean's interior dynamics, that we may expect the learning process to retrieve through the optimization of the forecasting performance. We vary the number of past dates, referred to as N_T , used as inputs and the number of aug-

Configuration	$dt = 1$	$dt = 2$	$dt = 3$	$dt = 4$	$dt = 5$
$N_T = 1, N_{Aug} = 0$	60.5% (68.5%)	62.4% (69.7%)	62.2% (67.1%)	60.4% (64.0%)	60.1% (63.0%)
$N_T = 5, N_{Aug} = 0$	70.8% (78.8%)	75.0% (78.8%)	74.3% (75.8%)	73.8% (73.3%)	70.0% (69.5%)
$N_T = 5, N_{Aug} = 50$	75.5% (81.2%)	80.5% (82.5%)	80.5% (80.4%)	78.5% (76.4%)	75.2% (72.3%)
$N_T = 5, N_{Aug} = 100$	75.9% (81.8%)	80.2% (82.6%)	81.2% (81.4%)	80.0% (78.6%)	76.8% (75.6%)
$N_T = 5, N_{Aug} = 200$	74.6% (81.8%)	80.0% (82.3%)	80.0% (80.2%)	77.8% (75.6%)	74.7% (70.9%)

Table 2. Short-term forecasting performance of the proposed end-to-end learning scheme: we evaluate the proposed end-to-end learning framework, where we vary the number N_T of past SSH fields used as inputs and the dimension N_{Aug} of the augmented components for variable x (i.e., $N_{Aug} = 0$ corresponds to a multivariate space-time field with the same dimension of the SSH sequence). As performance metrics, we compute a normalized mean square error (MSE) of the dt -day-ahead forecasting with dt from 1 to 5 for the SSH and its gradient norm. We let the reader to the main text for additional details.

mented components N_A in variable x to evaluate their impact on the forecasting performance (Tab.2). The latter is evaluated in terms of mean square error normalized by the variance of a persistence model (i.e., assuming the SSH field remains constant) for the SSH field and its gradient norm.

The reported results point out the relevance of a sequence of past SSH fields to improve the short-term forecasting performance (e.g., 60% of explained variance of the SSH for a 5-day-ahead forecasting using $N_T = 0$ w.r.t. 77% for the best configurations with $N_T = 5$). We may point out that the reported performance metrics are a relative forecasting performance score w.r.t. the persistence. This explains why this performance score is lower for a one-day-ahead forecasting than for a two-day-ahead one. Similarly to (Ouala et al., 2020), the configurations with augmented states also lead to an improvement with a relative gain between 10% and 15% in terms of forecasting performance. Interestingly, the trained models include the relative weights given to each observation time steps in variational model (1). For instance, for configuration $N_T = 5$ and $N_{Aug} = 50$, the largest weight is given to time step t with a value of 1.15, whereas time steps $t - 2$ and $t - 3$ are given weights close to 1.0. Besides, time steps $t - 1$, $t - 4$ and $t - 5$ have the lowest weights with values between 0.6 and 0.8. While tested here with gap-free fields as inputs, the proposed end-to-end scheme shall also apply with irregularly-sampled fields as inputs as well as multi-tracer inputs. This opens interesting research avenues for future work.

3.4 Learning where to sample SSH measurements

We also investigate the proposed end-to-end learning framework to learn where to sample additional measurements to improve the reconstruction of SSH fields. Such sampling design could be of key interest for scientific cruises to improve the knowledge of sea surface conditions given by satellite observations. To address this issue, we extend variational setting $U_{\Phi, \mathcal{H}}(x, y, \Omega)$ (3) as follows

$$\lambda_1 \|\mathcal{H}(\bar{y}) \cdot (x - y)\|^2 + \lambda_2 \|x - \Phi(x)\|^2 \quad (6)$$

where \bar{y} is a coarse estimate of the SSH field. In the reported experiments, we consider for \bar{y} the optimally-interpolated SSH field using along-track nadir altimeter data. Here, $\mathcal{H}(\bar{y})$ stands

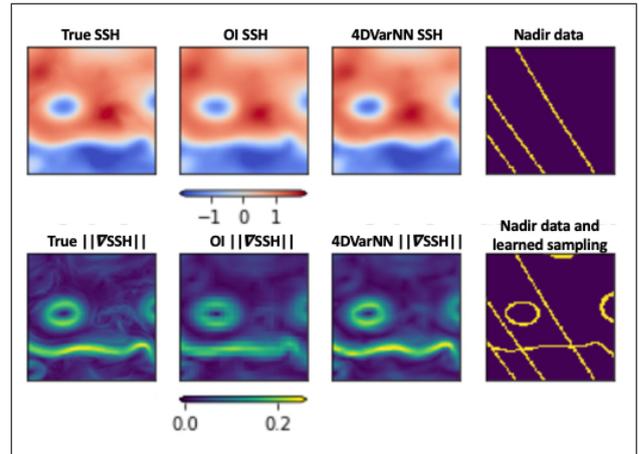


Figure 4. Illustration of the training of an adaptive sampling design for the reconstruction of SSH fields from irregularly-sampled data: in the first row, we report for a reference SSH field (top left), the associated OI reconstruction from nadir altimeter data (top right) along with the reconstruction issued from the proposed end-to-end learning framework. For the reported example, the trained sampling operator proposes to sample data along the main front as well the main eddies (bottom right) to complement the nadir altimeter data. As illustrated by the gradient norm fields (bottom row), the additional measurements lead to an improvement of the reconstruction of the gradient field. Quantitative results are in Tab.3.

for a state-adaptive sampling operator. For any coarse condition \bar{y} , $\mathcal{H}(\bar{y})$ shall comprise as few non-zeros positive values as possible. Such a constraint may be enforced during the training phase by adding to the training loss a new term based on the L_1 norm of the sampling operator $\max(1/N \sum_n \|\mathcal{H}(\bar{y}_n)\| - \xi, 0)$ with ξ a scalar parameter to control the expected sparsity level.

Regarding the parameterisation of sampling operator $\mathcal{H}(\cdot)$, we consider a CNN architecture whose penultimate layer involves a sigmoid activation to rescale the output between 0 and 1. To enforce sparsity, the last layer is a thresholding layer with a threshold set to 0.1, meaning that all values below 0.1 are set to zero. Regarding the training phase, we proceed similarly to the interpolation case-study except that we also learn sampling operator $\mathcal{H}(\cdot)$ in addition to operator Φ and solver Γ . We may also account for available observations over a given subdomain Ω and replace $\mathcal{H}(\bar{y})$ by $\max(\mathcal{H}(\bar{y}), \mathcal{H}_\Omega)$.

Here, we report OSSE experiments to learn a sampling pattern to improve the reconstruction of a SSH field issued from nadir along-track altimeter data. We use as conditioning variable \bar{y} the optimally-interpolated (OI) SSH field, which typically resolves horizontal scales up to 100km. For 5-day-long sequences, we investigate two different parameterizations of the sampling operator: a first one which constrains the sampling to horizontal and vertical lines, a second one with no specific spatial structure. To explicitly benefit from the available OI estimation, we apply the considered variational framework (3) where state variable x involves two components a coarse-scale component \bar{x} and a fine-scale anomaly δx such that the SSH field is $\bar{x} + \delta x$. Subsequently, the observation variable concatenates OI reconstruction \bar{y} and the anomaly of the along-track nadir data w.r.t. \bar{y} .

Method	Sampling	% of data	SSH	$\ \nabla SSH\ $
OI	NA	1.34%	1.22%	7.24%
4DVarNN	NA	1.34%	0.81%	4.95%
4DVarNN	NA+A	5.40%	0.64%	4.20%
4DVarNN	NA+B	5.90%	0.46%	3.70%

Table 3. Impact of the sampling design on reconstruction performance for SSH fields: for different sampling designs, we evaluate the reconstruction performance in terms of mean square error of the SSH field and of the norm of the gradient of the SSH field for the test dataset normalised by the variance of the groundtruthed reference. As baseline, we consider an Optimal interpolation using nadir along-track data. We then compare three configurations of the proposed end-to-end learning using nadir altimeter data solely, combined with a trained sparse sampling design constrained to horizontal and vertical lines (A) and combined with a trained sparse sampling design with no additional constraint (B).

Quantitative results reported in Fig.3 show that sampling between 4% and 5% of additional SSH measurements when appropriately selected through the trainable sampling operator can lead to decreasing by a factor of 3-to-4 of the MSE for the SSH and its gradient norm. As expected, the unconstrained sampling operator selects more informative points for the reconstruction than the sampling operator constrained to sampling rows and columns. We illustrate these results for one example in Fig.4. The end-to-end scheme has clearly learned that high gradient points shall be preferentially sampled to improve the reconstruction performance, as the trained sampling operator samples points along the main SSH front as well as along the boundaries of the main mesoscale eddies. These results support the relevance of the OI estimate, used here as input of the sampling operator, to indicate the areas of the true SSH field with the more-energetic fine-scale structures.

4. DISCUSSION

In this work, we have investigated the application of end-to-end learning schemes based on variational formulations for the exploitation of satellite ocean remote data, and more specifically satellite altimetry data. Due to sensors' characteristics, these observation datasets are characterized by an irregular space-time sampling of the sea surface and very large missing data rates, typically about 95-99% on a daily scale for nadir altimeters. The proposed scheme naturally applies to irregularly-sampled input data through a variational formulation. Such variational formulations are classically used to solve inverse problems and combine an observation term and a prior. Here, we implement this variational formulation in a neural network architecture, which also comprises a trainable gradient-based solver of the inverse problem. For a given task, we can jointly learn all the trainable components of the neural network architecture to optimize some predefined performance criterion. Through different case studies, we show that this generic end-to-end learning scheme can reach state-of-the-art performance for the reconstruction and forecasting of SSH fields.

The first two case-studies address the space-time interpolation and short-term forecasting of SSH fields. In these case-studies, we jointly learn a space-time prior and the solver of the reconstruction problem the observation term being predefined through the sampling patterns of the considered satellite altimeter settings. Regarding space-time interpolation issues, our experiments support the relevance of upcoming SWOT wide-swath

altimeter data to improve the reconstruction of sea surface dynamics for horizontal scales below 100km. We report a significant gain w.r.t. the operational OI product (DUACS) with a relative gain of about 20% in terms of mean square error. We also demonstrate how the proposed framework applies to short-term forecasting issues. Here, from a sequence of SSH fields, we use an augmented representation space inspired by (Ouala et al., 2020) and we recover up to 80% of the variance w.r.t. the persistence model for a 4-day-ahead prediction. The augmented representation combined with the ability to consider a sequence of daily inputs clearly contribute to the improvement of the forecasting performance. In these different case-studies, the training loss is stated as the reconstruction or forecasting performance on the entire domain assuming we are provided with a groundtruthed gap-free dataset. Future work will investigate how we may adapt this training procedure to training losses only computed over observed points. We expect the results presented in our experiments to extend to these real-world configurations if we can gather sufficiently large observation data to compensate for a subsampling of the observed points. Future work shall also investigate how the proposed framework may apply on larger regions, typically the global ocean or an ocean basin. One key question is the ability to learn priors which apply to all upper ocean regimes (Klein et al., 2009).

We investigate an other original case study through the learning of sampling patterns with a view to improving the reconstruction performance issued from satellite altimeter data. It comes to complement the considered variational model with a trainable observation operator. We design a sampling operator under sparsity constraint such that only a fraction of the domain is observed. Besides, this sampling operator may be conditioned by an auxiliary variable, here a low-resolution version of the SSH field. Through numerical experiments, we show that we may predict where to sample additional SSH measurements which may complement nadir altimeter data to significantly improve the reconstruction of the SSH field. These experiments may provide the basis for investigating context-adaptive observing systems which could adapt their sampling design based on a coarse prediction or forecast, or even a proxy, with a view to optimizing a predefined reconstruction and/or forecasting performance.

We believe the proposed scheme may be of broad interest beyond satellite altimetry and ocean remote sensing to earth observation and remote sensing applications. Similarly to the case-studies illustrated in this paper, it provides a generic framework to address reconstruction and forecasting problems. To broaden the application range, future work shall further investigate trainable observation operators, including when the observed variable is not directly a noisy version of the variable of interest.

ACKNOWLEDGEMENTS

This work was supported by LEFE program (LEFE MANU project IA-OAC), CNES (grant SWOT-DIEGO) and ANR Projects Melody and OceaniX. It benefited from HPC and GPU resources from Azure (Microsoft EU Ocean awards) and from GENCI-IDRIS (Grant 2020-101030).

REFERENCES

Abdalla, S., Abdeh Kolahchi, A., et al, 2021. Altimetry for the future: Building on 25 years of progress. *Advances in Space Research*.

- Aggarwal, H. K., Mani, M. P., Jacob, M., 2019. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, 38(2), 394–405.
- Alvera-Azcarate, A., Barth, A., Rixen, M., Beckers, J. M., 2005. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature. *Ocean Modelling*, 9(4), 325–346.
- Ballarotta, M., Ubelmann, C., Pujol, M.-I., Taburet, G., Fournier, F., Legeais, J., Faugère, Y., Delepoulle, A., Chelton, D., Picot, N., 2019. On the resolutions of ocean altimetry maps. *Ocean Science*, 15(4), 1091–1109.
- Beauchamp, M., Fablet, R., Ubelmann, C., Ballarotta, M., Chapron, B., 2020. Intercomparison of Data-Driven and Learning-Based Interpolations of Along-Track Nadir and Wide-Swath SWOT Altimetry Observations. *Remote Sensing*, 12(22), 3806.
- Chelton, D. B., Ries, J., Haines, B. J., Fu, L.-L., Callahan, P. S., 2001. Satellite Altimetry. A. Cazenave, L.-L. Fu (eds), *International Geophysics, Satellite Altimetry and Earth Sciences A Handbook of Techniques and Applications*, 69, Academic Press, 1–ii.
- Chen, Y., Yu, W., Pock, T., 2015. On Learning Optimized Reaction Diffusion Processes for Effective Image Restoration. *Proc. IEEE CVPR*, 5261–5269.
- Cicek, O., Abdulkadir, A., Lienkamp, S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. Springer, 424–432.
- Dong, C., Loy, C. C., He, K., Tang, X., 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 38(2), 295–307.
- Fablet, R., Chapron, B., Drumetz, L., Memin, E., Pannekoucke, O., Rousseau, F., 2020. Learning Variational Data Assimilation Models and Solvers. *arXiv:2007.12941*. <http://arxiv.org/abs/2007.12941>. arXiv: 2007.12941.
- Fablet, R., Drumetz, L., Rousseau, F., 2019. End-to-end learning of energy-based representations for irregularly-sampled signals and images. *arXiv preprint arXiv:1910.00556*.
- Fablet, R., Viet, P. H., Lguensat, R., 2017. Data-Driven Models for the Spatio-Temporal Interpolation of Satellite-Derived SST Fields. *IEEE Trans. on Computational Imaging*, 3(4), 647–657.
- Gaultier, L., Ubelmann, C., 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.
- Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A., 2020. Meta-learning in neural networks: A survey. *arXiv preprint arXiv:2004.05439*.
- Klein, P., Isern-Fontanet, J., Lapeyre, G., Roullet, G., Danioux, E., Chapron, B., Le Gentil, S., Sasaki, H., 2009. Diagnosis of vertical velocities in the upper ocean from high resolution sea surface height. *Geophysical Research Letters*, 36(12), L12603.
- Lguensat, R., Tandeo, P., Aillot, P., Fablet, R., 2017. The Analog Data Assimilation. *Monthly Weather Review*.
- Molines, J.-M., 2018. MEOM configurations NATL60-CJM165: NATL60 code used for CJM165 experiment.
- Ouala, S., Nguyen, D., Drumetz, L., Chapron, B., Pascual, A., Collard, F., Gaultier, L., Fablet, R., 2020. Learning Latent Dynamics for Partially-Observed Chaotic Systems. *Chaos*. <http://arxiv.org/abs/1907.02452>. arXiv: 1907.02452.
- Taburet, G., Sanchez-Roman, A., Ballarotta, M., Pujol, M.-I., Legeais, F., Fournier, F., Faugere, Y., Dibarboure, G., 2019. DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Sci.*, 18.
- Tranchant, B., Remy, E., Greiner, E., Legalloudec, O., 2019. Data assimilation of Soil Moisture and Ocean Salinity (SMOS) observations into the Mercator Ocean operational system: focus on the El Niño 2015 event. *Ocean Science*, 15(3), 543–563.