Multimodal 4DVarNets for the Reconstruction of Sea Surface Dynamics From SST-SSH Synergies

Ronan Fablet, Senior Member, IEEE, Quentin Febvre, and Bertrand Chapron

Abstract—The space-time reconstruction of sea surface dynamics from satellite observations is a challenging inverse problem due to the associated irregular sampling. Satellite altimetry provides direct observation of the sea surface height (SSH), which relates to the divergence-free component of sea surface currents. The associated sampling pattern prevents operational schemes from retrieving fine-scale dynamics, typically below 10 days. By contrast, other satellite sensors provide higher-resolution observations of sea surface tracers such as sea surface temperature (SST). Multimodal inversion schemes then arise as appealing approaches. Though theoretical evidence supports the existence of an explicit relationship between SST and sea surface dynamics under specific dynamical regimes, the generalization to the variety of upper ocean dynamical regimes is complex. Here, we investigate this issue from a physics-informed learning perspective. We introduce a trainable multimodal inversion scheme for the reconstruction of sea surface dynamics from multisource satellite-derived observations, namely satellite-derived SSH and SST data. The proposed multimodal 4DVarNet schemes combine a variational formulation involving trainable observation and a priori terms with a trainable gradient-based solver. An observing system simulation experiment (OSSE) for a Gulf stream region supports the relevance of our approach compared with state-of-the-art schemes. We report a relative improvement greater than 60% compared with the operational altimetry product in terms of root mean square error (MSE) and resolved space-time scales. We discuss further the potential and the limitations of the proposed approach for the reconstruction and forecasting of geophysical dynamics from irregularly-sampled satellite observations.

Index Terms—End-to-end learning scheme, inverse problem, meta-learning, multimodal observations, satellite imaging, sea surface dynamics, variational models.

I. INTRODUCTION

SATellite altimeters provide the main source of observation data to inform sea surface dynamics on a regional and global scale [1], [10]. As illustrated in Fig. 1, the associated daily sampling at sea surface on a global scale remains very scarce for current nadir altimeter constellations. This results in relatively low-resolution reconstructions of sea surface dynamics delivered by operational altimetry-derived products using model-driven and observation-driven schemes [8], [37], [49]. We can emphasize that retrieving mesoscale sea surface dynamics, typically from a few tens of kilometers in terms of horizontal scales, is key for a wide range of scientific topics and applications [1], [10] such as among others weather and ocean forecasting, climate modeling, ecological studies, maritime traffic routing, offshore activities.

While future altimetry missions such as SWOT will improve the spatial sampling with wide-swath sensors, the associated time sampling will likely remain too scarce for a while to retrieve fine-scale sea surface dynamics [27], [37]. Numerous other satellite sensors deliver sea surface observations that may inform sea surface dynamics. Among others, we may cite sea surface temperature (SST) products [17] derived from micro-wave and infrared sensors as well as ocean color (OC) products from multispectral sensors [12]. These wide-swath sensors result in a much denser space-time sampling of the sea surface and the associated satellite-derived fields can reveal sea surface dynamics for finer scales compared with altimetry-derived products. This has motivated a rich literature to exploit SST and OC fields solely or combined with altimeter-derived products to better inform sea surface dynamics [31], [45], [50]. The illustration reported in Fig. 1 supports such a synergistic approach to jointly exploit satellite-derived SST and altimetry observations to reconstruct sea surface currents, as they clearly share common geometric patterns as stressed by the main meander and the large eddy South to the later. Through the surface quasi-geostrophic (SQG) theory, analytical derivations further support the relevance of SST-SSH synergies as explored in [31], [32], and [34]. The SQG theory is generally more applicable in low-energetic regions such as the ocean’s interior, than in high-energetic region such as the Gulf stream [31], [56]. This limits the operational exploitation of the resulting SST-derived inversion of sea surface dynamics. As such, the design of multimodal inversion schemes which could exploit joint altimetry and SST observation remains a challenge [1].

From a methodological point of view, the computation of satellite-derived geophysical products generally relies on optimal interpolation (OI) and data assimilation schemes [19], [37], [49], [51]. Deep learning methods have emerged as appealing approaches to address inverse problems [11], [41], including for applications to space oceanography [20], [28]. Especially, end-to-end neural schemes provide plug-and-play solutions to train multimodal inversion schemes as they can directly learn from data the underlying multimodal
Section V.

Section III presents our numerical experiments on the multimodal learning-based inversion framework applied to satellite observation data and extend 4DVarNet schemes introduced in [21] to represent surface height (SSH) fields, from altimetry and SST observations. Recent advances in physics-informed learning [21] for inverse problems appear more suited to jointly benefit from some physical advances in physics-informed learning [21] for inverse problems. Recent advances in physics-informed learning [21] for inverse problems appear more suited to jointly benefit from some physical advances in physics-informed learning [21] for inverse problems.

The scarce sampling to be dealt with as illustrated in Fig. 1 however limits the applicability of off-the-shelf image-to-image translation schemes [61]. Recent advances in physics-informed learning [21] for inverse problems appear more suited to jointly benefit from some physical advances in physics-informed learning [21] for inverse problems.

II. PROBLEM STATEMENT AND RELATED WORK

The space-time reconstruction of geophysical fields at sea surface from irregularly-sampled satellite observations can be stated as a data assimilation problem [19], that is to say the reconstruction of the dynamics of a geophysical state given some observation data. Classicaly, given some physical prior on the dynamics, data assimilation relies on a state-space formulation

\[
\begin{align*}
\frac{\partial x(t)}{\partial t} &= M(x(t)) + \eta(t) \\
y_m(t) &= H_m(x(t)) + \epsilon_m(t) \quad \forall t, m
\end{align*}
\]

where \(x\) is the space-time process of interest defined over a space domain \(\mathbb{D}\) and a time interval \([0, T]\). \(x\) is governed by dynamical model \(M\). \(y_m\) is the space-time observation data for observation modality \(m\). Observation operator \(H_m\) relates state \(x\) to observation \(y_m\). \(\eta\) and \(\epsilon_m\) refer to noise processes, which account for modeling and observation errors.

From this state-space formulation, we can state the reconstruction of state dynamics \(x\) from some observation data \(\{y_m\}_{m=1}^{M} = \{y_m(t_i)\}_{i=1}^{I}\) sampled at time steps \([t_i]_I\) as the minimization of the following variational criterion:

\[
U_\Phi(x, \{y_m\}_m) = \sum_m \lambda_m \sum_i \|y_m(t_i) - H_m(x(t_i))\|^2 + \gamma \sum_i \|x(t_i) - \Phi(x(t_i))\|^2
\]

where \(\Phi(x(t_i))\) is the time integration of dynamical model \(M\) over time interval \([t_{i-1}, t_i]\) from state \(x(t_{i-1})\). \(\lambda_m\) and \(\gamma\) are the weighting parameters of the observation and prior terms. This variational formulation is referred to as a weak-constrained 4DVarNet scheme in the data assimilation literature [19]. Numerically speaking, the minimization of variational cost (2) classically involves gradient-based solvers using adjoint approaches [19]. Within a discrete-time formulation of (1), ensemble and Kalman methods are also among the state-of-the-art schemes [19]. Model-driven reconstructions of sea surface dynamics from satellite-derived observations exploit these approaches with dynamical models given by ocean circulation models [1], [8]. We may point out that such approaches address the reconstruction of 3-D ocean state series, which include sea surface variables. It results in a much more complex problem that the reconstruction of the sole sea surface dynamics.
OI [15] is a particular case of the above formulation with the following additional assumptions: operators \( \{ \mathcal{H}_m \} \) are masking operators, dynamical prior \( \mathcal{M} \) is a linear operator and noise processes \( \eta \) and \( \{ \epsilon_m \} \) are Gaussian processes. Under these hypotheses, one can derive the analytical solution of the minimization of the discrete-time formulation of (2). The state-of-the-art altimeter-derived SSH product [49] relies on such an OI scheme. The resolved space-time scales are in the same range as those resulting from the assimilation of ocean circulation models [8], meaning that horizontal scales below 100 km cannot be retrieved in general.

When considering multimodal observation data as considered here, we may explore two different options from state-space formulation (1). On the one hand, we may include in state \( x \) all the observed geophysical parameters (here, both SSH and SST fields) such that the synergies between the geophysical parameters shall be accounted for through dynamical model \( \mathcal{M} \) or associated flow operator \( \Phi \). On the other hand, we may also explore observation models which could inform the relationship between the geophysical parameter to be reconstructed (here, the SSH) and the one which is observed (here, the SST). As an example, for SSH-SST synergies, the SQG framework [31], [34] provides a theoretical motivation to this second option. As detailed in Section III, we benefit from the proposed trainable framework to explore and compare both options.

As mentioned above, we can consider various optimization algorithms for the minimization of variational cost (2). Interestingly, when considering numerical implementations in deep learning and differentiable frameworks (e.g., pytorch) of all the operators in play in (2), we can benefit from the embedded automatic differentiation tools to implement gradient descent algorithms. This also opens avenues for considering pretrained operators without analytical derivations of adjoint operators. This has recently been explored for computational imaging and signal processing problems both with plug-and-play priors and pretrained observation operators [57], [60]. We may point out however that in such schemes, there is no guarantee for the pretrained operators to be fully-relevant for the considered inversion task. End-to-end learning approaches can address these shortcomings as one may learn an inverse model using some reconstruction performance metrics in the training loss [2], [26], [40]. Deep learning methods for space-time inpainting issues [33] do not apply directly given the very high missing data rates to be accounted for with ocean remote sensing data. Physics-driven learning schemes naturally arise as appealing approaches to benefit from prior physical knowledge on sea surface dynamics and associated satellite-derived observations. While one may complement classic end-to-end neural architecture with physics-informed training losses as illustrated in [26] for pansharpening applications, we here explore neural approaches which explicitly rely on a variational formulation similar to (2) [21], [35]. Such approaches make explicit the exploitation of an underlying state-space formulation. More specifically, as detailed in the next section, we extend our prior work [21] to multimodal inversion problems with a view to exploiting SSH-SST synergies for the reconstruction of sea surface dynamics.

### III. Proposed Approach

This section introduces the proposed multimodal learning-based inversion scheme. We first present the proposed multimodal data assimilation formulation before introducing the resulting end-to-end learning scheme.

#### A. Multimodal Data Assimilation Formulation

As detailed below, we benefit from the versatility of the end-to-end learning framework introduced in [21] and explore SST-SSH synergies to enhance the reconstruction of sea surface dynamics multimodal SST-SSH observation terms.

We first introduce formally the SSH-only and SSH-SST state-space formulations as follows.

1) **SSH-Only State-Space**: Here, state \( x \) in (1) only involves the SSH. Following [7], [22], SSH component \( x_{SSH} \) decomposes as a coarse-scale component \( \bar{x}_{SSH} \) and two fine-scale components \( \delta x_{SSH} \) and \( \delta x_{SSH} \). We assume to be provided with two altimeter-derived data sources: the irregularly-sampled altimeter-derived SSH data denoted as \( \bar{y}_{SSH} \) and an optimally-interpolated product from altimeter-derived data denoted as \( \delta y_{SSH} \). From (2), we derive the following matrix-based variational data assimilation formulation:

\[
U_{\Phi}(x, \bar{y}_{SSH}, \delta y_{SSH}) = \lambda_1 \left\| \bar{y}_{SSH} - \bar{x}_{SSH} \right\|^2 + \lambda_2 \left\| H(x_{SSH}) \cdot (\bar{y}_{SSH} - \bar{x}_{SSH} - \delta x_{SSH}) \right\|^2 + \gamma \left\| x - \Phi(x) \right\|^2 \tag{3}
\]

where \( x = (\bar{x}_{SSH}, \delta x_{SSH}, \delta x_{SSH}) \). \( H(x_{SSH}) \) is the masking operator to account for the sampling pattern of altimeter-derived SSH data. In this state-space formulation, the reconstructed SSH field is given by \( \bar{x}_{SSH} = \bar{x}_{SSH} + \delta x_{SSH} \). Component \( \delta x_{SSH} \) appears only in the observation term, whereas component \( \delta x_{SSH} \) only appears in the reconstruction equation. This parameterization is proven efficient to remove geometric artifacts associated with the sampling patterns of nadir altimeter data. We let the reader refer to [7] for an additional discussion on this point.

2) **SST-SSH State-Space**: Here, we benefit from the versatility of trainable schemes so that state variable \( x \) includes both a SSH component \( x_{SSH} \) and a SST component \( x_{SST} \). We consider the same parameterization as above for SSH component \( x_{SSH} \) such that \( \bar{x}_{SSH} = (\bar{x}_{SSH}, \delta x_{SSH}, \delta x_{SSH}) \). In this SST-SSH state-space formulation, we naturally complement variational formulation (3) with an additional SST-specific observation term such that

\[
U_{\Phi}(x, \bar{y}_{SSH}, \delta y_{SSH}) = \lambda_1 \left\| \bar{y}_{SSH} - \bar{x}_{SSH} \right\|^2 + \lambda_2 \left\| H(\bar{y}_{SSH}) \cdot (\bar{y}_{SSH} - \bar{x}_{SSH} - \delta x_{SSH}) \right\|^2 + \lambda_3 \left\| y_{SST} - \bar{x}_{SST} \right\|^2 + \gamma \left\| x - \Phi(x) \right\|^2 \tag{4}
\]

1Here, we will consider the operational processing referred to as DUACS which applies an OI to altimeter-derived data to deliver gap-free SSH fields [49].
In this formulation, prior term \(\|x - \Phi(x)\|^2\) will account for SSH-SST synergies. In both formulations, operator \(\Phi\) states the prior onto the state to be reconstructed. While in a model-driven configuration it derives from known governing equations for state \(x\), our previous work suggests that the parameterization of \(\Phi\) using state-of-the-art neural network architectures such as U-Nets might lead to better inversion performance as the form of the prior can adapt to the considered inversion problem and observation patterns [7], [21]. Overall, in these two state-space formulations, the trainable parameters refer to the trainable parameters of operator \(\Phi\) as well as weighing factors \(\lambda_{1,2,3}\). Regarding the parameterization of operator \(\Phi\), we follow the same approach as in [7] and [21]. We consider a two-scale residual U-Net architecture [13] with bilinear blocks to account for the non-linearities expected in upper ocean dynamics. Given \(T\)-day time windows and a \(W \times W\) spatial grid, state \(x\) for the SSH-only state-space (resp. the SST-SSH state-space), is given as a \((3 \times T) \times W \times W\) (resp. \((4 \times T) \times W \times W\)) tensor to apply 2-D convolution layers.

**Multimodal Observation Term:** As advocated by the SQG theory [31], [36], we also investigate a multimodal observation term to explicitly state that SST observations may inform SSH fields. We consider the following synergistic term added to (3) or (4) depending on the considered state-space:

\[
U_{MM}(x, y) = \|G_{MM}^1(y) - G_{MM}^2(x)\|^2
\]

with \(G_{MM}^1\) and \(G_{MM}^2\) convolutional operators acting, respectively, on SST observation \(y_{SST}\) and state \(x\). We may remind that \(y_{SST}\) and \(x\) refer to space-time tensors. The SQG theory would lead to parameterize as \(G_{MM}^1\) as a passband filter and \(G_{MM}^2\) as a combination of a passband filter and of a fractional Laplacian operator such that (5) would lead to

\[
\left\|\mathcal{F}^1 \ast y_{SST} - \mathcal{F}^2 \ast (\Delta^{1/2} (\tilde{x}_{SSH} + \delta x^k_{SSH}))\right\|^2
\]

with \(\mathcal{F}^{1,2}\) linear passband filters to select the scale range to which the SQG theory applies and \(\Delta^{1/2}\) the fractional Laplacian operator which is a linear filter defined in the spectral domain. Overall, this parameterization results in the extraction of linear features to match SST and SSH patterns. Here, we investigate a generalization of the SQG-based parameterization, where operators \(G_{MM}^1, G_{MM}^2\) are trainable linear or nonlinear operators. As detailed hereafter, we consider simple ConvNets with a single layer in the linear case and four layers with tanh activations in the nonlinear case. Importantly, whereas the SQG theory involves space-only filters, our trainable operators may exploit space-time filters. One may exploit this multimodal observation term both in the SSH-only state-space formulation and in the SSH-SST one.

**B. End-to-End Inversion Model and Associated Learning Scheme**

From the proposed multimodal variational formulation with trainable components, we design an end-to-end inversion scheme which implements a gradient-based iterative solver as proposed in [21]. For a given state-space formulation and associated variational cost \(U_\phi\), the proposed end-to-end neural architecture performs a predefined number of iterations of an iterative gradient-based update. More precisely, at iteration \(k\), we apply

\[
\begin{align*}
    h^{k+1} & = S \left[ \nabla_x U \left( x^{(k)}, [y_m]^n \right), h^{(k)}, c^{(k)} \right] \\
    x^{(k+1)} & = x^{(k)} - \mathcal{L}(h^{(k+1)})
\end{align*}
\]

where \(h^{(k)}\) and \(c^{(k)}\) are the internal states of LSTM cell \(S\) at iteration \(k\) and \(x^{(k+1)}\) the updated state. \(\mathcal{L}\) is a linear layer to map the LSTM state to the space spanned by state \(x\). This gradient-based iterative update is similar to neural parameterization for the learning of optimizers [30]. As state \(x\) is implemented as a multivariate 2-D tensor, we consider 2-D convolutional LSTM cells. Experimentally, we cross-validated the use of 150-D convolutional LSTM cells.

Overall, the resulting end-to-end architecture uses as inputs observation data \(y_1 = y_{SST}, y_2 = y_{SSH}\) and \(y_3 = y_{SST}\) as well as some state initialization \(x^{(0)}\) to output the reconstructed state. This architecture implements a predefined number \(K\) of the above gradient steps, typically from 5 to 15. We denote by \(\hat{x} = \Psi_{\theta}(x^{(0)}, [y_m]^n)\) the output of the end-to-end inversion scheme after \(K\) iterations of (7) (i.e., \(\hat{x} = x^{(K)}\)) where \(\theta\) stands for the set of all trainable parameters. Depending on the considered multimodal configuration, the trainable components of the architecture comprise those of prior \(\Phi\) and of the LSTM-based solver, possibly complemented by those of multimodal observation operators \(G_{MM}^{1,2}\).

We exploit a supervised learning strategy to train our model. Given a training dataset comprising triplets of true states, observation data, and initial conditions \(\{x_{SSH,n}^{true}, [y_m,n]^m, x_{SSH,n}^{(0)}\}_n\), the training loss typically involves a weighted sum of the reconstruction error for reconstructed state \(x_{SSH}\) and its gradient

\[
\begin{align*}
    \mathcal{L}_r & = \sum_n \|x_{SSH,n}^{true} - \hat{x}_{SSH,n}\|^2 \\
    \mathcal{L}_v & = \sum_n \|\nabla x_{SSH,n}^{true} - \nabla \hat{x}_{SSH,n}\|^2.
\end{align*}
\]

As proposed in [21], we also include additional regularization terms

\[
\mathcal{L}_p = \sum_n \|x_{n}^{true} - \Phi(x_{n}^{true})\|^2 + \sum_i \|\tilde{x}_n - \Phi(\tilde{x}_n)\|^2
\]

such that the overall training loss is computed as a weighted sum of these different terms: \(v_r \mathcal{L}_r + v_v \mathcal{L}_v + v_\phi \mathcal{L}_p\) with \(v_r, v_v\) and \(v_\phi\) weighing parameters. The training procedure exploits Adam optimizer with a 1e-3 learning rate over 400 epochs. We select the best model according to reconstruction metrics evaluated on the validation dataset at each training epoch. The Pytorch code of our implementation, including all parameter values, is available along with trained models [23].

**IV. RESULTS**

This section details the numerical experiments we run to evaluate the reconstruction performance of the proposed approach. We first detail the considered experimental setting.
We then present our results, including a comparison to state-of-start schemes.

A. Dataset and Experimental Setting

We evaluate the proposed multimodal inversion framework for the space-time interpolation of SSH fields, which relate to the geostrophic component of sea surface velocities,\(^2\) using the benchmarking setting introduced in [37]. It relies on an OSSE for nadir and wide-swath satellite altimetry data. We exploit a one-year NATL60 numerical simulation dataset [3] for a 10° × 10° area along the Gulf Stream from October 2012 to September 2013 with a daily time resolution and a 1/20° spatial resolution. Nadir altimetry data involves the space-time sampling of a real four-altimeter configuration, whereas wide-swath SWOT data rely on a SWOT simulator [27]. In both cases, we consider noise-free observations. For SST observations, we assume gap-free observations of daily SST fields. During the training stage, we consider SST observations with a 1/20° spatial resolution. For evaluation purposes, we also investigate subsampled versions with resolutions of 1/10°, 1/5°, 1/4°, and 1/2°.

Regarding the training procedure and the evaluation framework, we split the OSSE dataset into training, validation, and test datasets as follows. We use as training data the data from February 2013 to September 2013 and as validation dataset data from January 2012. We use the validation dataset to monitor performance metrics and select the best model during the training procedure. The overall evaluation procedure relies on performance metrics computed for the test dataset which refers to the 40-day period from 22 October 2012 to 2 December 2012. As evaluation metrics, we first consider the metrics introduced in [37] to benchmark the proposed schemes with respect to the state-of-the-art approaches\(^3\).

1) \(\mu\), the normalized root-mean-square-error-based metrics equals 1 for a perfect reconstruction.
2) \(\lambda_t\), the minimum time scale resolved in days.
3) \(\lambda_s\), the minimum spatial scale resolved in degrees.

As described in [37], the last two metrics are computed in the spectral domain. With a view to enhancing the differences between the different configurations of the proposed multimodal inversion scheme, we also evaluate the relative improvement with respect to the operational SSH baseline (DUACS) [49] using the following three metrics.

1) \(\tau_{SSH}\), the relative gain with respect to DUACS baseline for the MSE of the reconstruction of the SSH.
2) \(\tau_{\nabla SSH}\), the relative gain with respect to DUACS baseline for the MSE of the reconstruction of the gradient of the SSH.
3) \(\tau_{\Lambda SSH}\), the explained variance for the Laplacian of the reconstructed SSH fields.

\(^2\)Geostrophic sea surface currents can be approximated from the gradient of SSH fields. For an area such as the one considered in our experiments, this results in the reconstruction of the divergence-free component of sea surface velocities.

\(^3\)We refer the reader to the following link for the detailed presentation of the evaluation experiment and benchmarked approaches https://github.com/ocean-data-challenges/2020a_SSH_mapping_NATL60.

We compute all these metrics from the reconstruction error of the SSH fields with respect to the daily-averaged SSH fields of the numerical simulations.

B. Impact of the Parameterization of the Multimodal Framework

We first investigate how the parameterization of the proposed multimodal framework affects the reconstruction performance. It includes both the parameterization of the multimodal observation term (5) as well as complementing or not the reconstructed state with the SST (1). Regarding multimodal observation term (5), we consider both linear and nonlinear parameterizations for operators \(G_{1,2}^{MM}\).

1) Linear Parameterization: In the linear setting, we exploit linear \(7 \times 3 \times 3\) space-time convolution kernels for operators \(G_{1,2}^{MM}\) applied to SST observations. Similarly, operators \(G_{1,2}^{MM}\) extract linear features from state \(x\) with \(3 \times 3\) kernels. We vary the number of extracted features from 1 to 50.

2) Nonlinear Parameterization: In the nonlinear setting, operators \(G_{1,2}^{MM}\) are convolutional networks with four layers and tanh activation functions. We also tested classic ReLu activations which led to slightly worse reconstruction performance. We also vary the number of extracted features from 1 to 50.

In Table I, we report the synthesis of the reconstruction performance of the different configurations of the proposed multimodal framework (5). For the configurations with multimodal observation terms (5), we report the performance of the best configuration, here a 20-D nonlinear observation operator. We further analyze below how this parameterization affects the performance (see Table II). As a comparison baseline, we consider the 4DVarNet scheme using only altimetry data and no SST observations. Overall, all multimodal configurations clearly outperform the altimetry-only baseline with a significant gain up to 8% to 16% in terms of the reconstruction of the SSH and its derivatives. The greatest improvement occurs for the resolved time scale (up to 2.47 versus 5.30 days). These results point out that the best performance comes from a SSH-only state-space where we exploit SST data in a trainable multimodal observation term (5). This suggests that we better account for SSH-SST synergies through multimodal observation term (5) than through a U-Net parameterization for prior \(\Phi\) in the SST-SSH state-space formulation. The lower performance of the latter combined with a multimodal observation term (5) may relate to overfitting issues as this parameterization is more complex. Larger training datasets and data augmentation may provide relevant solutions to overcome these issues.

We further analyze the sensitivity of the reconstruction performance with respect to the parameterization of multimodal observation term (5). As reported in Table II, we vary the dimensionality of the observation term for both linear and nonlinear observation terms from 1-D operators to 50-D ones. Interestingly, 1-D parameterizations already lead to very good performance. Whereas the performance is very similar from 1-D to 5-D linear operators, more complex linear operators
TABLE I
RECONSTRUCTION PERFORMANCE OF THE PROPOSED 4DVarNet FRAMEWORK FOR DIFFERENT CONFIGURATIONS: WHEN THE STATE IN (1) ONLY RELATES TO SSH, THE MULTIMODAL COMPONENT COMES FROM MULTIMODAL (MM) TERM (5); WE ALSO CONSIDER A MULTIMODAL SETTING WHERE THE STATE IN (1) COMPRISES BOTH SSH AND SST. IN THE LATTER CONFIGURATION, ONE MAY USE OR NOT THE MULTIMODAL TERM. WE REFER THE READER TO THE MAIN TEXT FOR THE DEFINITION OF THE PERFORMANCE METRICS. WE HIGHLIGHT IN BOLD THE BEST SCORE

| State | MM term (5) | μ   | λₐ (°) | λₜ (days) | τSSH | τₓSSH | τΔSSH |
|-------|-------------|-----|--------|-----------|------|-------|-------|
| SSH   | No          | 0.96| 0.67   | 5.3       | 71.4%| 65.1% | 83.7% |
|       | Yes         | 0.97| 0.50   | 2.47      | 83.8%| 81.4% | 91.8% |
| SST+SSH| No          | 0.96| 0.60   | 3.18      | 75.7%| 73.9% | 88.4% |
|       | Yes         | 0.96| 0.57   | 2.49      | 78.7%| 75.9% | 89.2% |

lead to a poorer performance which indicates some overparameterization. By contrast, when considering nonlinear operators, we observe an increasing performance trend from a 1-D operator to a 20-D one. This likely reflects the ability of the nonlinear operators to learn relevant specific SST features. As mentioned above, we consider hyperbolic tangent activation functions. When considering other activation functions such as ReLu activations, we worsen the reconstruction performance. This likely relates to the regularity of the higher-order derivatives of geophysical fields which are rather low-contrast fields compared to natural images.

C. Comparison to State-of-the-Art Schemes

Based on the previous experiments, we synthesize the performance of the best configurations of the proposed multimodal framework with respect to that of state-of-the-art schemes according to the benchmarking framework presented in [37]. As detailed in the associated data challenge, we first include approaches which only rely on altimetry data. They include the operational optimal-interpolation-based method (DUACS) [49], model-driven interpolations using variational data assimilation schemes [37], [52] and a multiscale interpolation approach [51]. We also implement a SQG-based inversion schemes to complement the optimally-interpolated SSH fields for horizontal scales below 1.2°. The last category of approaches we consider in our benchmarking experiment refers to direct learning-based inversion schemes. Here, we train U-Net architectures [13] to reconstruct the SSH fields from gappy data using an initial zero-filling strategy. We consider both a SSH-only and SSH-SST configurations to assess the relevance of 4VarNet architectures. Regarding the parameterization of the U-Nets, we evaluate both the U-Net architecture considered for prior \( \Phi \) in the implemented 4DVarNet schemes as well as a standard three-scale U-Net architecture with ReLu activations [13]. We only report the results for the former parameterization which led to the best reconstruction performance.

We report in Table III the synthesis of the performance metrics for all the benchmarked approaches. Among all the methods using only altimetry data, the 4DVarNet scheme clearly leads to the best reconstruction performance with a very significant gain with respect to the operational OI baseline greater than 50% for MSE scores and 40% for the resolved space-time scale. Multimodal 4DVarNet schemes further improve the reconstruction score. We obtain the greatest improvement for the resolved time scale (2.47 versus 5.30 days for the altimeter-only 4DVarNet and 11.15 days for the operational baseline) and the SSH gradient (81.4% versus 65.1% for the altimeter-only 4DVarNet regarding the relative gain with respect to the operational baseline). We also observed some improvement though smaller for the resolved spatial scale (0.50° versus 0.68° and 1.22°). These results also highlight the benefits of 4DVarNet schemes compared to the application of state-of-the-art image-to-image neural architectures. Hence, they support the relevance of the proposed variational formulation to make explicit underlying observations and prior operators. Especially, the 4DVarNet scheme using SSH-only data delivers better reconstruction metrics than U-net-based inversion using jointly SSH and SST data. The latter leads to a very high value of the resolved time scale which indicates the presence of artifacts in the associated interpolation.

As an illustration, we display in Fig. 2 the norm of the gradient of the reconstructed SSH fields. Visually, the operational processing [49] leads to a blurry reconstruction as stressed by the resolved spatial scale above 1°. By contrast, all 4DVarNet schemes lead to much sharper gradients which are more similar to those of the true SSH fields. We draw similar conclusions from the reconstructed Laplacian fields. This example also provides a clear illustration of the added value of SSH-SST synergies. When considering the SSH-only 4DVarNet scheme, we cannot perfectly recover the geometry of the main meander (see zoom on the upper-left region in Figs. 3 and 4) as well as the orientation of the large eddy South of the main meander due to the scarce sampling of the altimeter data. We may notice that the SST field in Fig. 1 clearly reveals these geometrical patterns. Interestingly,
multimodal 4DVarNet schemes successfully extract these features to improve the reconstruction of the SSH field. Zooms in Figs. 3 and 4 further emphasize the better performance of the 4DVarNet scheme with a nonlinear multimodal observation term as well as local checkboard artifacts generated by a direct inversion using U-Nets. The latter suggests that the scarce sampling of the altimetry data impedes the generalization performance of such direct learning-based inversion schemes. We also visualize the SST features learned by the best linear and nonlinear multimodal 4DVarNet schemes [resp. using 3-D and 20-D multimodal observation terms (5)]. All features enhance fine-scale patterns. Nonlinear features involve even

**TABLE II**

Reconstruction Performance for Different Parameterizations of Multimodal (MM) Term (5): We report the Performance Metrics of the Proposed 4DVarNet Framework for Linear and Nonlinear Operators in (5) With 1, 2, 3, 5 or 10 Multimodal Features ($N = \{1, 2, 3, 5, 10\}$). We consider the same Performance Metrics as in Table III

| MM term (5) | $N_{Feat}$ | $\mu$ | $\lambda_x$ (°) | $\lambda_t$ (days) | $\tau_{SSH}$ | $\tau_{VSSH}$ | $\tau_{\Delta SSH}$ |
|-------------|------------|-------|-----------------|-------------------|-------------|-------------|-----------------|
| Linear      | 1          | 0.96  | 0.50            | 2.99              | 79.6%       | 78.0%       | 90.9%           |
|             | 2          | 0.97  | 0.60            | 3.12              | 82.0%       | 78.7%       | 90.5%           |
|             | 3          | 0.97  | 0.57            | 3.14              | 80.9%       | 79.1%       | 91.3%           |
|             | 5          | 0.97  | 0.60            | 2.50              | 80.3%       | 76.7%       | 89.5%           |
|             | 10         | 0.96  | 0.75            | 29.4              | 79.0%       | 75.6%       | 89.1%           |
|             | 20         | 0.97  | 0.61            | 3.20              | 81.3%       | 77.5%       | 89.9%           |
|             | 50         | 0.96  | 0.76            | 29.8              | 77.3%       | 74.1%       | 88.7%           |
| Non-linear  | 1          | 0.97  | 0.60            | 3.23              | 81.5%       | 77.8%       | 90.3%           |
|             | 2          | 0.97  | 0.59            | 3.31              | 81.6%       | 77.5%       | 89.7%           |
|             | 3          | 0.97  | 0.55            | 2.63              | 81.3%       | 78.6%       | 90.6%           |
|             | 5          | 0.97  | 0.58            | 3.12              | 81.1%       | 78.0%       | 90.0%           |
|             | 10         | 0.97  | 0.54            | 2.46              | 81.5%       | 79.5%       | 91.1%           |
|             | 20         | 0.97  | 0.50            | 2.47              | 83.8%       | 81.4%       | 91.8%           |
|             | 50         | 0.97  | 0.53            | 2.49              | 82.9%       | 79.5%       | 91.0%           |

**TABLE III**

Synthesis of the Reconstruction Performance of the Benchmarked Approaches: We report the Performance Metrics of the Benchmarked Approaches for the Reconstruction of Image Time Series of Sea Surface Currents From Satellite Data. We Refer the Reader to the Main Text for the Description of the Different Metrics. We Highlight in Bold the Best Score

| Approach      | Data used       | $\mu$ | $\lambda_x$ (°) | $\lambda_t$ (days) | $\tau_{SSH}$ | $\tau_{VSSH}$ | $\tau_{\Delta SSH}$ |
|---------------|-----------------|-------|-----------------|-------------------|-------------|-------------|-----------------|
| DUACS [49]    | SSH only        | 0.92  | 1.22            | 11.15             | 0%          | 0%          | 44.9%           |
| DYMOST [52]   | SSH only        | 0.93  | 1.20            | 10.07             | -           | -           | -               |
| MIOST [51]    | SSH only        | 0.94  | 1.18            | 10.14             | -           | -           | -               |
| BFN [37]      | SSH only        | 0.93  | 0.8             | 10.09             | -           | -           | -               |
| SQG [31]      | SSH and SST     | 0.93  | 1.12            | 11.16             | -           | -           | -               |
| U-Net         | SSH only        | 0.94  | 1.21            | 10.21             | 38.4%       | 36.2%       | 70.7%           |
|               | SSH and SST     | 0.95  | 1.09            | 37.0              | 54.6%       | 56.6%       | 79.6%           |
| 4DVarNet      | SSH only        | 0.96  | 0.67            | 5.3               | 71.4%       | 65.1%       | 83.7%           |
| (ours)        | SSH-SST-L       | 0.97  | 0.57            | 3.14              | 80.9%       | 79.1%       | 91.3%           |
|               | SSH-SST-NL      | 0.97  | 0.50            | 2.47              | 83.8%       | 81.4%       | 91.8%           |
Fig. 2. Comparison of the gradient of the reconstructed SSH fields on 25 October 2012. (Left to right) and (Top to bottom) We compare the map of the norm of the gradient of the true SSH field. (Top left) To those of interpolated SSH fields for benchmarked schemes, namely DUACS product [49], an SQG-based inversion [31], direct end-to-end neural inversion using U-Nets [13] with altimetry-only data (U-Net-SSH-only) and multimodal input data (U-Net-SSH-SST), a 4DVarNet scheme using altimetry-only data (4DVarNet-SSH-only) [7] and proposed multimodal 4DVarNet schemes using a multimodal state (4DVarNet-MMState) and multi-modal observation terms with linear operators (4DVarNet-MMObs-Lin) and nonlinear ones (4DVarNet-MMObs-NonLin).

We use the same color bar for all fields. We let the reader refer to the main text for the description of the different schemes. The red box in the top left field refers to the subarea which we zoomed-in view in Figs. 3 and 4.

finer-scale patterns and seem to reveal a greater diversity of patterns. For instance, some features exhibit the large eddy South of the main meander while others do not. In line with the SQG theory [31], [36], it seems that we can interpret the learned SST feature extraction step as a combination of passband filters and template-driven detection filters.

D. Impact of the Resolution of SST Observations

We further evaluate how the resolution of SST observations affects the reconstruction of the SSH. We may remind that satellite-derived microwave SST data typically lead to relatively low-resolution data with a typical 1/4° horizontal resolution, but almost gap-free observations on a daily scale [17], [43]. By contrast, infrared satellite sensors lead to much higher-resolution observations even below 1/100°, however at the expense of possibly large missing data rates due to the sensitivity to the cloud coverage [17], [43]. Here, using the trained multimodal model with gap-free 1/20° SST fields, we assess the reconstruction performance when providing during the evaluation procedure with coarsened versions of the...
Fig. 3. Zoom on reconstructed SSH gradient fields on 25 October 2012. We display zooms for (top left) subregion of the fields displayed in Fig. 2.

SST data. The coarsening proceeds as follows. We apply an average pooling by a factor ranging from 2 to 10 followed by a linear interpolation onto the original 1/20°. As such, we simulate SST pseudo-observations with different resolutions from 1/20° to 1/2. We synthesize the associated reconstruction performance in Table IV. As expected, the lower the resolution of the SST, the lower the reconstruction performance. This indicates that all spatial scales in SST fields from 1/20° to 1/2 contribute to inform fine-scale SSH patterns. This is in line with previous studies based on the SQG theory [31], which considered a scale-invariant hypothesis. Interestingly, up to 1/4°, we report a very significant improvement with respect to the altimeter-only baseline (e.g., 4.00 versus 5.30 days for the resolved time scale). These results support the potential application of the proposed framework to operational satellite-derived L4 SST products [43], which typically resolve horizontal scales between 1/20° and 1/4° on a daily resolution.

Future work could explore how gappy SST observations would affect the reconstruction performance. In this respect, we could consider the SSH-SST state-space formulation with multimodal observation terms to address the joint interpolation of SST and SSH fields. We could also explore multimodal configuration with different SST data sources for instance from microwave and infrared satellite sensors. As pointed out above, one may consider larger training datasets to overcome overfitting issues.

V. DISCUSSION

This article has introduced a novel multimodal learning-based inversion framework for the reconstruction of space-time sea surface dynamics from irregularly-sampled multisource
Fig. 4. Zoom on reconstructed SSH Laplacian fields on 25 October 2012. We depict zooms for (Top left) subregion of the gradient norm fields displayed in Fig. 2.

Table IV

| SST resolution | µ    | λσ (°) | λτ (days) | τSSH | τσSSH | τΔSSH |
|----------------|------|--------|-----------|------|-------|-------|
| 1/20°          | 0.97 | 0.50   | 2.47      | 83.8%| 81.4% | 91.8% |
| 1/10°          | 0.97 | 0.55   | 2.54      | 82.0%| 78.1% | 90.1% |
| 1/5°           | 0.96 | 0.67   | 3.87      | 77.7%| 72.2% | 87.0% |
| 1/4°           | 0.96 | 0.68   | 4.00      | 75.7%| 69.8% | 85.5% |
| 1/2°           | 0.95 | 0.88   | 5.65      | 61.5%| 53.1% | 77.3% |

satellite data. Reported numerical experiments support its relevance compared to the state-of-the-art approaches to exploit SST-SSH synergies and improve the reconstruction of finer-scale geostrophic sea surface dynamics. We discuss in this section how the proposed framework relates to and complements previous works according to three different
Fig. 5. Learned SST features extracted on 25 October 2012. We depict the output of operator $G^{MM}_{\text{SST}}$ in (5) computed for the seven-day SST field sequence centered on 25 October 2012 for the best 4DVarNet schemes with multimodal observation terms (5) using (top) linear and (bottom) nonlinear operators. From Tables I and II, we select a 3-D linear configuration and a 20-D nonlinear one.

Aspects: computational imaging for geoscience, deep learning for data assimilation, deep learning, and spaceborne Earth observation. We also discuss limitations and future work for the reconstruction of sea surface dynamics.

A. Computational Imaging for Geoscience

End-to-end learning strategies have become the state-of-the-art approaches for a variety of computational imaging problems, including among others denoising [57], [60], super-resolution [16] and inpainting issues [38], [58]. While numerous applications to the observation and monitoring of geophysical processes exploit state-of-the-art deep learning schemes, the underlying physical laws naturally advocate for the design of physics-aware approaches. This is particularly true for space-time interpolation issues with very high missing data rates as targeted in our study. We may also point out that video inpainting remain a challenge for deep learning [33]. Here, we exploit a classic inverse problem formulation to design our neural architecture. This allows us to make explicit the definition of the observation operators and the dynamical...
prior. As supported by our numerical experiments, the former is key to fully exploit SSH-SST synergies to improve the reconstruction of sea surface dynamics at finer scales. The proposed multimodal 4DVarNet scheme also relates to deep unfolding schemes [41]. Our neural architecture implements an iterative gradient descent of the trainable variational cost. Rather than considering a reaction-diffusion formulation [11] or an optimization scheme associated with proximal operators [35], we exploit the embedded automatic differentiation of the variational cost with a trainable gradient-based solver to speed up the optimization. This strategy also allows us to train jointly linear and nonlinear observation operators with the dynamical prior and the solver. This is expected to contribute to reducing inversion biases [21], [29], [39] and improving the overall interpretability of the neural architecture.

B. Deep Learning and Data Assimilation

The proposed framework can be regarded as a neural resolution of a variational data assimilation formulation. While data assimilation schemes [19] are widely used in geoscience for the reconstruction of space-time dynamics, they require the explicit knowledge of the underlying dynamics and of the observation operators in (5). In our case-study, the classical choice would be to consider an ocean general circulation model [8], [19] and identity observation operators with covariance priors. This results in a much more computationally-demanding inversion scheme compared with our approach. Besides, such schemes implemented in operational systems typically lead to reconstruction performance for sea surface dynamics in the same range as the operational baseline considered in our experiments [49]. Our study illustrates the versatility of deep learning frameworks to define a state-space formulation that only comprises the variables of interest (here, the SSH possibly complemented by the SST). The significant improvement over model-driven approaches such as [37], [51] stresses the ability and relevance to learn the underlying variational representation from data. This seems particularly appealing to explore multimodal synergies between different geophysical tracers, when the derivation of physical laws to relate the processes of interest reveal complex. Hence, the proposed framework may also contribute to the identification of such laws from data as the calibration of the observation operators is a by-product of the training process.

C. Deep Learning and Ocean Remote Sensing

Spaceborne Earth observation and ocean remote sensing greatly benefit from deep learning advances. Given orbiting characteristics as well as the sensitivity of satellite sensors to the atmospheric conditions, spaceborne Earth observation data often result in an irregular space-time sampling which may involve very large missing data rates as illustrated here for satellite altimetry data. While OI schemes remain the state-of-the-art processing for a wide range of operational gap-free satellite-derived products [17], [49], deep learning schemes emerge as relevant approaches [6], [20], [28]. This study further supports their relevance to best exploit multimodal synergies which are not easily accounted for with OI schemes. The ever-increasing availability of observation data and numerical simulations also greatly contribute to the development and evaluation of learning-based and data-driven approaches as illustrated by the considered experimental setting based on an open data challenge.2 We could apply and extend the proposed framework to other space-time geophysical products such as OC [46], [55], sea surface turbidity [44], [54], sea and land surface temperature [6], sea surface currents [12], [43]. Future challenges also involve joint calibration and interpolation issues for future satellite missions [24] as well as multimodal synergies between satellite data and other remote sensing and in situ data sources such as drifters [5], [48], underwater acoustics data [9], moored buoys [25], Argo profilers [14], [18]. Especially, the latter might provide new ways to better monitor the interior of the ocean which cannot be directly observed from space.

D. Limitations and Future Work for SSH Mapping

This study supports the relevance of multimodal 4DVarNets for the space-time reconstruction of SSH fields from SST-SSH synergies. Our numerical experiments rely on an idealized simulation-based case-study. Especially, we do not account for noisy observations. Real altimetry observations involve both observation noises and high-frequency fine-scale geophysical signals, such as internal tides and internal gravity waves [4], [59]. Numerous studies support the potential robustness of deep learning schemes to noisy patterns, when the training dataset involves appropriate noise simulations [47]. The availability of realistic tide-resolving submesoscale-permitting ocean simulations provides the basis to address these issues in a future work. The extension of the considered state-space formulation could also open new directions for the separation of tide-related and tide-free SSH signals. Regarding SST data, future work could account for more realistic simulations of available SST products. This could include experiments with simulation of gap-free interpolated SST products as well as the extension of the proposed scheme to the joint interpolation of SSH and SST fields. Both aspects appear as a relatively direct extension of the proposed scheme. The generalization from the specific dynamical regime observed in the Gulf stream region to the global ocean appears as a more complex challenge. Although we do not expect the considered parameterization to apply on a global scale, future work could explore the combination of region-specific models as considered in the operational parameterization of global-scale OI products [49]. Another strategy could investigate more complex parameterization of observation and prior terms in (3)–(5) to better reflect the variety of dynamical regimes at sea surface on a global scale. Attention-based architectures [53] seem particularly appealing in this context to introduce multiregion and/or multiregime parameterizations. We expect the availability of our code and trained models [23] to support future studies addressing these challenges.

REFERENCES

[1] S. Abdalla et al., “Altimetry for the future: Building on 25 years of progress,” Adv. Space Res., vol. 68, pp. 319–363, Mar. 2021.

[2] https://github.com/ocean-data-challenges/2020a_ssh_mapping_NATL60
[2] H. K. Aggarwal, M. P. Mani, and M. Jacob, “MoDL: Model-based deep learning architecture for inverse problems,” IEEE Trans. Med. Imag., vol. 38, no. 2, pp. 394–405, Feb. 2019.

[3] A. Ajayi et al., “Spatial and temporal variability of the North Atlantic eddy field from two kilometric-resolution ocean models,” J. Geophys. Res., Oceans, vol. 125, no. 5, May 2020, Art. e2019JC015827.

[4] B. K. Arbic, A. J. Wallcraft, and E. J. Metzger, “Concurrent simulation of the eddying general circulation and tides in a global ocean model,” Ocean Model., vol. 32, nos. 3–4, pp. 175–187, 2010.

[5] G. Baaklini et al., “Blending drifters and altimetric data to estimate surface currents: Application in the levantine Mediterranean and objective validation with different data types,” Ocean Model., vol. 166, Oct. 2021, Art. no. 101850.

[6] A. Barth, A. Alvera-Azcárate, M. Licer, and J.-M. Beckers, “DINCAE 1.0: A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations,” GeoSci. Model Develop., vol. 13, no. 3, pp. 1609–1622, Mar. 2020.

[7] M. Beauchamp, Q. Febvre, H. Georgenthum, and R. Fablet, “End-to-end neural interpolation of satellite altimetry data using 4DVarNet schemes,” GeoSci. Model Develop., vol. 16, no. 8, pp. 2191–2147, Apr. 2023.

[8] M. Benkiran et al., “Assessing the impact of the assimilation of SWOT observations in a global high-resolution analysis and forecasting system part 1: Methods,” Frontiers Mar. Sci., vol. 8, Jul. 2021, Art. no. 691955.

[9] D. Cazeau, J. Bonnell, and M. Baumgärtner, “Wind speed estimation using acoustic underwater glider in a near-shore marine environment,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 4, pp. 2097–2106, Apr. 2019.

[10] D. B. Chelton, J. C. Ries, B. J. Haines, L.-L. Fu, and P. S. Callahan, “Satellite altimetry,” in Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications (International Geophysics), vol. 69, A. Cazenave and L.-L. Fu, Eds. New York, NY, USA: Academic, 2001.

[11] Y. Chen, W. Yu, and T. Pock, “On learning optimized reaction diffusion processes for effective image restoration,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 5261–5269.

[12] D. Ciani, E. Charles, B. B. Nardelli, M.-H. Rio, and R. Santoleri, “Ocean currents reconstruction from a combination of altimeter and ocean colour data: A feasibility study,” Remote Sens., vol. 13, no. 12, p. 2389, Jun. 2021.

[13] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning dense volumetric segmentation from sparse annotation,” in Proc. MICCAI, 2016, pp. 424–432.

[14] G. Cossarini et al., “Towards operational 3D-VAR assimilation of chlorophyll biogeochemical-argo float data into a biogeochemical model of the Mediterranean Sea,” Ocean Model., vol. 133, pp. 112–128, Jan. 2019.

[15] N. Cressie and C. K. Wikle, Statistics for Spatio-Temporal Data. Hoboken, NJ, USA: Wiley, 2015.

[16] C. Dong, C. K. Loy, and Y. LeCun, “Deep learning for image denoising,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, Feb. 2015.

[17] C. K. Dong, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and O. Ronneberger, “Faster R-CNN: Towards real-time object detection with region proposals,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 6378–6387.

[18] C. K. Dong, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and O. Ronneberger, “Image super-resolution using high-pass residual learning,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 5261–5269.

[19] G. Lapeyre, J. C. Ries, B. J. Haines, L.-L. Fu, and P. S. Callahan, “Wind speed estimation using acoustic underwater glider in a near-shore marine environment,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 4, pp. 2097–2106, Apr. 2019.

[20] D. B. Chelton, J. C. Ries, B. J. Haines, L.-L. Fu, and P. S. Callahan, “Satellite altimetry,” in Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications (International Geophysics), vol. 69, A. Cazenave and L.-L. Fu, Eds. New York, NY, USA: Academic, 2001.

[21] Y. Chen, W. Yu, and T. Pock, “On learning optimized reaction diffusion processes for effective image restoration,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 5261–5269.

[22] D. Ciani, E. Charles, B. B. Nardelli, M.-H. Rio, and R. Santoleri, “Ocean currents reconstruction from a combination of altimeter and ocean colour data: A feasibility study,” Remote Sens., vol. 13, no. 12, p. 2389, Jun. 2021.

[23] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning dense volumetric segmentation from sparse annotation,” in Proc. MICCAI, 2016, pp. 424–432.

[24] G. Cossarini et al., “Towards operational 3D-VAR assimilation of chlorophyll biogeochemical-argo float data into a biogeochemical model of the Mediterranean Sea,” Ocean Model., vol. 133, pp. 112–128, Jan. 2019.

[25] N. Cressie and C. K. Wikle, Statistics for Spatio-Temporal Data. Hoboken, NJ, USA: Wiley, 2015.

[26] C. Dong, C. K. Loy, and Y. LeCun, “Deep learning for image denoising,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, Feb. 2015.

[27] C. K. Dong, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and O. Ronneberger, “Faster R-CNN: Towards real-time object detection with region proposals,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 5261–5269.

[28] G. Lapeyre, J. C. Ries, B. J. Haines, L.-L. Fu, and P. S. Callahan, “Wind speed estimation using acoustic underwater glider in a near-shore marine environment,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 4, pp. 2097–2106, Apr. 2019.

[29] D. B. Chelton, J. C. Ries, B. J. Haines, L.-L. Fu, and P. S. Callahan, “Satellite altimetry,” in Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications (International Geophysics), vol. 69, A. Cazenave and L.-L. Fu, Eds. New York, NY, USA: Academic, 2001.

[30] Y. Chen, W. Yu, and T. Pock, “On learning optimized reaction diffusion processes for effective image restoration,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 5261–5269.

[31] D. Ciani, E. Charles, B. B. Nardelli, M.-H. Rio, and R. Santoleri, “Ocean currents reconstruction from a combination of altimeter and ocean colour data: A feasibility study,” Remote Sens., vol. 13, no. 12, p. 2389, Jun. 2021.

[32] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning dense volumetric segmentation from sparse annotation,” in Proc. MICCAI, 2016, pp. 424–432.

[33] G. Cossarini et al., “Towards operational 3D-VAR assimilation of chlorophyll biogeochemical-argo float data into a biogeochemical model of the Mediterranean Sea,” Ocean Model., vol. 133, pp. 112–128, Jan. 2019.

[34] N. Cressie and C. K. Wikle, Statistics for Spatio-Temporal Data. Hoboken, NJ, USA: Wiley, 2015.

[35] C. Dong, C. K. Loy, and Y. LeCun, “Deep learning for image denoising,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, Feb. 2015.

[36] C. K. Dong, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and O. Ronneberger, “Faster R-CNN: Towards real-time object detection with region proposals,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 5261–5269.
Ronan Fablet (Senior Member, IEEE) received the bachelor's degree from the Ecole Nationale Supérieure de ronanique et de l’Espace, Toulouse, France, in 1997, and the Ph.D. degree in signal processing and telecommunications from the University of Rennes/INRIA Rennes, Rennes, France, in 2001. In 2002, he was an INRIA Post-Doctoral Fellow with Brown University, Providence, RI, USA. From 2003 to 2007, he held a full-time research position with IFREMER, Brest, France, in the field of signal and image processing applied to fisheries science. In 2008, he joined the Signal and Communications Department, IMT Atlantique Bretagne-Pays de la Loire (formerly com Bretagne), Nantes, France, as an Associate Professor, and has been holding a Full Professor position, since 2012. He was a Visiting Researcher with Institut de Recherche pour le Développement (IRD) and Peruvian Sea Research Institute (IMARPE), Santiago de Surco, Peru, in 2011 and a Visiting Professor at IMEDEA (CSIC/UB, Esportes, Spain), in 2016. He has led national and international research initiatives at the crossroads of ocean remote sensing, ocean observation and ocean modeling. His research interests cover physics-informed deep learning for ocean modeling, monitoring and surveillance.

Quentin Febvre received the bachelor’s degree from the Ecole Supérieure d’Electricité, Paris, France, and the M.Sc. degree in computer science from Tsinghua University, Beijing, China, in 2015. He pursued the Ph.D. degree with the Mathematical and Electrical Engineering Department, IMT Atlantique Bretagne-Pays de la Loire, Plouzané, France.

He then worked as a Data Scientist at Internatinoal Advertising Company, in 2016, before joining Theodo Group, Paris, France, where he worked on web development, data engineering and computer vision applications using data driven approaches. His research focuses on the benefits deep learning approaches can bring to ocean observation systems, and more specifically on altimetry data analysis.