Observation-Only Deep Learning for Gappy Satellite-Derived Ocean Color Data Using 4DVarNet

Clément Dorffer, Frédéric Jourdin, Thi Thuy Nga Nguyen, Rodolphe Devillers, David Mouillot, and Ronan Fablet, *Senior Member, IEEE*

***Abstract*—Monitoring optical properties of coastal and open ocean waters is crucial to assessing the health of marine ecosystems. Deep learning offers a promising approach to address these ecosystem dynamics, especially in scenarios where gap-free ground-truth data is lacking, which poses a challenge for designing effective training frameworks. Using an advanced neural variational data assimilation scheme (4DVarNet), we introduce a comprehensive training framework designed to effectively train directly on gappy datasets. Using the Mediterranean Sea as a case study, our experiments not only highlight the high performance of the chosen neural network (NN) in reconstructing gap-free images from gappy datasets but also demonstrate its superior performance over state-of-the-art algorithms such as data interpolating empirical orthogonal function (DInEOF) and end-to-end neural mapping scheme-based CNN or UNet architectures.**

***Index Terms*—Bio-optical parameter estimation, data assimilation, data-driven model, deep learning in satellite imagery, end-to-end deep learning, image gap filling, observing system experiment (OSE), ocean color remote sensing, space–time interpolation.**

I. Introduction

WATER optical properties are key parameters in understanding and monitoring ocean biogeochemistry and its dynamics [1], in particular in the long term [2], including the fate of particles and dissolved material from the continents [3]. These optical properties allow measurement of phytoplankton biomass and bulk concentration of suspended matter in the water [4]. The water transparency, which can be affected by human activities and reveals the amount of light accessible at a given depth available for ocean primary production [5], can also be monitored.

Measuring water optical properties is challenging and requires multiple and repeated observations to get a good representation of the signal, especially within coastal areas where turbidity is highly fluctuating. While some optical parameters are traditionally measured in situ [6], satellite multispectral imaging can provide high-resolution optical observations that can be used over regions of various sizes. For example, the moderate Resolution Imaging Spectroradiometer (MODIS) sensors [7] provide images every 16 days at 300 m–1 km spatial resolutions for 36 spectral bands (wavelengths: 405–14 385 nm), while the Ocean and Land Color Instrument (OLCI) sensors [8] provide images every 27 days at 300-m spatial resolution, for 21 spectral bands (400–1020 nm). Once preprocessed (e.g., atmospheric corrections), different seawater parameters can be derived from these multispectral data [9], [10], [11], [12], [13]. Here, the optical particulate backscattering coefficient at wavelength 443 nm (labeled BBP443) has been selected for its ability to derive information on a wide range of suspended marine particles, which play different roles in ocean ecosystems and biogeochemical cycling [14]. The Copernicus Marine Environment Monitoring Service (CMEMS) makes daily "level-3" (L3) products of multisensor remote sensing observations of BBP443 in particular [15], freely accessible online via its platform [16]. However, these operational sea surface parameter products can have large sampling gaps, typically ranging between 30% and 70% of missing data for a region such as the Mediterranean Sea, due to the impact of cloud cover on satellite-derived measurements [17]. Improved gap-filling accuracy is therefore essential for a wide range of end-user applications, as the assimilation of satellite observations is needed to reduce model bias in ocean models [18]. The uninterrupted data streams can be critical for biogeochemical models that forecast carbon cycling and for ecosystem assessments tracking the health of fisheries. Furthermore, operational services rely on these complete datasets to monitor water quality and support maritime activities in near real time [3]. Specifically, the enhanced resolution provided by our method better enables the detection and tracking of dynamic, fine-scale phenomena such as river plumes, sediment resuspension events, and submesoscale turbidity features [19], bridging the gap between technical reconstruction and crucial ecological applications.

A variety of algorithms have been proposed to deliver gap-free datasets (daily "level-4," or L4, products), using among others low-rank matrix completion methods [20], [21], geostatistical methods like spatiotemporal Kriging [22], optimal

Received 12 March 2025; revised 14 July 2025, 15 September 2025, and 16 October 2025; accepted 17 October 2025. Date of publication 23 October 2025; date of current version 7 November 2025. This work was supported in part by French National Research Agency (ANR) through the Artificial Intelligence (IA)-Biodiv Challenge: Fish-Predict and the Oceanix Project, in part by High-Performance Computing (HPC) and Graphics Processing Unit (GPU) Resources from Grand Équipement National de Calcul Intensif (GENCI)-Institut du Développement et des Ressources en Informatique Scientifique (IDRIS) under Grant 2021-101030, in part by the Regional Council of Brittany and [European Regional Development Fund (FEDER) Fonds Européen de Développement Régional] through the Access of Contrat de Plan État–Région (CPER) AIDA GPU Cluster, and in part by European Union (EU) Horizon Europe Project European Digital Twin Ocean (EDITO) Model Lab under Grant 101093293. *(Corresponding author: Thi Thuy Nga Nguyen.)*

Clément Dorffer, Thi Thuy Nga Nguyen, and Ronan Fablet are with IMT Atlantique, UMR, CNRS, Lab-STICC, INRIA Team Odyssey, 29280 Brest, France (e-mail: nga.nguyen@imt-atlantique.fr).

Frédéric Jourdin is with the Service Hydrographique et Océanographique de la Marine (Shom), 29200 Brest, France.

Rodolphe Devillers is with the Espace-Dev (IRD, Univ Montpellier, Univ Guyane, Univ La Réunion, Univ Nouvelle Calédonie, Univ Perpignan Via Domitia), Station SEAS-OI, Saint-Pierre, 34093 La Réunion, France.

David Mouillot is with MARBEC, CNRS, IFREMER, IRD, University of Montpellier, 34095 Montpellier, France.

Digital Object Identifier 10.1109/TGRS.2025.3624465

interpolation (OI) [23], data interpolating empirical orthogonal function (DInEOF) [24], and eDInEOF [25]. Recently, neural mapping schemes have emerged as an attractive solution to address sampling gaps in ocean remote sensing datasets [26], [27], [28], [29], [30]. They often suggest a potentially significant improvement in the reconstruction of gap-free products, especially regarding higher resolution patterns. Common deep learning strategies for gap filling often involve training on simulated data from observing system simulation experiments (OSSEs) [26], [31]. However, this approach relies on the critical assumption that simulations can fully capture the complex variability of real-world ocean color, which often does not hold. The key novelty of our work is to introduce a framework that trains directly on observation-only data. By learning the underlying physical dynamics from the gappy satellite measurements themselves, our approach is designed to be more robust and applicable where perfect, gap-free reference data are unavailable.

In this article, we explore state-of-the-art neural mapping methods, especially 4DVarNet schemes [32], to deliver gap-free sea surface BBP443 fields from multisensor satellite observations. We used the Mediterranean Sea as a case study with the associated multisensor sea surface CMEMS ocean color product [33]. Our main contributions are as follows.

1) We introduce a patch-based resampling approach to train neural mapping schemes directly from observation-only gappy datasets, i.e., thereby eliminating the need for real or simulated gap-free reference datasets.
2) Our benchmarking experiments demonstrate that our proposed training framework and the chosen neural mapping 4DVarNet significantly enhance the reconstruction of sea surface BBP443 fields.
3) We conducted extensive experiments with various satellite sensors (MODIS, VIIRS, and OLCI) to evaluate their contributions to data quality. Our analysis confirmed that using all available satellite sensors improves the overall performance. The VIIRS-JPSS1 sensor was found to be particularly crucial thanks to its large swath that covers a larger area.

This article is organized as follows. Section II presents the case study, i.e., the parameter, data product, and domain of interest. Section III presents the different interpolation methods that will be tested. Section IV then presents the different data sampling strategies used for learning. Section V reports on a performance analysis of 4DVarNet algorithm considering the different learning/testing setups. Results are compared to state-of-the-art approaches eDInEOF and direct inversion interpolation. This section also proposes to evaluate the interpolation improvement reached using different combinations of satellite observations. Finally, Section VI synthesizes our main findings and discusses future work.

II. Case Study

A. Parameter of Interest: BBP443

The optical particulate backscattering coefficient (labeled BBP) depends on a wide range of particle sizes: from submicrometer to centimeter particles [34]. The diversity of these particles is also large. There are living and nonliving particles. The nonliving component includes the submicrometer colloids, which contribute significantly (around 40%) to BBP [35], and a bulk of generally larger particles known as nonalgal particles (NAPs), which are composed of detritus, minerogenic particles, and aggregates [36]. The living component includes, apart from viruses and bacteria, the phytoplankton, which is a major contributor to BBP [37], and even the zooplankton, which also contributes to BBP [38].

When measuring BBP, either with in situ or remote sensors, other processes affect the signal as well. These can come from wave-induced bubbles (including microbubbles) and turbulence-related optical effects, which are more related to the atmosphere and ocean hydrodynamics. Finally, liquid oil (in the form of oil slicks, droplets, and oil bubbles), which is mainly anthropogenic, can be encountered [34]. All these components have distinct dynamics, obeying either biological systems (from phytoplankton growth to the diel migration of zooplankton, the microbial loop and detrital production), lithogenic systems (with river plumes and sediment resuspension along the coast), along with purely hydrodynamic systems at different scales (for instance, mesoscale and submesoscale features, or waves and currents generating and transporting bubbles) or even anthropogenic pollution. This explains why we need to accurately reconstruct BBP spatial fields and their temporal evolution in order to help scientific users in disentangling the different signal contributions to BBP.

In climate science in particular, the main interest in recovering BBP is because it is closely related to the particulate organic carbon (POC) content [39], [40] and phytoplankton carbon biomass [41]. Also, BBP is one parameter needed in the important assessment of export of organic carbon by the ocean biological carbon pump [42]. However, the observed BBP signal finds itself superimposed with all the other signals mentioned previously. In the Mediterranean Sea, for instance, BBP appears to also depend, in the eastern basin especially, on the amount of Saharan dust deposited on the surface of the ocean [43].

In our present experiment, we are exploiting the observed BBP at the light wavelength of 443 nm, hereafter noted BBP443, coming from the spatial product mentioned in the next section.

B. Product Used

We used the product [15] provided by the CMEMS that consists of a collection of daily, multisensor (MODIS-aqua, VIIRS-JPSS1, VIIRS-SNPP, and OLCI-sentinel3a and OLCI-sentinel3b) satellite images acquired from September 1997 to date. This dataset consists of daily "level-3" (L3) data, for which images have been regridded, preprocessed (atmospheric corrections were applied), and daily merged. It includes backscattering coefficients BBP443, processed from the raw multispectral reflectance data. The resulting fields involve large missing data rates, typically from 10% to 80% depending on the considered space–time location in the case-study region.

C. Data Splitting

The product covers the entire Mediterranean Sea, from 30 °N to 46 °N and from 6 °W to 36.5 °E, with a 1×1 km

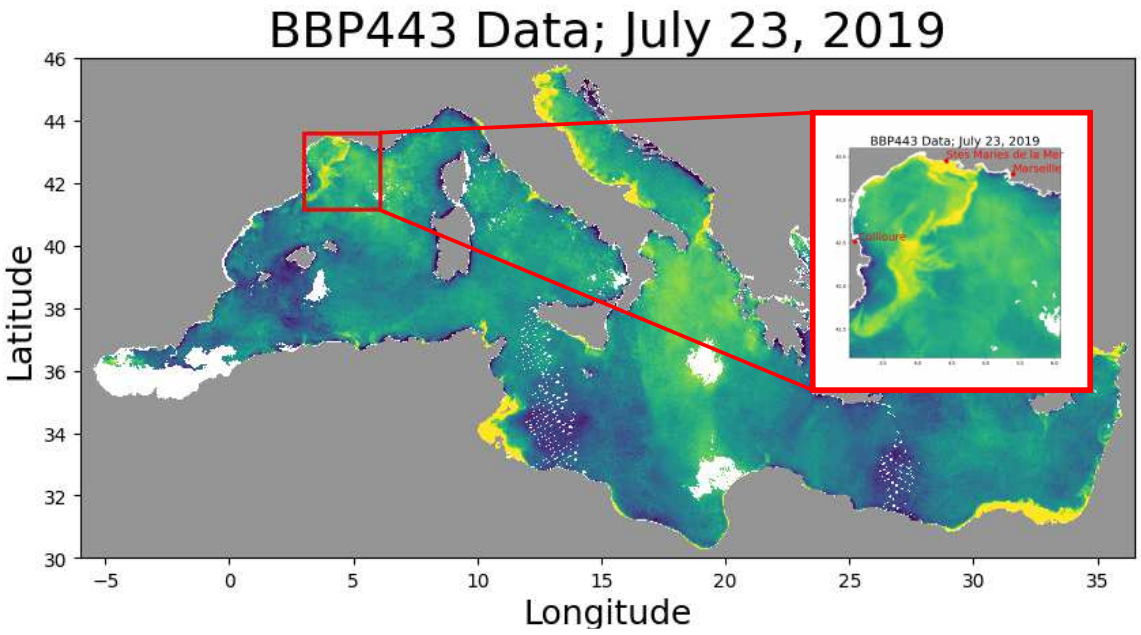

Fig. 1. Selected area of interest used for learning and evaluation tasks. Example of BBP443 data, in common log values, from July 23, 2019. Gray pixels are the continental mask, while white ones correspond to missing values.

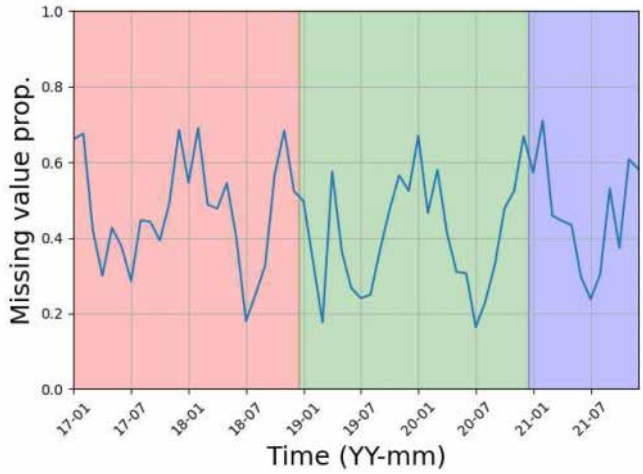

Fig. 2. Evolution of the monthly mean missing value proportion in the 240 × 240 area from January 2017 to December 2021 and data splitting schemes used for learning-based approaches. Red, blue, and green time windows correspond, respectively, to the training, validation, and testing periods.

resolution, leading to a grid size of 1580 × 3308 pixels. We define a 240 × 240 focus region that extends from 41 °N to 43.5 °N and from 3 °E to 6 °E, i.e., along the French coast.

Fig. 1 shows an example of $\log_{10}$ BBP443 concentration for both the entire area covered by the CMEMS product and the reduced area of interest. The image used is from July 23, 2019, for which missing values account for around 10% of the area, a really small level of incompleteness when compared to the mean missing value proportion of 45% over the time period considered (from January 2017 to December 2021). Evolution of the monthly mean missing value proportion over the restricted area is shown in Fig. 2. The seasonality of the cloud coverage is clearly visible, with more clouds during the winter than the summer periods. Fig. 2 also shows the data splitting used in this study to distinguish the train/test/validation periods.

III. Interpolation Methods

We tested three interpolation methods: DInEOF [24], a popular approach for the operational production of L4 satellite images [44], its enhanced version [25] that encompasses a temporal filtering step to force the temporal correlation into the reconstructed field, a UNet-based neural mapping scheme [45], and the NN-based variational data assimilation algorithm (4DVarNet) [32] that has recently shown state-of-the-art interpolation performance for satellite-based sea surface temperature (SST) and sea surface height (SSH) reconstructions [31].

Here, we first briefly introduce the first two approaches, and then, we provide more details about the 4DVarNet method.

A. DInEOF

Considering $x_1, \ldots, x_n$ a collection of incomplete images focused on the same area with m pixels, then one can build $X \in \mathbf{R}^{m \times n}$ the sparse data matrix whose columns correspond to the vectorized images. DInEOF [23] is an iterative matrix completion approach that consists of the following.

1) Computing $\tilde{X}$ a low-rank approximation of X (using truncated singular value decomposition).
2) Filling missing entries in X by corresponding values in $\tilde{X}$.
3) Repeating until convergence.

While relatively simple, this approach provides interesting results, especially when the observed physical phenomenon is slowly changing over time, thus ensuring that most of the signal can be efficiently recovered using only few principal components. However, when using this approach, each image

is recovered independently and does not benefit from the temporal dependencies within the dataset. As a consequence, totally blind images or highly sparse images are usually not correctly interpolated using DInEOF. An enhanced version of this method, eDInEOF, has been proposed in [25], consisting of a filtering step applied after the low-rank approximation to ensure a temporal correlation within matrix $\tilde{X}$ and thus tending to maintain a temporal continuity in the interpolated images.

B. End-to-End Neural Mapping

A classic way to use deep learning to solve a data interpolation problem consists of feeding a neural scheme with sparse images and optimizing the network weights to minimize the reconstruction error, i.e., the difference between the output of the neural network (NN) and the target data to be recovered. Once the training stage is done, the interpolated field associated with incomplete time series is simply provided by the output of the neural scheme fed by the time series. Different neural architectures can be considered. When dealing with image data structures, CNN and UNet schemes [45] are among the state-of-the-art architectures. Hereafter, we refer to those end-to-end neural mapping schemes as direct inversion schemes.

C. 4DVarNet

4DVarNet is an NN-based version of the traditional data assimilation approach 4DVar [46]. It mainly consists in solving the problem

$$\tilde{x} = \arg\min_{x} U(x) = \lambda_1 \|x - y\|_{\Omega}^2 + \lambda_2 \|x - \phi(x)\|_{\mathcal{F}}^2 \quad (1)$$

where x is the state variable to be estimated (e.g., a 3-D field containing a sequence of 2-D spatial fields over time); y represents the incomplete or noisy observations; $\|\ \|_{\Omega}$ stands for the $l2$ (possibly $l1$) norm computed on the observation domain Ω; λ_1 and λ_2 are the variational cost parameters; and ϕ is the dynamical model that could be a differential equation model, a physical model, or an NN-based model.

1) Architecture Details: The 4DVarNet architecture comprises three main trainable components, which are explained below and illustrated in Fig. 3.

1) Dynamical model component ϕ relies on CNN or UNet architectures. It represents the underlying dynamics of the system and contributes to the prior term of the variational cost.
2) The parameters λ_1 and λ_2 balance the observation fidelity term and the dynamical prior term, respectively. These weights are learned during training rather than being manually tuned.
3) Regarding the gradient descent solver, instead of using a fixed gradient step-size learning rate, 4DVarNet employs a long short-term memory (LSTM) network [47] to drive the gradient descent process, including determining step sizes and modifying the gradient for efficient convergence as explained in more detail next.

In 4DVarNet, inspired by meta-learning [48], an iterative gradient solver process is implemented by the LSTM network, which exhibits the capability in capturing the temporal evolution patterns in time-series data

$$\begin{cases} \mathbf{h}(k+1) & = \mathcal{R}\left[\nabla_x U\left(x^k, y\right), \mathbf{h}(k), \mathbf{c}(k)\right] \\ x^{k+1} & = x^k - \mathcal{L}\left(\mathbf{h}(k+1)\right) \end{cases} \quad (2)$$

where $\mathcal{R}$ is an LSTM cell; $\mathbf{h}(k)$ and $\mathbf{c}(k)$ are the hidden state and cell state at iteration k, respectively; and $\mathcal{L}$ is the following mapping:

$$\mathcal{L}\left(\mathbf{h}(k+1)\right) = \frac{1}{k+1}\mathbf{h}(k+1) + \alpha\frac{k+1}{K}\nabla_x U\left(x^k, y\right) \quad (3)$$

where α is a weight for $\nabla_{\mathbf{x}} U$ and K is the total number of iterations (i.e., LSTM cells). Each LSTM cell is equivalent to one iteration of the gradient descent process. Automatic differentiation framework in PyTorch is used to compute the gradients of the variational cost (1) at each iteration given its neural implementation.

Considering all the above three learnable components, the 4DVarNet algorithm consists in an end-to-end architecture that is presented in Fig. 3. Interestingly, if one considers using a NN for model ϕ, then its parameters can be learned simultaneously with variational cost weights and LSTM inner parameters so as to provide the best interpolation. Following the learning scheme presented in Algorithm [26] and Fig. 4, the parameters to be updated with backpropagation are λ_1, λ_2, ϕ, and LSTM, i.e., those appearing in red in Fig. 3.

2) Training Loss: Inspired by the concept of meta-learning [48], in 4DVarNet, the training losses contain two different losses: *reconstruction loss* and *variational cost* (1). This choice relates to a bi-level optimization formulation in which the optimal solution of reconstruction task is also the minimizer of the variational cost.

Reconstruction loss measures the fidelity of the reconstructed field compared to the gappy ground truth (GT) when data are available (supervised learning manner)

$$\alpha\|\tilde{x} - x_{\text{true}}\|^2 + \beta\|\nabla\tilde{x} - \nabla x_{\text{true}}\|^2 + \gamma\|\tilde{x} - \phi(\tilde{x})\|.$$

This loss consists of three components.

1) The first term uses mean square error (mse) to measure pixel-wise fidelity.
2) For applications like ocean remote sensing, we add the gradient norm to ensure consistency in spatial derivatives and preserve fine-scale structures.
3) The prior cost term $\|\tilde{x} - \phi(\tilde{x})\|^2$ enforces physical constraints learned by the autoencoder ϕ. Here, α, β, and γ are coefficients that balance the relative importance of the three terms in the reconstruction loss.

3) Training Configuration: In term of configuration, Table I indicates our chosen configuration.

Fig. 4 shows a pseudocode of the 4DVarNet scheme.

IV. Learning Strategies

NN-based approaches require observations and target data to be fit during the training phase. Depending on the application, those targets can be, e.g., known GT—hardly available for satellite multispectral imaging—or reference model outputs such as in [31]. Here, we aim to train NN-based mapping

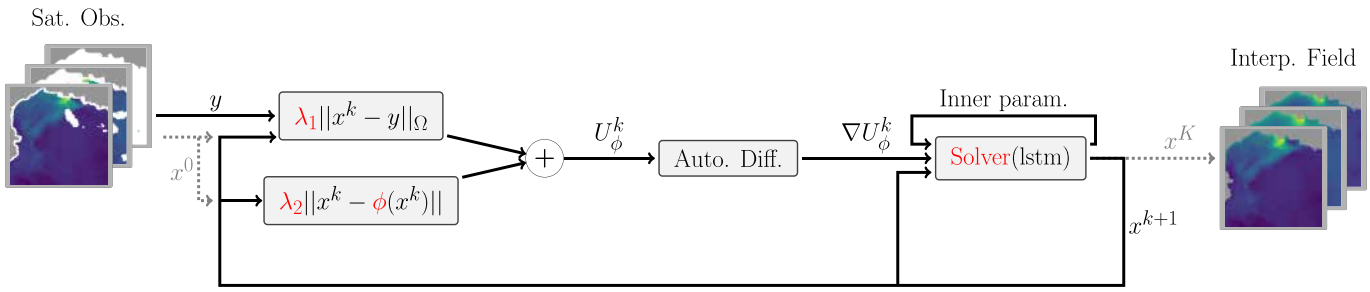

Fig. 3. Diagram of the 4DVarNet algorithm. Trainable components, i.e., the LSTM solver, dynamic model Φ, and variational cost parameters (λ_i), are presented in red.

Algorithm 1 4DVarNet Algorithm

Require: Observation data y, number of iterations K
Ensure: Reconstructed state $\tilde{x}$
  Initialize:
    Trainable components:
      - Dynamical function ϕ (CNN-based or UNet-based)
      - Learnable parameters λ_1 and λ_2
      - Learnable Gradient Solver with LSTM network
procedure 4DVarNetSolver(y)
  $x^0 \leftarrow y$, replaced NaN values in y by zero
  $x \leftarrow x^0$
  for $k = 0$ **to** $K - 1$ **do**
    // Compute variational cost and its gradient
    Observation cost $\leftarrow \lambda_1 \|x - y\|_\Omega^2$
    Prior cost $\leftarrow \lambda_2 \|x - \phi(x)\|^2$
    $U \leftarrow$ Observation cost + Prior cost ▷ Variational cost
    Compute $\nabla_x U$ with Pytorch automatic differentiation
    // Learnable gradient descent step
    $\Delta x \leftarrow \mathcal{L}(\mathcal{R}[\nabla_x U(x, y), \mathbf{h}, \mathbf{c}])$
    $x \leftarrow x - \Delta x$ ▷ Update state
  end for
  return x ▷ Return reconstructed state
end procedure
// Training phase
procedure Train(Training dataset)
  for each batch (y, x_{true}) **do**
    $\tilde{x} \leftarrow$ 4DVarNetSolver(y)
    $Loss \leftarrow \alpha\|\tilde{x} - x_{true}\|^2 + \beta\|\nabla\tilde{x} - \nabla x_{true}\|^2 + \gamma\|\tilde{x} - \phi(\tilde{x})\|$
    ▷ Reconstruction loss
    Update $\{\phi, \lambda_1, \lambda_2$ and LSTM parameters$\}$ via backpropagation
  end for
end procedure

Fig. 4. Pseudocode of 4DVarNet scheme.

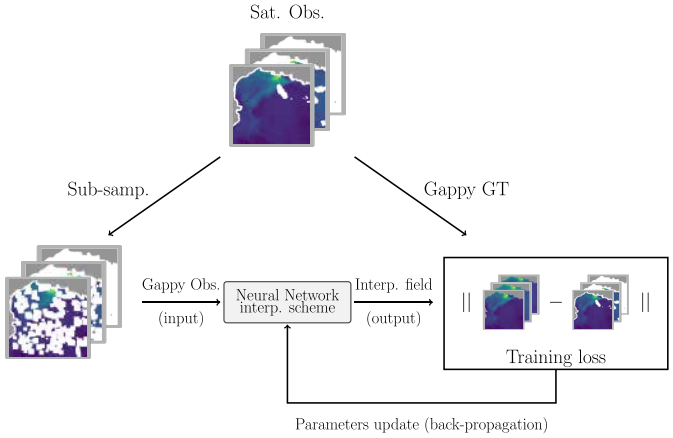

Fig. 5. Proposed learning strategy scheme for NN-based approaches.

algorithms using only available gappy satellite images. We cannot feed the training algorithms with the same data as input and target without any additional assumptions as the NN-based approaches will tend to learn the identity. We propose

TABLE I
Deep Learning Scheme Training Configuration

Parameter	Value
General Settings	
Temporal Resolution (Δt)	7 days
Training epochs	100
Gradient Steps K	15 steps
Data Domain	
Latitude Range	[41.17, 43.6]
Longitude Range	[3.00, 6.09]
Resolution	1 km
Train Period	2017–2018
Validation Period	2021
Test Period	2019–2020
Training Settings	
Batch Size	2
Number of parameters	500 thousands
GPU info	`1 NVIDIA A100 80GB GPU`
Training time	4.5 hours
Inference time (for 2 years test data)	130 seconds

using the gappy satellite images as a foundation to generate additional sparse observation patterns. The training data then consist in considering the gappy satellite images as the target and the subsampled version of these images as the input data. As such, the training loss is evaluated over all available pixels from the satellite images. An overview of this learning strategy is presented in Fig. 5.

One could think that a pure random subsampling of the satellite images—i.e., randomly removing pixels from the images—is an easy way to build a training dataset. While this "salt-and-pepper" noise is simple to implement, it defines a simple interpolation problem that often fails to force the network to learn the underlying, spatially coherent physical dynamics. To create a more challenging and realistic training objective that promotes better generalization, we therefore propose two resampling strategies based on removing spatially structured data: one using real satellite observation patterns and another based on removing random square patches.

A. Real Satellite-Based Observation Patterns

The data product considered [15] is a multisensor product, meaning that the images result from the fusion of optical data acquired by different satellite sensors. A sensor mask informing the different sensors contributing to each pixel is provided for each daily image.

Our first data subsampling strategy exploits these sensor masks to generate gappy patterns associated with different multisensor configuration, as illustrated in Fig. 6. As an

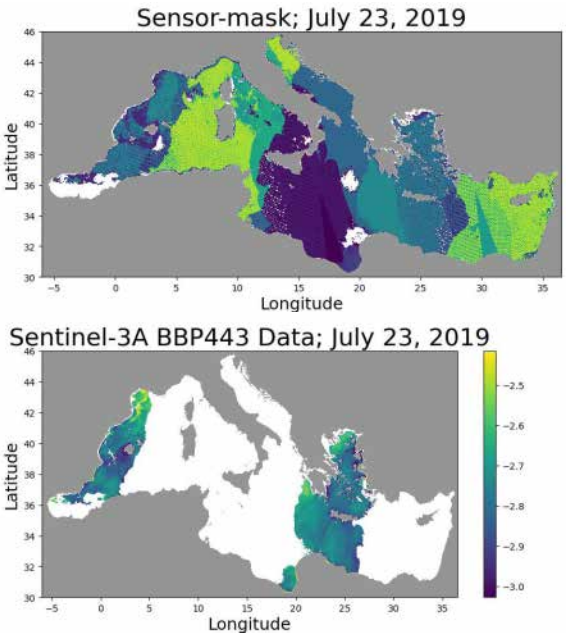

Fig. 6. Example of satellite-based observations sampling. (Top) Available sensor mask. Colors are composed with respect to available sensors presented in Section II. (Bottom) Observations sampled using the Sentinel-3-A mask. The unit of measurement is depicted in $\log_{10}$ scale of m^{-1}.

example, we also display the resulting field when considering only the OLCI-S3A sensor mask is also presented in Fig. 6. Such a sampling strategy provides realistic observation in the way that produced observations correspond to real observation patterns. As a drawback, it makes the proportion of missing data and spatial covering—that is preferred to be homogeneous—difficult to tune. This strategy also provides only one observation mask per day and does not allow randomized simulations. Moreover, some satellites follow trajectories that do not ensure a dense coverage in time and/or space. For example, Sentinel-3A revisits the same site every two days, which means that for a given pixel, an observation is available at most one day out of two, depending on the cloud cover.

B. Randomized Patch-Based Patterns

Another approach that allows us to better control missing data proportion, spatial sampling, and other factors consists in removing random patches instead of removing random individual pixels. The size and shape of the removed patches can also be controlled. This strategy seems appealing as a tradeoff between fully random patterns and realistic ones.

In the latter, we propose generating the observations by removing 50% of data from the multisensor images, with original missing value proportions that are below 75%. In other words, images from the original dataset that have more than 75% of missing data are kept in full as observations, while half the content of the other images is removed. Patch sizes (heights and widths) were randomly generated between 5 and 25 pixels, which in our case represents 0.04%–1% of the image. Overall, this approach allows us to discard enough observation data so that the inputs and targets depict significant differences while preserving very gappy images in the observation subset.

For illustration purposes, Fig. 7 shows one example of the two resampling approaches.

Our code and data are open source and available online at https://github.com/CIA-Oceanix/4dvarnet-core/tree/turbidity-fish-predict

V. Experiments

In our numerical experiments, we benchmark the considered interpolation methods presented using different learning and testing configurations in terms of observation patterns. More specifically, we aim to the following.

1) Compare the interpolation performance achieved by DInEOF, direct inversion, and 4DVarNet using a common dataset.
2) Analyze the impact of training and testing observation patterns on the performance of NN-based models.
3) Assess a potential enhancement of the interpolation performance when using multisensor datasets.

For evaluation purposes, we focus on the restricted area shown in Fig. 1. We compute the interpolated fields for the two-year testing time period that ranges from January 1, 2019 to December 31, 2020 (see Fig. 2) with the two following metrics. The root-mean-squared log error (RMSLE)

$$\text{RMSLE} = \sqrt{\frac{1}{\#\Omega} \cdot \sum_{i \in \Omega} \left(\log_{10}(x(i)) - \log_{10}(\tilde{x}(i))\right)^2} \tag{4}$$

and the mean relative error (MRE)

$$\text{MRE}\,(\%) = \frac{1}{\#\Omega} \cdot \sum_{i \in \Omega} 100 \cdot \left| \frac{x(i) - \tilde{x}(i)}{x(i)} \right|. \tag{5}$$

In both cases, Ω is the considered spatiotemporal domain on which the error is evaluated, i.e., pixels from the Gappy GT that are discarded from the observations. $x(i)$ is the GT values at spatiotemporal location i and $\tilde{x}(i)$ is the associated interpolated value. Note that in all illustrations and results, the unit of measurement is expressed in the $\log_{10}$ scale of m^{-1}, unless explicitly indicated otherwise.

A. Benchmarking of the Interpolation Methods

To compare eDInEOF, NN-based interpolation, and 4DVarNet, we propose to consider the dataset composed of the multisatellite images as the Gappy GT used for the learning step and testing steps, and observations obtained with the random patch-removing strategy for the training and testing steps. Two NNs were considered for both the direct inversion and 4DVarNet, a UNet, and a CNN. In both cases, networks were trained up to 100 epochs and patch-based observations were randomly generated online during the training stage.

To be fair with nonlearning-based approaches, especially with eDInEOF, the same data were used, i.e., the gappy GT from the learning stage and the observations from the testing step. In that way, the calibration of the EOF was performed using both training and testing data.

For each algorithm, the interpolation scores were computed over the randomly removed patches of the testing period and were shown in Table II. One can see that 4DVarNet

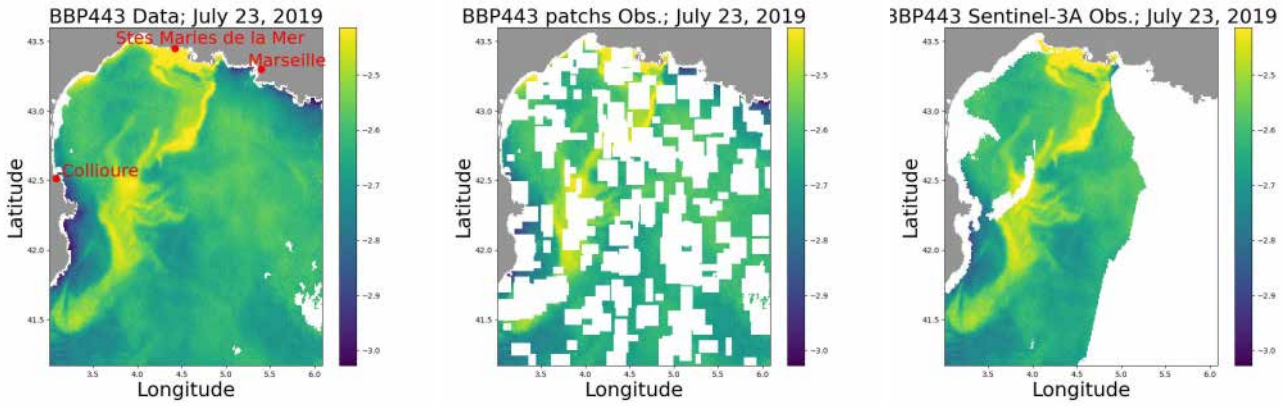

Fig. 7. Example of observation masks to be considered in the training phase. (Left) Gappy GT, (middle) random patch obs., and (right) Sentinel-3A-based obs.

TABLE II
COMPARISON OF THE DIFFERENT INTERPOLATION METHODS USING PATCH-BASED OBSERVATIONS

Algo.	RMSLE	MRE
DInEOF	0.0758	11.7
eDInEOF	0.0745	12.11
CNN	0.128	21.9
UNet	0.119	18.9
4DVarNet (CNN)	0.0501	7.3
4DVarNet (UNet)	0.0520	7.65

provides a large improvement in performance when compared to other approaches tested, with a 34% gain in terms of RMSLE, when compared to eDInEOF interpolation. Fig. 8 shows a reconstruction example for each approach. One can easily see that 4DVarNet approaches better recover small structures and provides more detailed reconstructions. The smooth aspect of DInEOF and eDInEOF typically comes from the underlying low-rank reconstruction strategy, while reconstructions from neural direct inversion approaches seem to suffer from numerous artifacts. While 4DVarNets also rely on convolutional architectures, they do not exhibit similar artifacts. We hypothesize that these artifacts arise because direct inversion models attempt to map the gappy input to a complete output in a single forward pass, which can struggle to enforce physical consistency. In contrast, 4DVarNet's integration of a variational cost may guide the network to learn a dynamical prior, steering the iterative reconstruction toward a more physically consistent state.

Fig. 9 presents a spatial comparison of the reconstruction errors for the six benchmarked interpolation methods over the restricted area of the Mediterranean Sea. These maps reveal critical spatial patterns in algorithm performance, particularly highlighting how direct inversion methods (CNN and UNet) exhibit higher errors in coastal regions and areas with complex dynamics. The DInEOF and eDInEOF methods display intermediate performance with moderate errors throughout the region but also have quite high error near the coast. In contrast, both 4DVarNet implementations show more uniform and generally lower error distributions across the domain.

Fig. 10 illustrates the statistical distribution of reconstruction errors [presented as Log_{10}(RMSLE)] for all six interpolation methods across the two-year test period. We take $\log_{10}$ of RMSLE to get better visualization for the distribution. This figure reveals not just average performance but the full error profile of each algorithm over the two years. The 4DVarNet with CNN method (green) demonstrates a distinct bimodal distribution with its main peak shifted toward lower error values compared to other methods, indicating superior performance across a significant portion of the dataset. Other methods (DInEOF, eDInEOF, and direct inversion approaches either with CNN or UNet) show similar distributions with peaks at higher error values. The 4DVarNet with UNet displays intermediate performance. This distribution analysis complements the spatial maps in Fig. 9 and shows that 4DVarNet schemes (especially 4DVarNet with CNN) perform better than other existing methods both in spatial and temporal interpolations.

As the 4DVarNet scheme using a CNN as model ϕ provides the best performance in these first experiments, we propose to consider only this method for the following tests, and we assess the impact of different observation patterns used during the training and testing stages.

B. Impact of the Learning Setup

Here, we compare the efficiency of 4DVarNets trained with the different observation patterns proposed in Section IV. To that end, we trained the same 4DVarNet architecture, i.e., 4DVarNet with the same CNN-based model, using input observations generated with: 1) the patch-based strategy; 2) a purely random pixel-level strategy where 50% of data were randomly removed; and 3) the S3A-sensor-based strategy. Each trained 4DVarNet was then used for the interpolation of all observation patterns. Interpolation scores were computed on the removed observations, i.e.,: 1) on the removed patches for patch-based observations; 2) on the randomly removed pixels for purely random-based observations; and 3) on the non-S3A-provided data for satellite-based observations.

We report in Table III a summary table of the interpolation scores reached for all models on all observation strategies. As the scores are computed on different data, depending on the testing observation patterns, one cannot directly compare the performances reached for the different testing observation setups. However, one can compare the performances of the

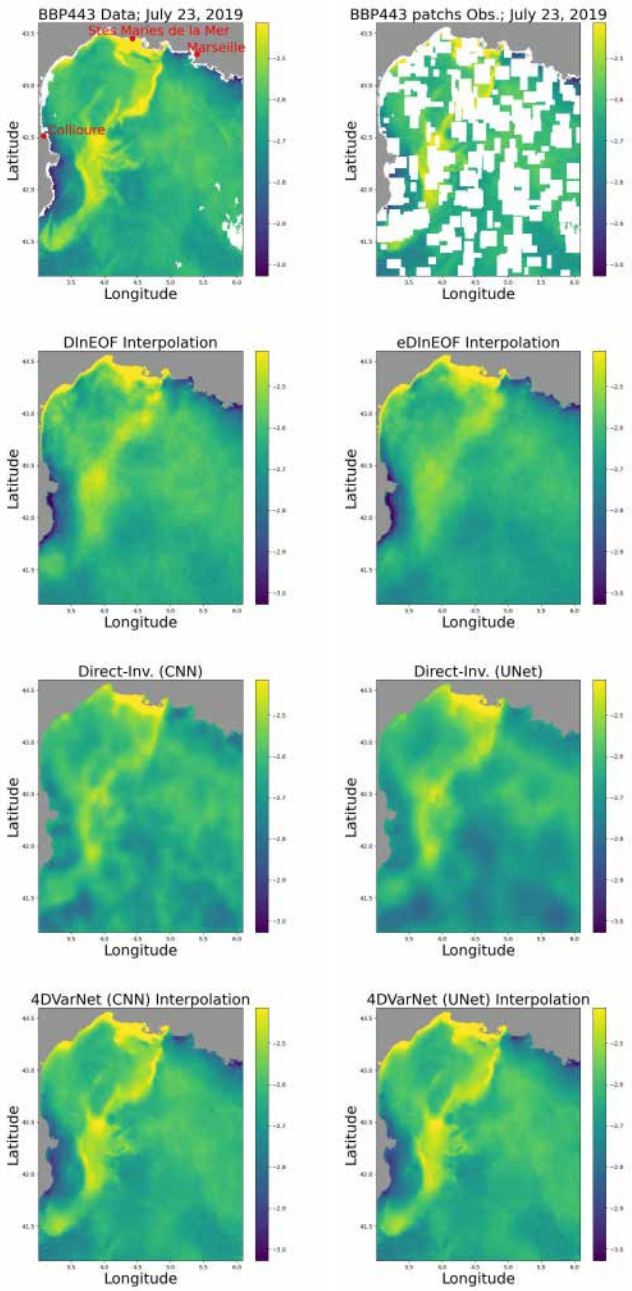

Fig. 8. Interpolation examples using different interpolation methods, highlighting the improved performance of 4DVarNet, which recovers fine-scale, physically plausible structures across the entire domain. (Top left) All available observation (target) and (top right) considered input obs. (Second line, left) DInEOF interpolation. (Second line, right) eDInEOF interpolation. (Third line, left) Direct inversion with CNN. (Third line, right) Direct inversion with UNet. (Bottom left) 4DVarNet interpolation with CNN model. (Bottom right) 4DVarNet interpolation with UNet model.

different 4DVarNet models when considering the same testing setup. We point out two interesting results from Table III. First, 4DVarNet trained using a given observation pattern always provides the best performance when applied to the interpolation of the same observation pattern. That result shows that during the training phase, 4DVarNets learn interpolation schemes optimized for the associated training observation pattern. Second, one can see that when it is used for training, the patch-based observation strategy seems to provide an interpolation scheme that generalizes well, considering the

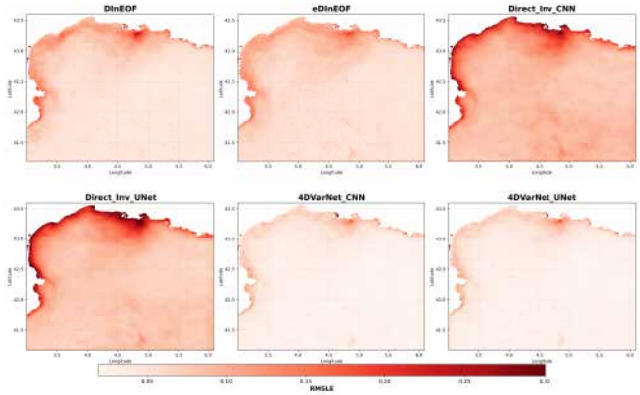

Fig. 9. RMSLE error maps of all interpolation algorithms averaged over the entire test period (2019–2020), comparing performance against gappy satellite observations. The maps highlight consistently lower errors for the 4DVarNet schemes across the domain. (Top row, from left to right) DInEOF, eDInEOF, and direct inversion with CNN. (Bottom row, from left to right) Direct inversion with UNet, 4DVarNet with CNN model, and 4DVarNet with UNet model.

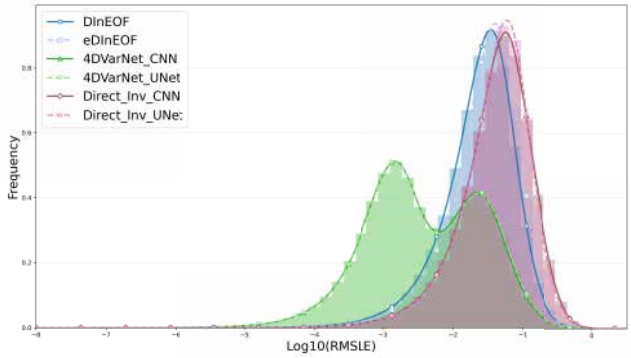

Fig. 10. Log_{10}(RMSLE) distribution over the entire test period (2019–2020) for all six interpolation methods, showing the frequency of different error magnitudes across the test set. This figure highlights the superior performance and reliability of 4DVarNet.

TABLE III

PERFORMANCE METRICS REACHED BY THE 4DVARNET ALGORITHM, CONSIDERING DIFFERENT OBSERVATION SETUP FOR TRAINING AND TESTING STEPS. NOTE: A DIRECT COMPARISON OF ABSOLUTE ERROR VALUES ACROSS COLUMNS SHOULD BE MADE WITH CAUTION, AS EACH TEST PATTERN USES A DIFFERENT EVALUATION MASK

Algo.	Train pattern	Test Pattern	RMSLE	MRE
4DVarNet (CNN)	Patch-based	Patch-based	0.0501	7.3
4DVarNet (CNN)	Random-based		0.1479	20.7
4DVarNet (CNN)	Sensor-based		0.1167	21.5
4DVarNet (CNN)	Patch-based	Random-based	0.0594	8.1
4DVarNet (CNN)	Random-based		0.0271	4.0
4DVarNet (CNN)	Sensor-based		0.1035	18.2
4DVarNet (CNN)	Patch-based	Sensor-based	0.1669	28.6
4DVarNet (CNN)	Random-based		0.2361	42.8
4DVarNet (CNN)	Sensor-based		0.1373	25.8

other two testing observation patterns. The 4DVarNet trained with patch-based observations provides results that are always between the best and worst models, contrary to models learned with purely random or satellite-based observations, which provide poor interpolation scores when used on other setups. We would like to emphasize that a direct comparison of absolute error values across columns of Table III should be made with caution, as each test pattern uses a different evaluation

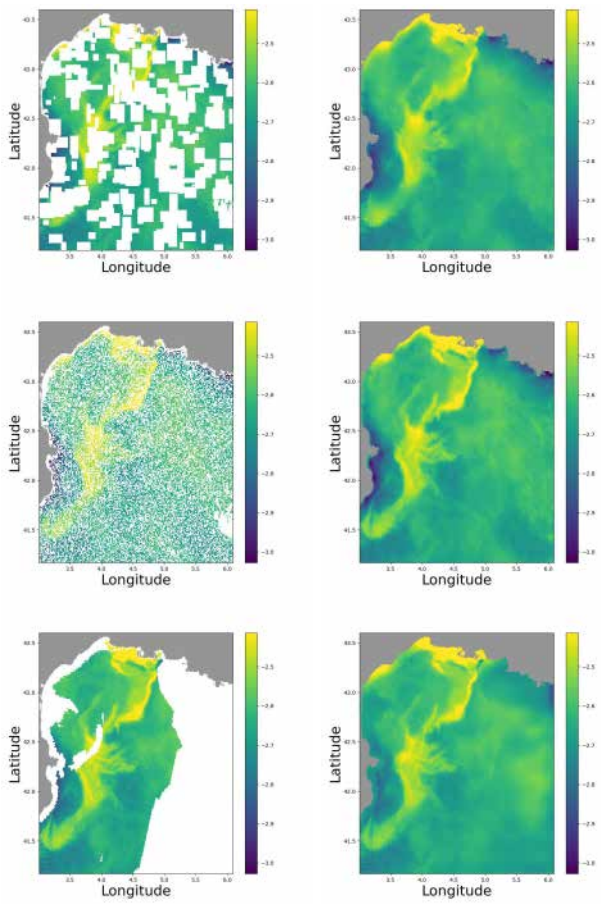

Fig. 11. Examples of reconstructions obtained with 4DVarNet for different observation patterns. (Left, top bottom) Considered input observations with random patch obs., random pixel level obs., and Sentinel-3A-based obs. (Right, top bottom) Associated 4DVarNet reconstructions.

mask. However, given the large volume of data in the test set, these metrics still provide a useful indication of the relative performance and difficulty associated with interpolating each type of observation pattern.

Fig. 11 shows the reconstruction examples for all kinds of observation patterns using the 4DVarNet model that provides the best RMSLE score for each setup.

Table III demonstrates the impact of different data removal schemes on 4DVarNet performance. The sensor-based scheme removes data in fixed patterns, limiting training flexibility. In contrast, the random patch-based removal serves as effective data augmentation, preventing overfitting and leading to better performance when tested on matching patterns and better generalization performance when tested on other patterns.

Although not included in the experiments of this work, it is worth mentioning that data availability significantly affects interpolation quality. As shown in [26] for altimetry data, very sparse observations (like using only along-track nadir data) substantially degrade performance. Their study demonstrates that combining complementary sparse observation sources (nadir + SWOT) significantly improves reconstruction quality compared to using single-source sparse observations. Based on their findings and our findings shown in Table III, we suggest that effective removal schemes should preserve sufficient information while incorporating randomness to enhance data diversity, particularly when working with limited datasets.

TABLE IV

4DVarNet Interpolation Performance Considering Different Satellite Combinations. Columns Correspond to (From Left to Right) the Considered Sensor Combination, the RMSLE, the MRE, and the Missing Value Proportion (MV prop)

Sensor combinaison	RMSLE	MRE	MV prop.
All-sensors	0.04426	6.71	0.46
1 : OLCI-S3A	0.1372	21.82	0.84
2 : MODIS-Aqua	0.08516	12.77	0.67
3 : VIIRS-JPSS1	**0.06891**	10.14	0.64
4 : VIIRS-SNPP	0.07434	10.84	0.66
5 : OLCI-S3B	0.1444	21.60	0.84
1+2	0.06900	10.17	0.61
1+3	0.06158	8.99	0.59
1+4	0.06386	9.26	0.60
1+5	0.09895	14.01	0.70
2+3	0.05905	8.62	0.57
2+4	0.05900	8.67	0.57
2+5	0.07301	10.67	0.61
3+4	**0.05154**	7.62	0.54
3+5	0.06039	8.85	0.59
4+5	0.06441	9.31	0.60
1+2+3	0.05330	7.84	0.53
1+2+4	0.05237	7.78	0.53
1+2+5	0.05868	8.54	0.55
1+3+4	**0.04756**	7.14	0.51
1+3+5	0.05367	7.87	0.54
1+4+5	0.05527	8.05	0.55
2+3+4	0.05045	7.46	0.51
2+3+5	0.05369	7.87	0.53
2+4+5	0.05373	7.90	0.54
3+4+5	0.04807	7.16	0.51
1+2+3+4	0.04689	7.03	0.49
1+2+3+5	0.04865	7.22	0.50
1+2+4+5	0.04791	7.17	0.50
1+3+4+5	**0.04471**	6.77	0.48
2+3+4+5	0.04724	7.05	0.49

C. Impact of the Satellite Selection

To assess the improvement due to different sensor contributions that impact the spatiotemporal coverage of the area, we propose a new evaluation framework. First, to compare interpolation performance using different input observations, we propose to discard 10% of data from the Gappy GT, ensuring a common evaluation support for each dataset. We then used the 4DVarNet algorithm trained with the above patch pattern to interpolate from the observation generated by keeping pixels provided by different sensor combinations, i.e., we first interpolate data provided by the different individual sensors, we then moved to data provided by all combinations of two sensors, then by all combinations of three sensors. As the dataset contains data provided by five sensors for the considered testing period, the proposed experiment consisted in interpolating 31 datasets and evaluating the reconstruction performances on the common missing data, i.e., on the 10% removed patches. RMSLE and MRE performances are shown in Table IV.

Considering the single sensor setup, one can see that interpolations based on the VIIRS-JPSS1 sensor provide the best reconstruction performances, with relative gains of approximately 52%, 50%, 19%, and 7% compared to OLCI-S3B, OLCI-S3A, MODIS-aqua, and VIIRS-SNPP, respectively. Interestingly, some sensors seem to provide more complementary information than others for the interpolation. Considering the three-sensor setup, one can note that each combination

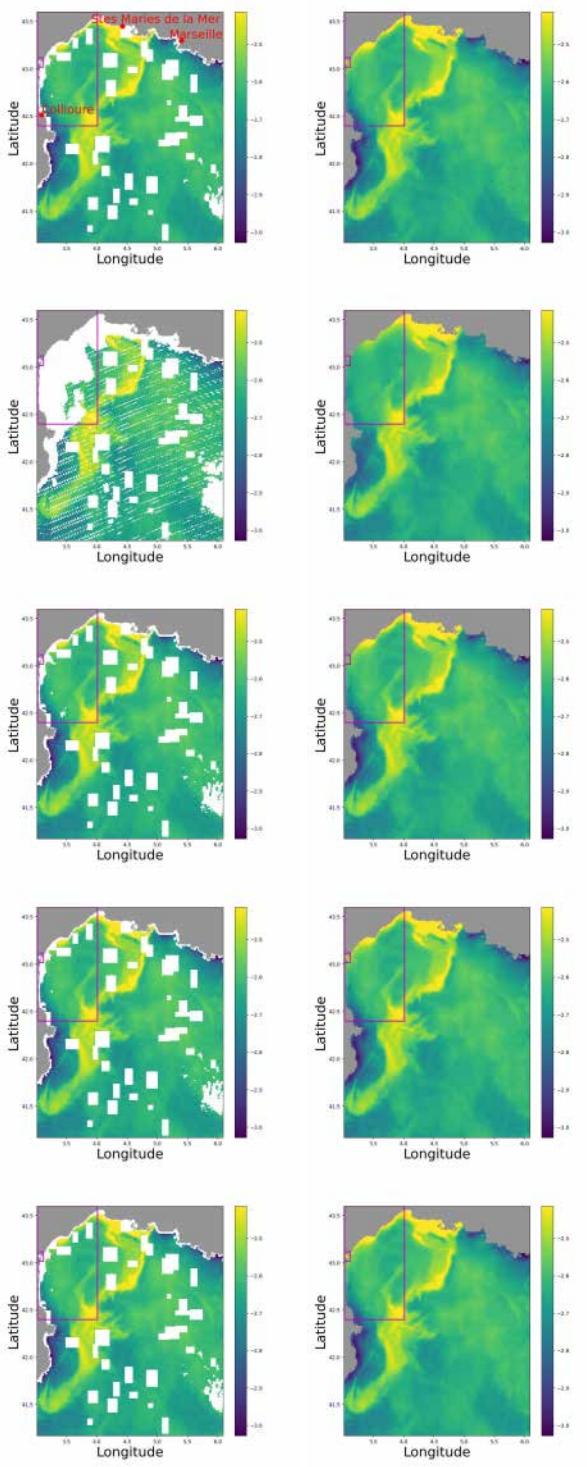

Fig. 12. Interpolation examples using different sensor combinations. (Left) Observations and (right) 4DVarNet reconstruction. (Top) All sensors observation, (second line) VIIRS-JPSS1 Obs., (third line) VIIRS-JPSS1 + VIIRS-SNPP Obs., and (fourth line) VIIRS-JPSS1 + VIIRS-SNPP +S3A Obs. (Bottom) VIIRS-JPSS1 + VIIRS-SNPP + S3A + S3B Obs. The two boxes emphasize areas with noticeable differences.

of sensors provides approximately the same proportion of missing values. However, interpolation performance is quite different, e.g., combinations $1+3+4$ and $2+3+4$ both get 51% of missing data but, respectively, raise RMSLE of 0.047 and 0.050, thus showing that the distribution of missing values largely impacts the interpolation results. These results seem consistent with the satellite properties, as sensor 1 (OLCI-S3A) provides finer observations near the coast compared to sensor 2 (MODIS).

Fig. 12 shows interpolation examples obtained with 4DVarNet algorithm using increasing sensor combinations, i.e., by gradually including sensors into the observations. One can see that the VIIRS-JPSS1 sensor provides a large coverage of the area but is subject to missing values caused by the acquisition process (striping and bow-tie effects [49]). The addition of VIIRS-SNPP observations leads to a significant improvement for the North-Eastern part of the considered region. The addition of S3A and S3B only brings minor changes, as highlighted for the inland waters along the northeastern sea shore. This is in line with the relative improvements brought by the different sensors reported in Table IV.

VI. Conclusion, Discussion, and Future Work

Here, we demonstrated the effectiveness of training NNs 4DVarNet schemes directly on gappy data from satellite observations. Through extensive testing across various sampling scenarios, we find that the best-performing approach uses random rectangular patches. For specific applications with BBP443 reconstructions, our chosen deep NN, 4DVarNet, outperforms other state-of-the-art methods (e.g., DInEOF and NN-based direct inversion), thanks to its integration of data assimilation within the network. Regarding computational cost, as illustrated in Table I, training 4DVarNet schemes on a single NVIDIA A100 80-GB GPU takes around 4.5 h, while the inference on two years of test data takes approximately 2 min, making it particularly efficient during the inference phase.

Regarding the use of multisensor data integration, based on extensive experiments with different satellite sensors, we have made two key observations: 1) utilizing data from all available satellite sensors yields the best performance thanks to the more available data; and 2) VIIRS sensors provide the most information, followed by MODIS, and finally the OLCI sensors (see the upper part of Table IV). However, these large differences in terms of sensor importance are unlikely to be explained by differences in view angles, which are selected to minimize sun glint. Also, they may not come from the sun-synchronous orbits that are comparable between all these satellites. Furthermore, a differential in radiometric performance of each sensor may not be a major cause. Although satellite sensors have different designs, all of them must comply with the binding requirements of operational ocean color satellite missions [50]. For instance, MODIS and VIIRS sensors are affected by two major imagery artifacts, striping and bow-tie distortions, but postprocessing algorithms complete the induced missing values and correct the main noise errors [49]. The viewing geometry (especially the sensor zenith angle) combined with the spectral configuration has an impact on the gap patterns, the atmospheric corrections, and on the uncertainties in deriving geophysical parameters [51]. However, when comparing the error budget of radiometric measurements coming from OLCI, MODIS, and VIIRS sensors, Pardo et al. [52] demonstrate that the primary error source from all these radiometric measurements comes from the atmospheric correction, seen as a common postprocess

applied to all sensors. Finally, the major difference, in terms of observability, between these three sensors used here are their swath widths. VIIRS has the widest ground swath of 3000 km, followed by MODIS with 2330 km, and finally OLCI with 1150 km. We can see that our global gain in RMSLE is nearly proportional to the gain in swath. Obviously, it appears here that the importance of a sensor in our reconstruction simply lies in its swath width. The wider it is, the more amount of data the sensor can bring into the overall process.

As demonstrated in another study [53], the proposed framework also shows strong generalization capabilities across different regions and ocean color parameters without requiring fine-tuning. Although this article focuses on a restricted area of the Mediterranean Sea, it is important to note that, according to [53], the proposed training framework with gappy GT data and random patch-sampling schemes allows the 4DVarNet architecture to generalize well to other regions (such as the North Sea or other areas of the Mediterranean Sea). It also generalizes well to other ocean color parameters (such as suspended particulate matter or chlorophyll-a concentration). Especially, while training over the entire Mediterranean Sea can be computationally demanding due to its large area, the study [53] with the framework proposed in this study (i.e., gappy GT data and random patch-sampling schemes) demonstrates a cost-effective strategy. The model trained on the relatively small selected area transfers well to the entire Mediterranean Sea without *no fine-tuning step*. Similarly, our study supports the potential training of generic interpolation schemes that could apply to a wide range of observation patterns. Future work will aim to scale this framework to deliver global-scale ocean color products. A key step in this process will be the integration of additional, available satellite-derived variables, such as SST, SSH, surface winds, and bathymetry, since they are known to correlate with ocean biogeochemical dynamics. These fields could be incorporated as conditional inputs into the 4DVarNet architecture, enabling the model to learn the physical relationships between ocean dynamics and biogeochemistry.

Though we do not specifically address uncertainty quantification in this study, we acknowledge the retrieval of uncertainty estimates to be of key interest for operational products. Recent studies have explored the extension of the considered 4DVarNet scheme to address jointly mapping problems and uncertainty quantification [26], [54]. For instance, in [54], the 4DVarNet architecture explicitly relies on stochastic partial differential equation prior to deliver an uncertainty quantification through an ensemble approach. A critical next step could be to define a clear operational pathway for its use. The uncertainty maps could be packaged alongside the gap-filled data products and distributed through operational services like the Copernicus Marine Service (CMEMS). These maps would serve as an essential quality flag, enabling end users to make more informed use of the data. For instance, these uncertainty quantification estimates could be used to define the observation error covariance matrix, giving less weight to reconstructions with higher uncertainty. For ecosystem modeling, they could be used to mask out regions where the data are not deemed reliable enough for sensitive analyses.

For future research, beyond the surface of the ocean, combining the learning-based method of [55] with our results could open new research avenues to retrieve the full 3-D $+t$ picture of BBP443 on a global scale and fully exploit the potential of the global network of Bio-Argo floats [56]. However, this presents unique challenges, chief among these are the sparse vertical coverage. Our neural data assimilation framework 4DVarNet might be well-suited to address this. In this case, we can use a 3-D convolutional architecture for the dynamical prior and explicitly providing the space–time coordinates of each observation as inputs. This would pave the way for a 4-D view of ocean biogeochemistry.

ACKNOWLEDGMENT

This study has been conducted using E.U. Copernicus Marine Service Information: https://doi.org/10.48670/ moi-00299

REFERENCES

[1] C. R. McClain, "A decade of satellite ocean color observations," *Annu. Rev. Mar. Sci.*, vol. 1, no. 1, pp. 19–42, Jan. 2009.

[2] S. Dutkiewicz, A. E. Hickman, O. Jahn, S. Henson, C. Beaulieu, and E. Monier, "Ocean colour signature of climate change," *Nature Commun.*, vol. 10, no. 1, p. 578, Feb. 2019.

[3] F. Gohin et al., "Satellite and in situ monitoring of Chl-*a*, turbidity, and total suspended matter in coastal waters: Experience of the year 2017 along the French coasts," *J. Mar. Sci. Eng.*, vol. 8, no. 9, p. 665, Aug. 2020.

[4] H. M. Dierssen and K. Randolph, "Remote sensing of ocean color," in *Earth System Monitoring: Selected Entries From the Encyclopedia of Sustainability Science and Technology*. Cham, Switzerland: Springer, 2012, pp. 439–472.

[5] F. Mohseni et al., "Ocean water quality monitoring using remote sensing techniques: A review," *Mar. Environ. Res.*, vol. 180, Sep. 2022, Art. no. 105701.

[6] A. Bricaud, C. Roesler, and J. R. V. Zaneveld, "In situ methods for measuring the inherent optical properties of ocean waters," *Limnology Oceanogr.*, vol. 40, no. 2, pp. 393–410, Mar. 1995.

[7] W. L. Barnes, X. Xiong, and V. V. Salomonson, "Status of Terra MODIS and Aqua modis," *Adv. Space Res.*, vol. 32, no. 11, pp. 2099–2106, Dec. 2003.

[8] C. Donlon et al., "The global monitoring for environment and security (GMES) Sentinel-3 mission," *Remote Sens. Environ.*, vol. 120, pp. 37–57, May 2012.

[9] M. Hieronymi et al., "Ocean color atmospheric correction methods in view of usability for different optical water types," *Frontiers Mar. Sci.*, vol. 10, Jul. 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmars.2023.1129876

[10] Z. Lee, K. L. Carder, and R. A. Arnone, "Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters," *Appl. Opt.*, vol. 41, no. 27, pp. 5755–5772, 2002.

[11] J.-F. Berthon and G. Zibordi, "Bio-optical relationships for the Adriatic sea," *Int. J. Remote Sens.*, vol. 25, nos. 7–8, pp. 1527–1532, Apr. 2004.

[12] A. Di Cicco, M. Sammartino, S. Marullo, and R. Santoleri, "Regional empirical algorithms for an improved identification of phytoplankton functional types and size classes in the Mediterranean Sea using satellite data," *Frontiers Mar. Sci.*, vol. 4, May 2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmars.2017.00126

[13] G. Volpe et al., "Mediterranean ocean colour level 3 operational multisensor processing," *Ocean Sci.*, vol. 15, no. 1, pp. 127–146, Feb. 2019. [Online]. Available: https://os.copernicus.org/articles/15/127/2019/

[14] E. Boss, D. Stramski, T. Bergmann, S. Pegau, and M. Lewis, "Why should we measure the optical backscattering coefficient?," *Oceanography*, vol. 17, no. 2, pp. 44–49, Jun. 2004.

[15] CMEMS, "Ocean colour mediterranean biogeochemical layers, L3 product for the mediterranean sea," Copernicus Marine Service / Italian Nat. Res. Council, Rome, Italy, 2023. [Online]. Available: https://data.marine.copernicus.eu/product/OCEANCOLOUR_MED_BGC_L3_MY_009_143/description

[16] *Copernicus Marine Service Website*. Accessed: 2023. [Online]. Available: https://data.marine.copernicus.eu/products

[17] E. Ioannidis, C. J. Lolis, C. D. Papadimas, N. Hatzianastassiou, and A. Bartzokas, "On the intra-annual variation of cloudiness over the Mediterranean region," *Atmos. Res.*, vol. 208, pp. 246–256, Aug. 2018.

[18] D. Ford and R. Barciela, "Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products," *Remote Sens. Environ.*, vol. 203, pp. 40–54, Dec. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425717301438

[19] J.-M. Vient, R. Fablet, F. Jourdin, and C. Delacourt, "End-to-end neural interpolation of satellite-derived sea surface suspended sediment concentrations," *Remote Sens.*, vol. 14, no. 16, p. 4024, Aug. 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/16/4024

[20] J.-F. Cai, E. J. Candés, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010, doi: 10.1137/080738970.

[21] L. T. Nguyen, J. Kim, and B. Shim, "Low-rank matrix completion: A contemporary survey," *IEEE Access*, vol. 7, pp. 94215–94237, 2019.

[22] M. A. Oliver and R. Webster, "A tutorial guide to geostatistics: Computing and modelling variograms and kriging," *CATENA*, vol. 113, pp. 56–69, Feb. 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0341816213002385

[23] J. L. Høyer and J. She, "Optimal interpolation of sea surface temperature for the North Sea and Baltic sea," *J. Mar. Syst.*, vol. 65, nos. 1–4, pp. 176–189, Mar. 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924796306002910

[24] J.-M. Beckers, A. Barth, and A. Alvera-Azcárate, "DINEOF reconstruction of clouded images including error maps – application to the sea-surface temperature around Corsican island," *Ocean Sci.*, vol. 2, no. 2, pp. 183–199, Oct. 2006. [Online]. Available: https://os.copernicus.org/articles/2/183/2006/

[25] A. Alvera-Azcárate, A. Barth, D. Sirjacobs, and J.-M. Beckers, "Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF," *Ocean Sci.*, vol. 5, no. 4, pp. 475–485, Oct. 2009.

[26] M. Beauchamp, Q. Febvre, H. Georgenthum, and R. Fablet, "4DVarNet-SSH: End-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry," *Geosci. Model Develop.*, vol. 16, no. 8, pp. 1–37, Apr. 2023.

[27] A. Barth, A. Alvera-Azcárate, C. Troupin, and J.-M. Beckers, "DINCAE 2.0: Multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations," *Geosci. Model Develop.*, vol. 15, no. 5, pp. 2183–2196, Mar. 2022.

[28] A. Barth, J. Brajard, A. Alvera-Azcárate, B. Mohamed, C. Troupin, and J.-M. Beckers, "Ensemble reconstruction of missing satellite data using a denoising diffusion model: Application to chlorophyll a concentration in the black sea," *Ocean Sci.*, vol. 20, no. 6, pp. 1567–1584, Dec. 2024. [Online]. Available: https://os.copernicus.org/articles/20/1567/2024/

[29] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, May 2020, Art. no. 111716. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425720300857

[30] C. Dorffer, T. T. N. Nguyen, R. Fablet, and F. Jourdin, "Adaptive spatial and multi-variable generalization of 4dvarnet in ocean colour remote sensing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2024, pp. 7016–7019.

[31] Q. Febvre, J. Le Sommer, C. Ubelmann, and R. Fablet, "Training neural mapping schemes for satellite altimetry with simulation data," 2023, *arXiv:2309.14350*.

[32] R. Fablet, B. Chapron, L. Drumetz, E. Mémin, O. Pannekoucke, and F. Rousseau, "Learning variational data assimilation models and solvers," *J. Adv. Model. Earth Syst.*, vol. 13, no. 10, p. 2021, Oct. 2021.

[33] CMEMS.(2023). *Ocean Colour Mediterranean Biogeochemical Layers, L3 Product for the Med. Sea*. [Online]. Available: https://data.marine.copernicus.eu/product/OCEANCOLOURMEDBGCL3MY009143/description

[34] D. Stramski, E. Boss, D. Bogucki, and K. J. Voss, "The role of seawater constituents in light backscattering in the ocean," *Prog. Oceanogr.*, vol. 61, no. 1, pp. 27–56, Apr. 2004.

[35] X. Zhang, L. Hu, Y. Xiong, Y. Huot, and D. Gray, "Experimental estimates of optical backscattering associated with submicron particles in clear oceanic waters," *Geophys. Res. Lett.*, vol. 47, no. 4, p. 2020, Feb. 2020.

[36] W. Shi and M. Wang, "Characterization of particle backscattering of global highly turbid waters from VIIRS ocean color observations," *J. Geophys. Res., Oceans*, vol. 122, no. 11, pp. 9255–9275, Nov. 2017.

[37] E. Organelli, G. Dall'Olmo, R. J. W. Brewin, G. A. Tarran, E. Boss, and A. Bricaud, "The open-ocean missing backscattering is in the structural complexity of particles," *Nature Commun.*, vol. 9, no. 1, p. 5439, Dec. 2018.

[38] W. J. Burt and P. D. Tortell, "Observations of zooplankton diel vertical migration from high-resolution surface ocean optical measurements," *Geophys. Res. Lett.*, vol. 45, no. 24, pp. 13–396, Dec. 2018.

[39] K. M. Bisson, E. Boss, P. J. Werdell, A. Ibrahim, and M. J. Behrenfeld, "Particulate backscattering in the global ocean: A comparison of independent assessments," *Geophys. Res. Lett.*, vol. 48, no. 2, p. 2020, Jan. 2021. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL090909

[40] D. Koestner, D. Stramski, and R. A. Reynolds, "Improved multivariable algorithms for estimating oceanic particulate organic carbon concentration from optical backscattering and chlorophyll—A measurements," *Frontiers Mar. Sci.*, vol. 10, Jan. 2024, Art. no. 1197953.

[41] C. Serra-Pompei, A. Hickman, G. L. Britten, and S. Dutkiewicz, "Assessing the potential of backscattering as a proxy for phytoplankton carbon biomass," *Global Biogeochemical Cycles*, vol. 37, no. 6, p. 2022, Jun. 2023.

[42] M. Bellacicco, S. Marullo, G. Dall'Olmo, D. Iudicone, and B. Buongiorno Nardelli, "The oceanic physical injection pump of organic carbon," *Nature Commun.*, vol. 16, no. 1, p. 7100, Aug. 2025.

[43] H. Loisel, E. Bosc, D. Stramski, K. Oubelkheir, and P. Deschamps, "Seasonal variability of the backscattering coefficient in the Mediterranean Sea based on satellite SeaWiFS imagery," *Geophys. Res. Lett.*, vol. 28, no. 22, pp. 4203–4206, 2001.

[44] G. Volpe, B. B. Nardelli, S. Colella, A. Pisano, and R. Santoleri, "An operational interpolated ocean colour product in the Mediterranean Sea," *New Frontiers Oper. Oceanogr.*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:134745461

[45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[46] P. Courtier, J.-N. Thépaut, and A. Hollingsworth, "A strategy for operational implementation of 4D-Var, using an incremental approach," *Quart. J. Roy. Meteorological Soc.*, vol. 120, no. 519, pp. 1367–1387, 1994.

[47] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[48] M. Andrychowicz et al., "Learning to learn by gradient descent by gradient descent," 2016, *arXiv:1606.04474*.

[49] K. Mikelsons, M. Wang, L. Jiang, and M. Bouali, "Destriping algorithm for improved satellite-derived ocean color product imagery," *Opt. Exp.*, vol. 22, no. 23, pp. 28058–28070, 2014.

[50] , C. R. McClain and G. Meister, Ed., *Mission Requirements for Future Ocean-Colour Sensors* (Reports of the International Ocean Colour Coordinating Group), vol. 13. Dartmouth, Canada: IOCCG, 2012, p. 109, doi: 10.25607/OBP-104. [Online]. Available: http://www.ioccg.org/reports/IOCCG_Report13.pdf

[51] B. B. Barnes and C. Hu, "Dependence of satellite ocean color data products on viewing angles: A comparison between SeaWiFS, MODIS, and VIIRS," *Remote Sens. Environ.*, vol. 175, pp. 120–129, Mar. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425715302765

[52] S. Pardo et al., "Radiometric assessment of OLCI, VIIRS, and MODIS using fiducial reference measurements along the Atlantic meridional transect," *Remote Sens. Environ.*, vol. 299, Dec. 2023, Art. no. 113844.

[53] T. Thuy Nga Nguyen, C. Dorffer, F. Jourdin, and R. Fablet, "Generalization performance of neural mapping schemes for the space-time interpolation of satellite-derived ocean colour datasets," 2025, *arXiv:2503.11588*.

[54] M. Beauchamp, Q. Febvre, and R. Fablet, "Ensemble-based 4DVarNet uncertainty quantification for the reconstruction of sea surface height dynamics–ERRATUM," *Environ. Data Sci.*, vol. 2, 2023, Art. no. e33, doi: 10.1017/eds.2023.27.

[55] R. Sauzède et al., "A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth: Retrieval of the particulate backscattering coefficient," *J. Geophys. Res., Oceans*, vol. 121, no. 4, pp. 2552–2571, Apr. 2016.

[56] E. Boss et al., "Observations of pigment and particle distributions in the western North Atlantic from an autonomous float and ocean color satellite," *Limnology Oceanogr.*, vol. 53, no. 5part2, pp. 2112–2122, Sep. 2008.