Environmental Data Science (2026), 5: e43, 1–16
doi:10.1017/eds.2026.10063

METHODS PAPER

Neural ocean forecasting from sparse satellite-derived observations: A case study for SSH dynamics and altimetry data

Daria Botvynko[1], Pierre Haslée[1], Lucile Gaultier[2], Julien Le Sommer[3], Clément de Boyer Montégut[4], Anass el Aouni[5], Betrand Chapron[4] and Ronan Fablet[1]

[1]IMT Atlantique, UMR CNRS 6285 Lab-STICC, Brest, France
[2]OceanDatalab, Brest, France
[3]CNRS, UMR IGE, Grenoble, France
[4]LOPS, CNRS, IRD, Ifremer, Plouzané, France
[5]Mercator Ocean International, Toulouse, France
Corresponding author: Daria Botvynko; Email: daria.botvynko@imt-atlantique.fr

Received: 25 November 2025 UTC; **Revised:** 31 July 2026 UTC; **Accepted:** 05 August 2026 UTC

Keywords: end-to-end neural scheme; neural forecast; satellite altimetry; SLA; end-to-end neural ocean forecast; sea surface neural forecast

Abstract

We present an end-to-end deep learning framework for short-term forecasting of global sea surface dynamics based on sparse satellite altimetry data. Building on two state-of-the-art architectures: U-Net and 4DVarNet, originally developed for image segmentation and spatiotemporal interpolation, respectively, we adapt the models to forecast the sea level anomaly (SLA) and sea surface currents over a seven-day horizon using sequences of sparse nadir altimeters observations. The model is trained on data from the GLORYS12 operational ocean reanalysis, with synthetic nadir sampling patterns applied to simulate realistic observational coverage. The forecasting task is formulated as a sequence-to-sequence mapping, with the input comprising partial SLA snapshots and the target being the corresponding future full-field SLA maps. We evaluate model performance using (i) normalized root mean squared error (nRMSE), (ii) averaged effective resolution, (iii) percentage of correctly predicted velocities, magnitudes and angles, and benchmark results against the operational Mercator Ocean forecast product. Results show that end-to-end neural forecasts outperform the baseline across all lead times, with particularly notable improvements in high variability regions. Our framework is developed within the OceanBench benchmarking initiative, promoting reproducibility and standardized evaluation in ocean machine learning. These results demonstrate the feasibility and potential of end-to-end neural forecasting models for operational oceanography, even in data-sparse conditions.

Impact Statement

This study addresses a key challenge in ocean forecasting: the ability to produce accurate short-term predictions from sparse satellite observations. By adapting a state-of-the-art deep learning model, 4DVarNet, to a forecasting context, we demonstrate that neural networks can effectively predict global sea surface dynamics over a seven-day horizon, even from incomplete data. Our work contributes to the growing field of machine learning for Earth system science and is integrated into the OceanBench initiative, which aims to standardize and accelerate progress in ocean AI research. The approach has practical implications for operational oceanography, offering a scalable and data-efficient alternative to traditional numerical models, particularly valuable in real-time scenarios with limited observational input.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

1. Introduction

The short-term forecasting of the ocean is of key interest for a variety of applications, including among others offshore operations (Le Traon et al., 2019), maritime traffic routing (Davidson et al., 2009), ocean extremes (Hobday et al., 2016) and sampling strategies of scientific surveys. State-of-the-art operational systems rely on data assimilation schemes (Lellouche et al., 2018) to combine a physical model of the ocean dynamics with different observation sources, usually both in situ and satellite-derived observations. The resulting short-term ocean forecasts typically involve an estimation of the current observation given past available observations. The physical model then propagates the state over the targeted time horizon, typically over a week. Numerous studies support the relevance of the operational ocean forecasting schemes, but also reveal significant uncertainty levels (Lellouche et al., 2013).

Recently, neural forecasting schemes have emerged as appealing solutions for ocean dynamics, with striking examples for short-term ocean forecasting (Wang et al., 2024; Cui et al., 2025; El Aouni et al., 2025). These neural schemes involve emulators of the physical model, trained from reanalysis datasets. These schemes now reach state-of-the-art forecasting performance for short-term ocean forecasts and can even outperform data assimilation-based systems for specific case studies and metrics (El Aouni et al., 2025). While those neural forecasting schemes emerge, the characteristics of ocean observing systems may however question whether one should also address uncertainties in the estimation of initial state by data assimilation schemes. Satellite-derived observations, such as the sea surface height (SSH) and sea surface temperature (SST), involve large sampling gaps, which prevent data assimilation schemes to recover fine-scale dynamics, typically below one or 200 kilometers for sea surface currents (SSCs) (Ballarotta et al., 2019). The sampling of the interior of the ocean by ARGO floats or moorings is obviously even scarcer. Optimal interpolation (OI) has long been the reference approach to deliver gap-free observation-based products, but end-to-end neural mapping schemes (Beauchamp et al., 2021; Martin et al., 2023, 2024; Febvre et al., 2024) have shown a great potential to better exploit sparse observation datasets and improve the reconstruction of ocean processes.

This study explores how end-to-end neural schemes could contribute to improved short-term ocean forecasts through the direct exploitation of sparse ocean observation datasets. Following the above-mentioned advances in neural interpolation schemes, we state the short-term forecasting of ocean states as an end-to-end neural mapping problem from gappy observations to targeted ocean forecasts. As demonstration testbed, we focus on SSH forecasts, as SSH is a key signature of global ocean circulation and is associated with sparse satellite altimetry observations (Ducet et al., 2000). Adapting a state-of-the-art neural mapping scheme (Fablet et al., 2021; Febvre et al., 2024), we do not only adapt the 4DVarNet scheme, but also introduce the state-of-the-art UNet architecture for the forecasting task. Our approach integrates both neural architectures to predict SSH from sparse satellite observations. Numerical experiments, conducted with a one-year real satellite altimetry dataset, demonstrate the potential of these deep learning models to significantly enhance short-term ocean forecasts, compared to the state-of-the-art data assimilation-based methods. We discuss further these findings toward the evolution of neural ocean forecasting approaches.

This article is organized as follows: Section 2 provides a brief introduction to short-term ocean forecasting. We present the proposed deep learning approach in Section 3. Section 3.3 describes the considered dataset and benchmark, and we report our results in Section 4. Section 5 discusses our main findings with respect to the state-of-the-art.

2. Problem statement and related work

Operational forecasting systems for the earth systems rely on data assimilation schemes (Lahoz and Schneider, 2014; Buizza et al., 2018). They state the short-term forecasting of a state of interest as a two-step process. The analysis step solves an initial condition problem and the forecasting comes to propagate the initial condition over the targeted forecasting window using the physical dynamical model. Such assimilation-based systems leverage state-of-the-art general circulation models and data assimilation

schemes. The GLO12 ocean forecasting system (Lellouche et al., 2018) falls into this category. It combines the NEMO ocean model (Madec and the NEMO team, 2015) with a Kalman assimilation scheme (Brasseur and Verron, 2006) to deliver 10-day global forecasts at 1/12° resolution, providing both hourly and daily averages based on different observation datasets (mainly satellite altimetry, satellite-derived SST and ARGO float data). While representative of the state-of-the-art in terms for forecasting performance (Lellouche et al., 2013), assimilation-based forecasts may involve significant uncertainties especially for time scales below a week and horizontal scales below a few hundreds of kilometers. This may be critical for a variety of applications such as maritime traffic routing (Davidson et al., 2009; Lellouche et al., 2023) and the forecasting of ocean BGC dynamics (Fennel et al., 2022).

From a methodological point of view, assimilation-based system addresses the forecasting of the entire physical state and does not focus on specific variables of interest. We denote by $\mathbf{x}$ the space-time state of interest, whose evolution is governed by a dynamical operator $\mathcal{M}$ and by $\mathbf{y}$ the observation process, which relates to state $\mathbf{x}$ through an observation operator $\mathcal{H}_t$. This observation operator can be time-dependent to account for irregularly sampled observation data due to the geometry of the observing system (e.g., polar-orbiting earth observation satellites) as well as the impact of atmosphere conditions onto the passive satellite sensing of the ocean surface (e.g., infrared and multispectral satellite sensors). Let t^* denote the current time at which a forecast is provided, distinct from the running time variable t. Forecasting the state $\mathbf{x}$ from time t^* to time $t^* + N \cdot \Delta$ involves two sub-problems:

- the estimation of the state $\widehat{\mathbf{x}}(t^*)$, often stated as a minimization problem:

$$\widehat{\mathbf{x}}_0 = \arg\min_{\mathbf{x}(t^*)} \mathcal{J}\left(\mathbf{x}(t^*), \mathbf{y}\left(t^*_<\right), \mu\right) \tag{1}$$

with $\mathcal{J}$ the cost function, $\mathbf{y}\left(t^*_<\right)$ the observation time series available up to time t^* and μ the set of hyperparameters of the cost function. Variational data assimilation methods (Cummings and Smedstad, 2013) are examples of such schemes.

- the iterative application of a time-stepping operator $\Phi_{\mathcal{M},\Delta}$ for dynamical model $\mathcal{M}$ and time step Δ from initial condition $\widehat{\mathbf{x}}_0$ such that:

$$\mathbf{x}(t^* + n\Delta) = \Phi_{\mathrm{M},\Delta}[\mathbf{x}(t^* + (n-1)\Delta)], \forall n \in [0, N-1] \tag{2}$$

This formulation makes explicit two main sources of uncertainty in assimilation-based ocean forecasts: uncertainty in the estimate $\widehat{\mathbf{x}}_0$ of the initial state from available partial observations and uncertainty associated with the approximation of the true dynamics of state $\mathbf{x}$ by the time-stepping operator $\Phi_{\mathcal{M},\Delta}$. Quantifying these two sources of uncertainty jointly remains challenging, due to the nonlinear growth of initial-condition errors under $\mathcal{M}$ and to the difficulty of characterizing non-Gaussian model and observation errors statistics at the scales relevant to operational forecasting (Lermusiaux et al., 2006; Hoteit et al., 2025).

A very recent literature has emerged regarding neural short-term forecast for earth system dynamics (Wang et al., 2024; Cui et al., 2025; El Aouni et al., 2025), neural schemes for short-term weather forecasts being the most impactful examples (Lam et al., 2022; Bi et al., 2023). Interestingly, different neural architectures, including purely data-driven schemes (El Aouni et al., 2025) and hybrid models using neural correction terms (Liu et al., 2023), have rapidly reached state-of-the-art performance on a global scale compared with model-based forecasts. They typically exploit end-to-end learning strategies from reanalysis datasets and state the short-term forecasting as the prediction of a time series of states over a given time window from a gap-free initial condition. The best neural schemes generalize to real-time forecast configurations to compete or outperform operational systems such as ECMWF IFS schemes (Bi et al., 2023). Similar learning strategies have also been explored for short-term ocean forecasts (Xiong et al., 2023; Wang et al., 2024) and point to the potential improvement of assimilation-based ocean forecasts. Most neural forecasting schemes exploit autoregressive formulations and can be regarded as

replacing the model-based time-stepping operator by a neural time-stepping operator $\Phi_{NN,\Delta}$ (where the *NN* refers to neural-network configuration):

$$\mathbf{x}(t^* + n\Delta) = \Phi_{NN,\Delta}[\mathbf{x}(t^* + (n-1)\Delta)], \forall n \in [0, N-1] \tag{3}$$

As a consequence, such neural forecasting schemes focus on improving forecasting skill associated with uncertainties in the model-based time-stepping operator $\Phi_{\mathcal{M},\Delta}$, while still relying on a previously estimated initial condition $\widehat{\mathbf{x}}_0$.

Regarding ocean dynamics, the scarcity of the available satellite-derived and in situ observation datasets for the ocean may however call into question the underlying assumption that uncertainties in the estimation of the initial state can be neglected. Whereas OI methods and assimilation-based schemes remain the state-of-the-art operational approaches for most operational oceanography products, numerous studies point out how neural mapping schemes, trained from real observation or simulation-only datasets, could better exploit real gappy datasets for a wide range of missing data rates and patterns, as illustrated among others for satellite altimetry (Beauchamp et al., 2021; Martin et al., 2023), ocean color products (Dorffer et al., 2025) and SSTs (Barth et al., 2020). These findings invite to question how they translate to short-term ocean forecasts. Through a case study for satellite altimetry datasets, we explore neural ocean forecasts stated as an end-to-end neural mapping from a series of partial observations to the targeted time series of states in the future (see Figure 1), that is to say trainable neural operators as:

$$\{\mathbf{x}(t^*), \ldots, \mathbf{x}(t^* + N\Delta)\} = \Psi\left[\mathbf{y}\left(t^*_{<}\right)\right] \tag{4}$$

In this study we consider the neural operator Ψ to be based on: (i) the neural data assimilation scheme introduced in (Fablet et al., 2021) or (ii) the state-of-the-art UNet model.

Unlike the auto-regressive neural operator $\Phi_{NN,\Delta}$ of Equation (3), which emulates the physical time-stepping operator from a gap-free initial condition, the operator Ψ introduced here maps the full sequence of past partial observations directly to the future state trajectory in a single end-to-end step, without requiring an explicit initial condition.

In the remainder, we describe those two architectures adapted to the forecast task, as well as the proposed benchmark for the short-term neural forecast of the SSH from satellite altimetry data at the global scale.

3. Proposed approach

3.1. End-to-end neural schemes for short-term forecasting

The forecasting problem is formulated as a nonlinear mapping from observed sparse fields at time t to predicted gap-free fields over the subsequent seven days. Formally:

$$F : \mathcal{X}_t \rightarrow \mathcal{Y}_{t+1:t+7}$$

where $\mathcal{X}_t$ denotes the sparse input and $\mathcal{Y}_{t+1:t+7}$ the target future sequence.

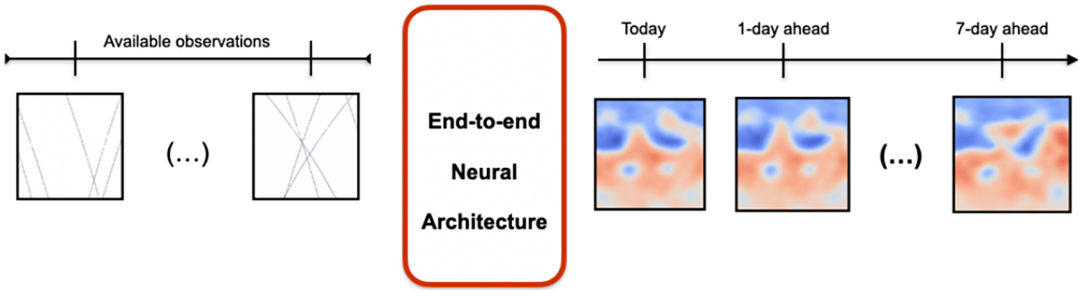


Figure 1. *Proposed end-to-end neural scheme for short-term forecasting: input satellite observations are fed to the end-to-end neural architecture providing a full SSH forecasted field.*

Inspired by recent advances in the neural mapping of gappy observation fields (Fablet et al., 2023; Martin et al., 2023; Febvre et al., 2024), we address the short-term forecasting problem (2.4) as the end-to-end learning of a neural parameterization of operator Ψ. The underlying statement resorts to considering the short-term forecasting problem as an interpolation problem with observation data available only in the past, that is, before time step t^*. Specifically, in this study, we consider a 14-day temporal window in the past.

Considering that the observation data are provided as gridded fields with missing data, any state-of-the-art image-to-image architectures could potentially apply. Among others, we may cite UNet, Transformer and ConvLTSM architectures (Ronneberger et al., 2015; Shi et al., 2015; Vaswani, 2017). As ocean observation data usually involve large missing data rates, neural mapping schemes recently explored for the space–time interpolation of the sea level anomaly (SLA) from satellite altimetry data (Fablet et al., 2023; Martin et al., 2023) also appear as appealing solutions.

As described above, in this study we leverage (i) the 4DVarNet architecture, originally designed for spatiotemporal interpolation, and (ii) UNet, the state-of-the-art neural architecture widely used in computational imaging, and adapt them to perform sequence-to-sequence forecasting from incomplete SLA observations. We refer the reader to Beauchamp et al. (2021) and Fablet et al. (2021) and to the associated Github page for the details on the 4DVarNet framework. The UNet code used in this study is an implementation of Ronneberger et al. (2015) and is available in this GitHub repository.

3.2. OSSE-based training scheme

Following Febvre et al. (2024), we consider a supervised training scheme from simulation data, while we assess the performance of trained model on real ocean observation data. We detail the considered training dataset in Section 3.3. Let us denote by $\left\{\mathbf{x}_n^{true}, \mathbf{y}_n\right\}_n$ the training dataset formed by pairs of true state and associated gappy observation data (see Figure 2), with the latter corresponding to real nadir altimeter sampling patterns. For the 4DVarNet neural scheme, the training loss combines the following three elementary losses for each sample index n (while for the UNet it only combines first two):

$$\mathcal{L}_{\mathbf{x},n} = \sum_{k=-K_O}^{+K_f} \omega_k \left\| \mathbf{x}_n^{true}(t^* + k\Delta) - \widehat{\mathbf{x}}_n(t^* + k\Delta) \right\|^2 \tag{5}$$

$$\mathcal{L}_{\nabla\mathbf{x},n} = \sum_{k=-K_O}^{+K_f} \omega_k \left\| \nabla\mathbf{x}_n^{true}(t^* + k\Delta) - \nabla\widehat{\mathbf{x}}_n(t^* + k\Delta) \right\|^2 \tag{6}$$

$$\mathcal{L}_{\Phi,n} = \left\| \mathbf{x}_n^{true} - \Phi\left(\mathbf{x}_n^{true}\right) \right\|^2 + \left\| \widehat{\mathbf{x}}_n - \Phi(\widehat{\mathbf{x}}_n) \right\|^2 \tag{7}$$

where $K_O = 14$ and $K_f = 7$, respectively, denote the number of past (observed) and future (forecast) daily time steps considered in the loss, consistent with the 14-day observation window and 7-day forecast horizon introduced in Section 3, and ω_k is a weight that increases linearly with k over the observation window ($k = -K_O, \ldots, 0$), reaching its maximum value at the reference time ($k = 0$), and decreases linearly over the forecast window ($k = 0, \ldots, +K_f$). The first two losses are MSE losses for state $\mathbf{x}$ and its spatial gradient $\nabla\mathbf{x}$, whereas the third loss acts as a regularization loss (Fablet et al., 2021). Overall, the training phase amounts to minimizing a weighted sum of these three losses: $\mathcal{L} = \sum_n \alpha_{\mathbf{x}} \mathcal{L}_{\mathbf{x},n} + \alpha_{\nabla\mathbf{x}} \mathcal{L}_{\nabla\mathbf{x},n} + \alpha_\Phi \mathcal{L}_{\Phi,n}$ (for the UNet, only the first two terms are used, i.e., $\alpha_\Phi = 0$). Weights $\alpha_{\mathbf{x}}$, $\alpha_{\nabla\mathbf{x}}$ and α_Φ were tuned empirically to 50, 1,000 and 1, respectively, for the 4DVarNet, and to 50 and 50, respectively, for $\alpha_{\mathbf{x}}$ and $\alpha_{\nabla\mathbf{x}}$ for the UNet.

3.3. Data preparation

Training is conducted using the GLORYS12 operational ocean reanalysis product, which provides daily global SLA fields at a 1/4° spatial resolution. To simulate real-world altimetric coverage, we generate synthetic nadir-like sampling patterns from the GLORYS12 SLA fields (Importantly, this SLA is derived

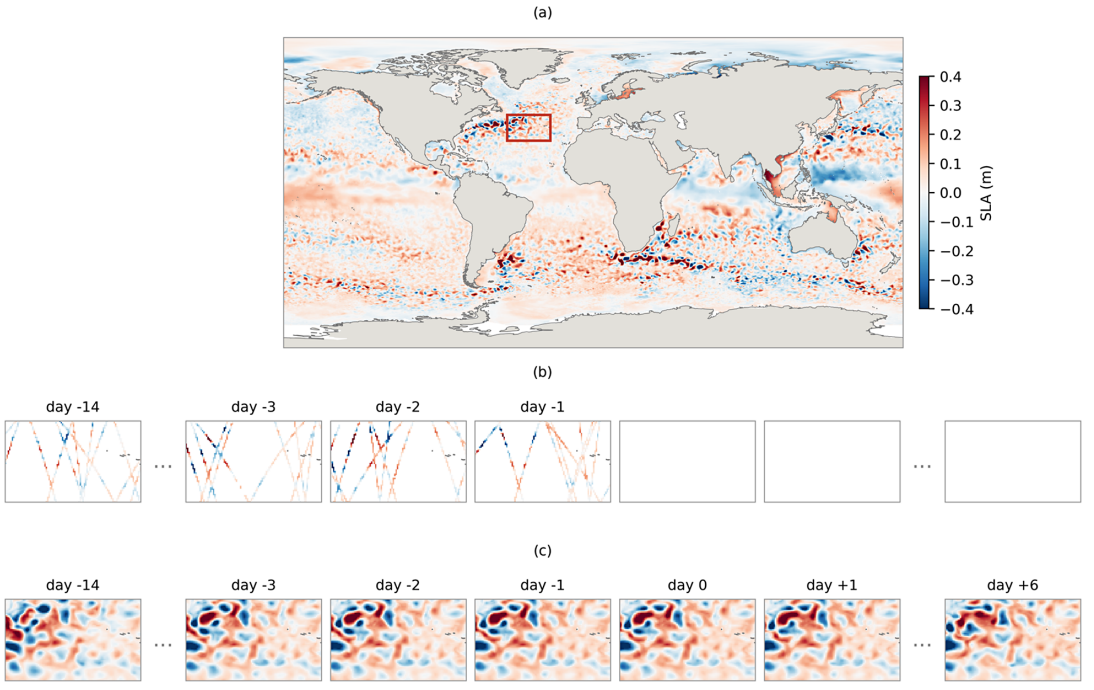


Figure 2. *Input and target of the proposed end-to-end neural scheme for short-term forecasting. (a) Global SLA field at the reference time t^*; the red box marks the region shown in (b, c). (b) Input to the end-to-end neural model: partial along-track SLA observations for a subset of the 14-day past observation window (blank panels indicate future days, for which no observation exists). (c) Target: the corresponding complete SLA fields across the same window, including the 7-day forecast horizon.*

from the related SSH field by substracting the corresponding mean dynamic topography (MDT) doi:10.48670/moi-00021) and couple them to the corresponding gap-free fields (see Figure 2). These sampled tracks serve as input to the model, while the corresponding future full-field GLORYS12 outputs are used as ground truth during training. During the inference phase, we use real satellite altimetry data from near-real-time dataset (doi:10.48670/moi-00147).

The GLORYS12 SLA fields are used as global fields rather than regional patches. Each input tensor has a shape of $1440 \times 680 \times 14$ and each target tensor has a shape of $1440 \times 680 \times 7$, corresponding to the global domain at 1/4° resolution.

3.4. Model training

The models were trained on data from the 2010 to 2018 period, with data from 2019 kept for the validation, using the Adam optimizer and a learning rate of 10^{-3}. Inputs consist of sequences of spatial SLA fields with missing values due to the nadir sampling mask application and outputs are the corresponding complete SLA sequences over a 7-day forecast horizon. Land points are excluded from the loss using a land-sea mask derived from the target field. The training process was performed using one A100 GPU and it took a couple of days for the 4DVarNet to converge and around a day for the UNet. We refer the reader to our PyTorch code for further details.

3.5. Evaluation metrics

The performance of the forecasting system is quantitatively evaluated through a set of diagnostics assessing the accuracy and spatial fidelity of the predicted SLA and SSC fields. All metrics are computed

Table 1. *Benchmarked models and their specifications*

Model	Type	Resolution	Nb of parameters
GLO12	Operational baseline	1/12° daily	—
GloNet	Neural emulator	1/12° daily	200 M
XiHe	Neural emulator	1/12° daily	80 M
4DVarNet	End-to-end neural forecast	1/4° daily	652 K
UNet	End-to-end neural forecast	1/4° daily	17 M

by comparing model forecasts against independent observational datasets: SLA from along-track delayed-time Saral/AltiKa dataset (doi:10.48670/moi-00146) and velocities of drifter trajectories at 15-m depth (doi:10.17882/86236), which were filtered by daily averaging, similarly to Garcia et al. (2025).

We evaluate the forecasted SSH outputs of (i) operational ocean forecast GLO12 provided in real time by Copernicus Marine (doi:10.48670/moi-00016), (ii) state-of-the-art neural ocean emulators (El Aouni et al., 2025; Wang et al., 2024), and (iii) the SLA of the studied end-to-end neural ocean forecasts 4DVarNet and UNet. As summarized in Table 1, these models span two native spatial resolutions: 1/12° for GLO12, GloNet and XiHe, and 1/4° for the proposed end-to-end schemes 4DVarNet and UNet.

Following the evaluation protocol of the Ocean Data Challenge for SSH mapping, we compare all forecasts at the observation locations rather than on a common grid. The forecasted fields are interpolated onto the sparse observation points. Specifically, for the SLA metrics, we interpolate the forecasted data along nadir tracks, while for SSC, we perform the interpolation at the drifters' locations. In contrast to the SSH mapping evaluation introduced in the mentioned Data Challenge, in this study we adapt the framework to respect a strictly forecast-oriented evaluation workflow. We define the leadtime τ as the number of days between the reference time t_0 and the target forecast date, and the nominal forecast leadtime as the calendar date $t_0 + \tau$. We only consider one forecasted field per week for each leadtime (where the first leadtime corresponds to Wednesday). As consequence, in order to ensure temporal consistency of the evaluation, for a given leadtime τ, we consider only observations acquired within a temporal window of ± 0.5 days around the nominal forecast leadtime $t_0 + \tau$ in the computation of the metrics described below, so that the observations used to score a given leadtime are temporally collocated with the target date. It is important to mention that a particular attention during the evaluation process is put into the choice of MDT (Le Traon et al., 2017) to derive SLA from provided SSH outputs and vice-versa. We consider the MDT distributed by Copernicus Marine in order to depict the SSH-derived variables provided by model-based forecasts (GLO12 [Lellouche et al., 2018], GloNet [El Aouni et al., 2025] and [Wang et al., 2024]) and the MDT provided in the corresponding altimetry product (doi:10.48670/moi-00146).

Following the oceanic regime masks of the SSH Mapping OSE Data Challenge (Ballarotta et al., 2019), coastal regions are defined as points located within 200 km of the coastline, while the offshore domain is split into high- and low-variability regions by a threshold of 200 cm^2 (0.02 m^2) on the SLA variance from the DUACS product.

We use the normalized root mean squared error (nRMSE) and the average effective resolution to assess the SLA forecast accuracy. The nRMSE quantifies the relative amplitude of model-observation discrepancies, normalized by the observed variability:

$$\text{nRMSE score} = 1 - \frac{\sum_{i=1}^{N} (\widehat{y}_i - y_i)^2}{\sum_{i=1}^{N} y_i^2}, \tag{8}$$

where $\widehat{y}_i$ and y_i represent the predicted and observed data, respectively, and N the total number of observation samples.

The average effective resolution, or λ_x, is defined as the first wavelength λ, where the noise to signal ratio (NSR) falls below the chosen threshold $\alpha = 0.5$ (Ballarotta et al., 2019; see Equation [9]). Here, the

NSR is computed as the ratio of the power spectral density of the difference between the forecasted and observed signals to the power spectral density of the observed signal, following Ballarotta et al. (2019).

$$\lambda_x = \min\{\lambda | \mathrm{NSR} \leq \alpha\}. \tag{9}$$

λ_x is a diagnostic of the spectral content reconstructed by each product and may be sensitive to the models' grid resolutions, although the resolution alone does not determine it, as models sharing the same grid exhibit different λ_x values (see Figure 4).

In order to obtain the forecasted geostrophic SSC, we differentiate the corresponding SLA forecasts (except the $\pm 2°$ equator area) and apply drifters-based corrections MDT_u and MDT_v from the MDT product doi:10.48670/moi-00150:

$$u = -\frac{g}{f}\frac{\partial \mathrm{SLA}}{\partial y} + MDT_u \tag{10}$$

$$v = \frac{g}{f}\frac{\partial \mathrm{SLA}}{\partial x} + MDT_v \tag{11}$$

where u and v are zonal and meridional components of geostrophic current; parameters g and f represent the gravitational acceleration and the Coriolis parameter.

Following the SSC evaluation described in Garcia et al. (2025), we use here two metrics: percentage of correctly predicted velocity directions (P_θ) and percentage of correctly predicted velocity magnitudes ($P_{|u|}$), both compared to the drifters' velocities at 15-m depth. P_θ quantifies the proportion of locations where the angular deviation between forecasted and observed velocities is within a tolerance angle $\Delta\theta_{\max}$ fixed at $45°$ (see Equation [12]) and $P_{|u|}$ measures the proportion of points where the relative error in speed remains below a threshold $\Delta_{|u|,\max}$ fixed at 25%, (see Equation [13]).

$$P_\theta = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(|\widehat{\theta}_i - \theta_i| < \Delta\theta_{\max}\right) \times 100, \tag{12}$$

$$P_{|u|} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(\frac{||\widehat{u}_i| - |u_i||}{|u_i|} < \Delta_{|u|,\max}\right) \times 100. \tag{13}$$

where N is the total number of collocated observation samples and $\mathbb{I}(\cdot)$ denotes the indicator function.

4. Results

In this section, we present the results of the evaluation of the benchmarked forecast models (see Section 4.1). We focus on the performance of the presented forecasts in comparison to the Mercator Ocean operational forecast. The evaluation is performed globally, followed by an analysis of regime-dependent performances across different dynamical regimes.

4.1. Benchmarked models

In this study, we benchmark state-of-the-art neural ocean emulators (Wang et al., 2024; El Aouni et al., 2025) and end-to-end neural ocean forecast models against the operational baseline GLO12 (Lellouche et al., 2018; Lellouche et al., 2023). The models considered are summarized in Table 1. They are categorized into two groups: neural ocean emulators, which aim to propagate in time the initial state provided by assimilation-based products, and end-to-end neural forecasts, which directly predict the SLA from sparse inputs without relying on any physical model. We evaluate all models at leadtimes ranging from 0 to 7 days, with their predictions compared against independent altimeter and drifter observations. The benchmark focuses on their ability to reproduce SLA and SSC on the set of the evaluation metrics presented in Section 3.5.

4.2. Global-scale performance

Figure 3 shows the SLA predicted by end-to-end neural ocean forecasts at leadtime 0 on January 18, 2023. Figure 4 shows the evaluation metrics across leadtimes 0, 3 and 5 for the benchmarked neural emulators models, all compared to the baseline GLO12, with the colorbar corresponding to the respective relative gains in %. The 4DVarNet consistently outperforms baseline GLO12 and other state-of-the-art neural forecasts across the nRMSE SSH score, remaining stable over leadtimes. Furthermore, 4DVarNet also improves the derived SSC, particularly improving the forecast of SSC angles compared to drifters velocity vectors, suggesting a more accurate capture of flow directions. It is worth recalling that all metrics are computed by interpolating the gridded forecasts onto the nadir track locations. Since this interpolation step cannot add any spectral content beyond what is already present in the forecasted field, the effective resolution metric may be sensitive to each model's native grid resolution, where a finer resolved grid may be penalized more than a coarser one. Interestingly, this is not observed in our benchmark, as the end-to-end schemes 4DVarNet and UNet at coarser resolutions achieve finer average effective resolutions than the finer grids GLO12 and GloNet. This may suggest that the reported gains could reflect a real forecast-quality improvement rather than the grid-resolution-related bias.

4.3. Regime-dependent performance

Tables 2–4 assess the performance of the benchmarked models across leadtimes 0, 3 and 5 and three oceanic regimes: coastal, offshore high-variability and offshore low-variability regions. Here we inter-compare the benchmarked models against the observation-based delayed-time product DUACS

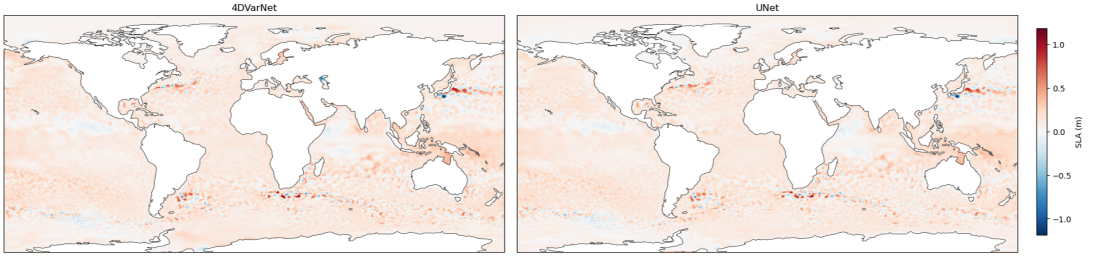


Figure 3. *SLA predicted by proposed end-to-end neural forecasts 4DVarNet (left column) and UNet (right column) at leadtime 0 on 2023-01-18.*

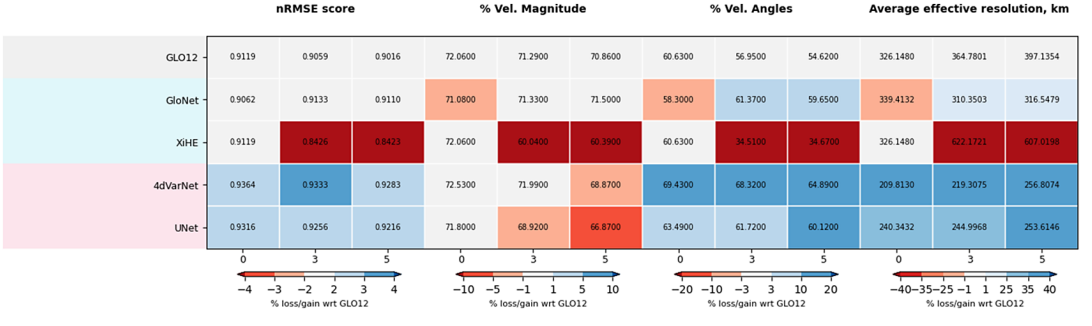


Figure 4. *Benchmark of SSH forecasts (and derived SSC), provided by GLO12, GloNet, XiHe and proposed end-to-end neural forecasts 4DVarNet and UNet. The benchmark presents the following metrics: nRMSE score and average effective resolution of the SSH with respect to SARAL/AltiKa, % correctly predicted SSC angles and magnitudes with respect to drifters. Colors indicate relative gain in % compared to the baseline GLO12, with blue indicating better performances.*

Table 2. *Relative error (%) of RMSE and nRMSE metrics of the forecasted SLA at leadtime 0 with respect to DUACS mapping product*

	Coastal		Offshore high var.		Offshore low var.	
Model	RMSE (%)	nRMSE (%)	RMSE (%)	nRMSE (%)	RMSE (%)	nRMSE (%)
GLO12	44.81	3.13	41.52	5.91	47.22	2.50
GloNet	52.26	3.64	52.54	7.48	57.33	3.04
XiHe	44.81	3.13	41.52	5.91	47.22	2.50
4DVarNet	**8.85**	**0.62**	**−1.35**	**−0.19**	**5.07**	**0.27**
UNet	9.97	0.70	8.95	1.27	15.91	0.84

Note. The metrics are calculated for three dynamical regions: coastal, offshore high-variability and offshore low-variability regions. Positive values indicate degradation relative to DUACS, negative values indicate improvement. Compared to the benchmarked forecasts, DUACS SLA relies on past and future altimetry observations for a given leadtime. Best results are shown in bold.

Table 3. *Relative error (%) of RMSE and nRMSE metrics of the forecasted SLA at leadtime 3 with respect to DUACS at the same leadtime*

	Coastal		Offshore high var.		Offshore low var.	
Model	RMSE (%)	nRMSE (%)	RMSE (%)	nRMSE (%)	RMSE (%)	nRMSE (%)
GLO12	49.23	3.52	54.25	7.72	55.51	2.97
GloNet	39.14	2.80	34.12	4.88	46.49	2.48
XiHe	130.18	9.32	183.34	26.14	157.44	8.41
4DVarNet	**11.20**	**0.80**	**5.37**	**−0.77**	**9.14**	**0.49**
UNet	20.47	1.46	19.49	2.78	22.79	1.21

Note. The metrics are calculated for three dynamical regions: coastal, offshore high-variability and offshore low-variability regions. Positive values indicate degradation relative to DUACS, negative values indicate improvement. Best results are shown in bold.

Table 4. *Relative error (%) of RMSE and nRMSE metrics of the forecasted SLA at leadtime 5 with respect to DUACS at the same leadtime*

	Coastal		Offshore high var.		Offshore low var.	
Model	RMSE (%)	nRMSE (%)	RMSE (%)	nRMSE (%)	RMSE (%)	nRMSE (%)
GLO12	55.21	3.85	64.23	9.39	63.32	3.35
GloNet	48.19	3.37	35.46	5.21	50.37	2.67
XiHe	131.97	9.20	182.31	26.71	160.08	8.64
4DVarNet	**15.48**	**1.08**	**16.52**	**2.42**	**17.72**	**0.94**
UNet	26.33	1.83	29.02	4.27	29.49	1.56

Note. The metrics are calculated for three dynamical regions: coastal, offshore high-variability and offshore low-variability regions. Positive values indicate degradation relative to DUACS, negative values indicate improvement. Best results are shown in bold.

(Taburet et al., 2019). DUACS uses an OI algorithm applied to sparse altimetry data (Ballarotta et al., 2019; Le Traon et al., 2025). We emphasize that, for a given leadtime, DUACS product exploits past and future altimetry observations. This is the reason why we do not include it in the benchmark among the

tested approaches. It provides us with a means to compute relative performance metrics in our regime-dependent analysis. DUACS can better reflect the spatial structures of SLA compared to assimilation-based model analyses, especially in highly dynamic regions (Ballarotta et al., 2019). To quantify forecasts' accuracies relative to the DUACS baseline, we compute percentage gains of each metric relative to the corresponding DUACS score (both evaluated with respect to Saral/AltiKa observations) (see Tables 2–4). At leadtime 0, 4DVarNet achieves the best performance compared to GLO12 baseline across all regions. It remains the closest to the DUACS reconstruction and is even able to outperform it in offshore high-variability regions. UNet performs slightly worse than 4DVarNet, however outperforms GLO12, especially in coastal regions. GloNet underperforms 4DVarNet and UNet, while XiHe is at the same level of performance as GLO12, as its first leadtime is defined as the one provided by the reanalysis. At leadtime 3, 4DVarNet maintains strong performance across all regimes, consistently achieving the lowest RMSE and highest nRMSE, reaching positive gains relative to DUACS. UNet shows competitive performance in low-variability areas, but is worse than GloNet in coastal regions, which gains in performance compared to the first leadtime. By leadtime 5, 4DVarNet still consistently outperforms GLO12 across all regimes and is considerably outperforming DUACS reconstruction especially in coastal and offshore low-variance regions. UNet remains competitive especially in low-variance areas. It is worth mentioning that the nadir data are known to contain additional errors closer to the coast (Ballarotta et al., 2019), which may introduce some bias in the present comparison. While this source of variability is shared by all the products evaluated near the coast, it still may introduce some additional penalization to fields presenting more small-scale variability compared to coarser products. Overall, all models lose performance relative to their first leadtime as forecasts advance, yet 4DVarNet keeps the lowest level of relative errors with respect to DUACS, highlighting the ability of this end-to-end neural architecture to remain close to a purely observations-based product even at longer leadtimes.

Figures 5 and 6 further illustrate this comparison by displaying the SLA fields from GLO12, proposed end-to-end forecasts 4DVarNet and UNet at leadtimes 0 and 5, but also DUACS L4 product for the corresponding dates, in two high-variability regions: Gulf Stream and Agulhas. As expected, GLO12 struggles to correctly position some dynamical structures in comparison to observation-based product DUACS. In the contrast, the end-to-end neural forecasts better recover spatial patterns and are more consistent with DUACS. At leadtime 5, however, the UNet forecast shows a noticeable loss of energy, reflected in a reduced SLA variance, a phenomenon which is not observed in the 4DVarNet forecasts.

5. Conclusion and discussion

This study introduces a deep learning-based framework for short-term ocean forecasting, proposing an end-to-end neural architecture, particularly the 4DVarNet scheme and a state-of-the-art UNet, in a novel forecasting setup. Trained on simulated nadir altimetry data, the model demonstrates state-of-the-art performance in predicting global SSH over a seven-day horizon. Integrated within the OceanBench benchmark setup (Johnson et al., 2023; El Aouni et al., 2026), our approach provides a standardized and reproducible framework for neural ocean forecasting. The results of the present benchmark open promising avenues for future improvements in operational oceanography using machine learning. Our findings demonstrate the relevance of end-to-end neural networks for short-term ocean forecasting. The ability of both 4DVarNet and UNet to handle sparse observations and to produce accurate forecasts is of particular interest for an operational context, where real-time data are often incomplete or noisy.

One of the limitations of the present benchmark concerns the comparison of forecasts provided at different spatial resolutions: 1/4° for the proposed end-to-end schemes (4DVarNet, UNet) against 1/12° (GLO12, GloNet and XiHe) (see Table 1). The presented point-wise evaluation protocol (see Section 3.5) scores every model against the same independent sparse SARAL/AltiKa and drifter observations rather than on a common grid. This avoids the degradation of the 1/12° products to coarser resolution. This approach cannot however fully separate the purely forecasting performance from the intrinsic resolution bias, such as the relative smoothness of the coarser fields. One of the implications may concern the strong performance of the 1/4° end-to-end schemes in offshore low-variability regions (Section 4.3) as a coarser

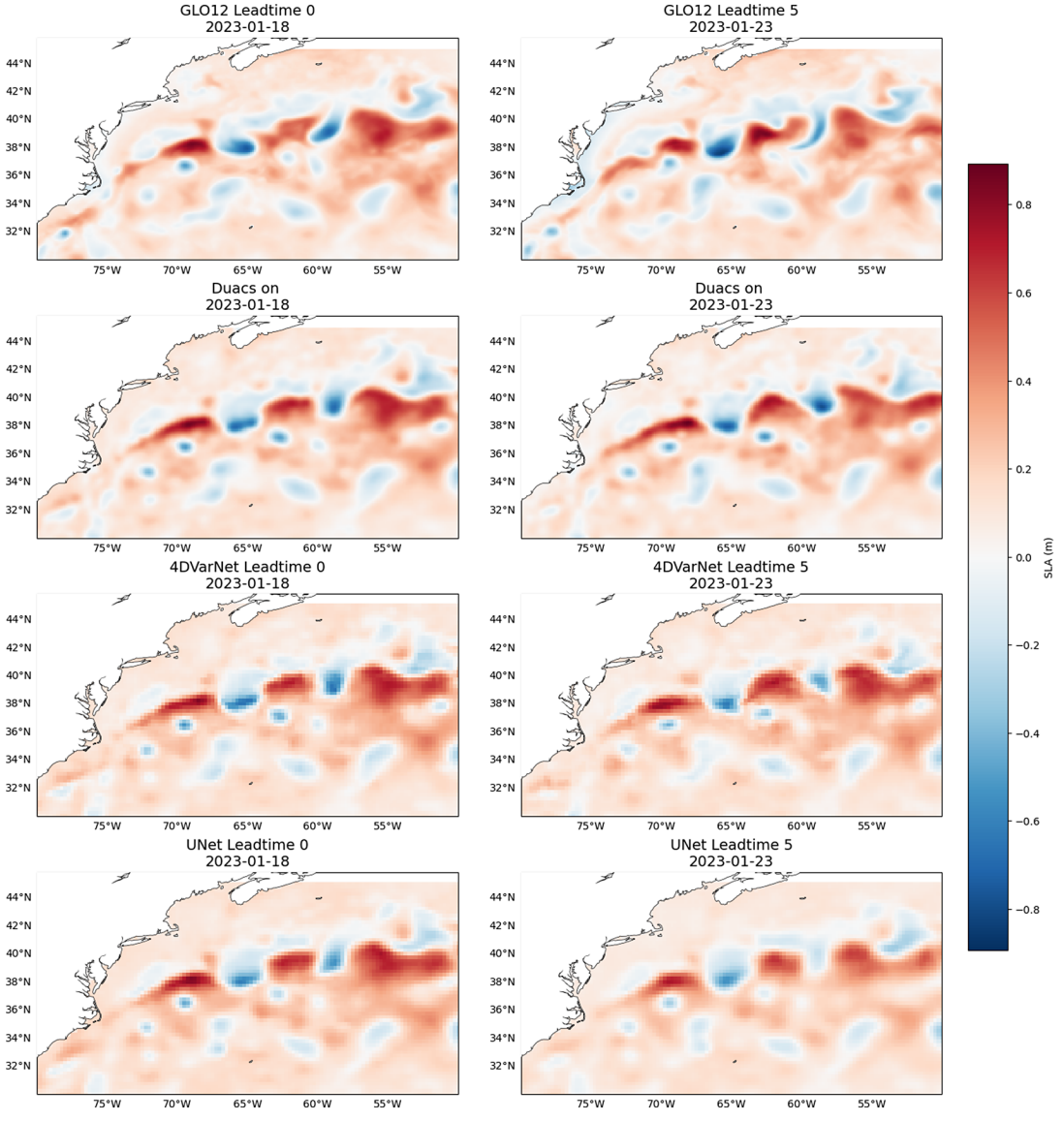


Figure 5. *SLA reconstructed by OI-based DUACS product vs baseline forecast GLO12 and proposed end-to-end neural forecasts 4DVarNet and UNet at leadtimes 0 and 5 in the Gulf Stream on 18 January, 2023 and 23 January, 2023.*

grid tends to smooth the dynamics and thus introduce less fine-scale variability. Furthermore, higher resolved products may be penalized twice, especially in regions where smaller structures may be correctly reconstructed in amplitude but be mispositioned (Crocker et al., 2020). In our benchmark, we observed that in the high-variability regions, 4DVarNet, despite its coarser resolution, outperforms both GLO12 and DUACS (see Section 4.3), in contrast to what we would expect from spatial averaging which would tend to degrade rather than improve performance in dynamically active regimes. Thus, we state this as evidence that at least part of the reported gains reflect the forecasting accuracy. Assessing the impact of resolution on the present benchmark would require a dedicated comparison, such as re-gridding all products to a common resolution before computing the presented metrics or after retraining the end-to-end schemes at higher resolution.

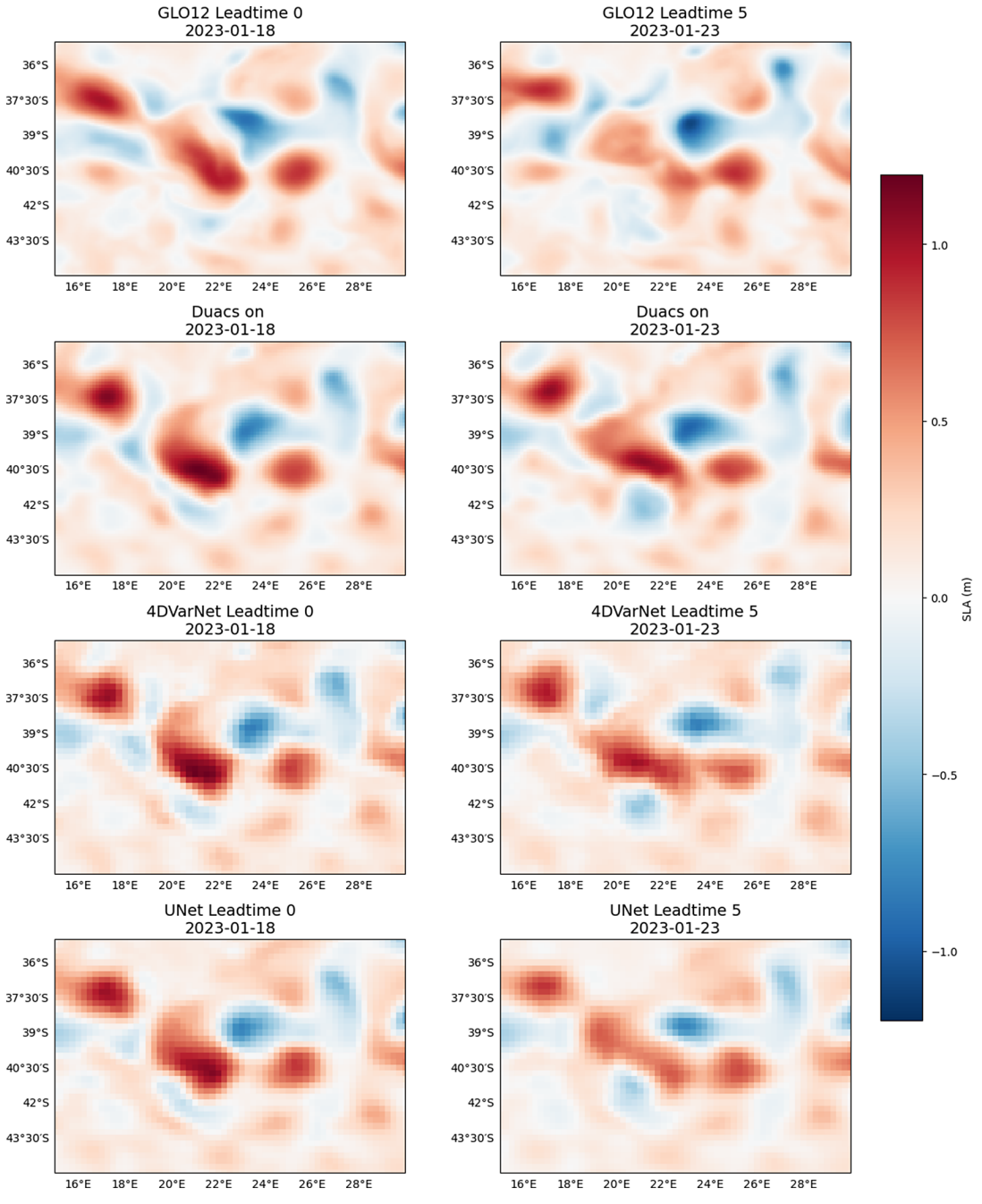


Figure 6. *SLA reconstructed by OI-based DUACS product vs baseline forecast GLO12 and proposed end-to-end neural forecasts 4DVarNet and UNet at leadtimes 0 and 5 in the Agulhas region on January 18, 2023 and 23 January, 2023.*

In addition, the present study benchmarks two state-of-the-art image-to-image architectures, 4DVarNet and UNet, chosen, respectively, for their variational data assimilation grounding and their common use in computational imaging. Beyond these two architectures, the transformer-based one constitutes a promising direction for the end-to-end forecasting task. Recent studies suggest that such models can

outperform convolutional architectures such as UNet on forecasting tasks (Gao et al., 2022). We note, however, that within the present benchmark, the transformer-based neural emulator XiHe underperforms the more conventional CNN-based GloNet, suggesting that architecture choice alone does not guarantee improved performance in the proposed setting, but the training strategy and dataset design play an important role.

While we only rely on training strategies based on reanalysis datasets and simulated observation data, future work could explore the potential of using real observation datasets (Martin et al., 2024; Dorffer et al., 2025). In this context, the benefit of wide-swath altimetry data could be highly valuable (Ballarotta et al., 2025; Fouchet et al., 2025). Accounting for uncertainty quantification in the proposed end-to-end neural forecasts is also a key challenge. We aim to explore the combination of 4DVarNet schemes and generative approaches (Li et al., 2024) as recently explored for SLA mapping (Beauchamp et al., 2025).

Data availability statement. The data used in this study are publicly available at Copernicus Marine Data Store. The code of the proposed benchmark workflow is available at https://github.com/CIA-Oceanix/4dvarnet-starter/tree/forecast-glorys12.

Author contribution. Conceptualization: D.B., L.G., J.L.S., C.d.B.M., A.e.A. and R.F.; Data curation: D.B.; Data visualization: D.B. and L.G.; Methodology: D.B., P.H., J.L.S. and R.F.; Writing original draft: D.B., L.G. and R.F.

Funding statement. This work has been carried out as part of the Copernicus Marine Service OceanBench-STOF project. Copernicus Marine Service is implemented by Mercator Ocean in the framework of a delegation agreement with the European Union. This work was partially supported by the French ANR OceaniX (ANR-19-CHIA-0016). It also benefited from HPC and GPU resources provided by GENCI-IDRIS (Grant 2021-101,030) and the CPER AIDA GPU cluster, supported by the Regional Council of Brittany, Brest Métropole and Fonds Européen de DEveloppement Régional (FEDER).

Competing interests. The authors declare no conflicts of interest.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

Ballarotta M, **Ubelmann C**, **Bellemin-Laponnaz V**, **Le Guillou F**, **Meda G**, **Anadon C**, **Laloue A**, **Delepoulle A**, **Faugère Y**, **Pujol MI**, **Fablet R and Dibarboure G** (2025) Integrating wide-swath altimetry data into level-4 multi-mission maps. *Ocean Science 21*(1), 63–80. https://doi.org/10.5194/os-21-63-2025

Ballarotta M, **Ubelmann C**, **Pujol MI**, **Taburet G**, **Fournier F**, **Legeais JF**, **Faugère Y**, **Delepoulle A**, **Chelton D**, **Dibarboure G and Picot N** (2019) On the resolutions of ocean altimetry maps. *Ocean Science 15*(4), 1091–1109. https://doi.org/10.5194/os-15-1091-2019

Barth A, **Alvera-Azcárate A**, **Licer M and Beckers JM** (2020) DINCAE 1.0: A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. *Geoscientific Model Development 13*(3), 1609–1622. https://doi.org/10.5194/gmd-13-1609-2020

Beauchamp M, **Amar MM**, **Febvre Q and Fablet R** (2021) End-to-end learning of variational interpolation schemes for satellite-derived SSH data. In *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Vol. *2021*. New York: IEEE, pp. 7418–7421.

Beauchamp M, **Fablet R**, **Benaichouche S**, **Tandeo P**, **Desassis N and Chapron B** (2025) Neural variational data assimilation with uncertainty quantification using SPDE priors. *Artificial Intelligence for the Earth Systems 4*(3), 240060. https://doi.org/10.1175/AIES-D-24-0060.1

Bi K, **Xie L**, **Zhang H**, **Chen X**, **Gu X and Tian Q** (2023) Accurate medium-range global weather forecasting with 3D neural networks. *Nature 619*(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

Brasseur P and Verron J (2006) The SEEK filter method for data assimilation in oceanography: A synthesis. *Ocean Dynamics 56*(5), 650–661. https://doi.org/10.1007/s10236-006-0080-3

Buizza R, **Balmaseda MA**, **Brown A**, **English S**, **Forbes R**, **Geer A**, **Haiden T**, **Leutbecher M**, **Magnusson L**, **Rodwell M**, **Sleigh M**, **Stockdale T**, **Vitart F and Wedi N** (2018) *The development and evaluation process followed at ECMWF to upgrade the integrated forecasting system (IFS)*. Reading: European Centre for Medium Range Weather Forecasts.

Crocker R, **Maksymczuk J**, **Mittermaier M**, **Tonani M and Pequignet C** (2020) An approach to the verification of high-resolution ocean models using spatial methods. *Ocean Science 16*(4), 831–845. https://doi.org/10.5194/os-16-831-2020

Cui Y, **Wu R**, **Zhang X**, **Zhu Z**, **Liu B**, **Shi J**, **Chen J**, **Liu H**, **Zhou S**, **Su L**, **Jing Z**, **An H and Wu L** (2025) Forecasting the eddying ocean with a deep neural network. *Nature Communications 16*(1), 2268. https://doi.org/10.1038/s41467-025-57389-2

Cummings JA and Smedstad OM (2013) Variational data assimilation for the global ocean. In *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, Vol. *II*. Berlin, Heidelberg: Springer, pp. 303–343. https://doi.org/10.1007/978-3-642-35088-7_13

Davidson FJ, **Allen A**, **Brassington GB**, **Breivik Ø**, **Daniel P**, **Kamachi M**, **Sato S**, **King B**, **Lefevre F and SuttonMand Kaneko H** (2009) Applications of GODAE Ocean current forecasts to search and rescue and ship routing. *Oceanography 22*(3), 176–181. https://doi.org/10.5670/oceanog.2009.76

Dorffer C, **Jourdin F**, **Nguyen TTN**, **Devillers R**, **Mouillot D and Fablet R** (2025) Observation-only deep learning for gappy satellite-derived ocean colour data using 4DVarNet. *IEEE Transactions on Geoscience and Remote Sensing 63*, 1–12. *https://doi.org/*10.1109/TGRS.2025.3624465

Ducet N, **Le Traon PY and Reverdin G** (2000) Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and-2. *Journal of Geophysical Research: Oceans 105*(C8), 19477–19498. https://doi.org/10.1029/2000JC900063

El Aouni A, **Gaudel Q**, **Johnson JE**, **Charly REGNIER**, **Le Sommer J**, **Van Gennip S, and Le Traon P** (2026) Oceanbench: A benchmark for data-driven global ocean forecasting systems. *Advances in Neural Information Processing Systems 38*. https://doi.org/10.52202/085713-0303

El Aouni A, **Gaudel Q**, **Regnier C**, **Van Gennip S**, **Le Galloudec O**, **Drevillon M**, **Drillet Y and Lellouche JM** (2025) GLONET: Mercator's end-to-end neural Global Ocean forecasting system. *Journal of Geophysical Research: Machine Learning and Computation 2*(3). https://doi.org/10.1029/2025JH000686

Fablet R, **Chapron B**, **Drumetz L**, **Mémin E**, **Pannekoucke O and Rousseau F** (2021) Learning variational data assimilation models and solvers. *Journal of Advances in Modeling Earth Systems 13*(10). https://doi.org/10.1029/2021MS002572

Fablet R, **Febvre Q and Chapron B** (2023) Multimodal 4DVarNets for the reconstruction of sea surface dynamics from SST-SSH synergies. *IEEE Transactions on Geoscience and Remote Sensing 61*, 1–14. https://doi.org/10.1109/TGRS.2023.3268006

Febvre Q, **Le Sommer J**, **Ubelmann C and Fablet R** (2024) Training neural mapping schemes for satellite altimetry with simulation data. *Journal of Advances in Modeling Earth Systems 16*(7). https://doi.org/10.1029/2023MS003959

Fennel K, **Mattern JP**, **Doney SC**, **Bopp L**, **Moore AM**, **Wang B and Yu L** (2022) Ocean biogeochemical modelling. *Nature Reviews Methods Primers 2*(1), 76. https://doi.org/10.1038/s43586-022-00154-2

Fouchet E, **Benkiran M**, **Le Traon PY and Remy E** (2025) Comparison of a global high-resolution ocean data assimilation system with SWOT observations. *Frontiers in Marine Science 12*, 1563934. https://doi.org/10.3389/fmars.2025.1563934

Gao Z, **Shi X**, **Wang H**, **Zhu Y**, **Wang Y**, **Li M**, **Yeung DY** (2022) Earthformer: Exploring space-time transformers for earth system forecasting. *NeurIPS 35*, 25390–25403.

Garcia P, **Larroche I**, **Pesnec A**, **Bull H**, **Archambault T**, **Moschos E**, **Stegner A**, **Charantonis A and Béréziat D** (2025) ORCAst: Operational high-resolution current forecasts. *Artificial Intelligence for the Earth Systems 4*(4). https://doi.org/10.1175/AIES-D-25-0002.1

Hobday AJ, **Alexander LV**, **Perkins SE**, **Smale DA**, **Straub SC**, **Oliver EC**, **Benthuysen JA**, **Burrows MT**, **Donat MG**, **Feng M**, **Holbrook NJ**, **Moore PJ**, **Scannell HA**, **Sen Gupta A and Wernberg T** (2016) A hierarchical approach to defining marine heatwaves. *Progress in Oceanography 141*, 227–238. https://doi.org/10.1016/j.pocean.2015.12.014

Hoteit I, **Chassignet E and Bell M** (2025) Improving accuracy and providing uncertainty estimations: Ensemble methodologies for ocean forecasting. *State of the Planet 5-opsr*, 21. https://doi.org/10.5194/sp-5-opsr-21-2025

Johnson JE, **Febvre Q**, **Gorbunova A**, **Metref S**, **Ballarotta M**, **Le Sommer J and Fablet R** (2023) Oceanbench: The sea surface height edition. *Advances in Neural Information Processing Systems 36*, 78275–78295.

Lahoz WA and Schneider P (2014) Data assimilation: Making sense of earth observation. *Frontiers in Environmental Science 2*, 16. https://doi.org/10.3389/fenvs.2014.00016

Lam R, **Sanchez-Gonzalez A**, **Willson M**, **Wirnsberger P**, **Fortunato M**, **Alet F**, **Ravuri S**, **Ewalds T**, **Eaton-Rosen Z**, **Hu W**, **Merose A**, **Hoyer S**, **Holland G**, **Vinyals O**, **Stott J**, **Pritzel A**, **Mohamed S and Battaglia P** (2022) *GraphCast: Learning skillful medium-range global weather forecasting.* Preprint, arXiv:2212.12794.

Le Traon PY, **Dibarboure G**, **Jacobs G**, **Martin M**, **Rémy E and Schiller A** (2017) Use of satellite altimetry for operational oceanography. In Stammer D and Cazenave A (eds), *Satellite Altimetry over Oceans and Land Surfaces*. CRC Press, pp. 581–608. https://doi.org/10.1201/9781315151779-18

Le Traon PY, **Dibarboure G**, **Lellouche JM**, **Pujol MI**, **Benkiran M**, **Drevillon M**, **Drillet Y**, **Faugère Y and Remy E** (2025) Satellite altimetry and operational oceanography: From Jason-1 to SWOT. *Ocean Science 21*(4), 1329–1347. https://doi.org/10.5194/os-21-1329-2025

Le Traon PY, **Reppucci A**, **Alvarez Fanjul E**, **Aouf L**, **Behrens A**, **Belmonte M**, **Bentamy A**, **Bertino L**, **Brando VE**, **Kreiner MB**, **Benkiran M**, **Carval T**, **Ciliberti SA**, **Claustre H**, **Clementi E**, **Coppini G**, **Cossarini G**, **De Alfonso Alonso-Muñoyerro M**, **Delamarche A**, **Dibarboure G**, **Dinessen F**, **Drevillon M**, **Drillet Y**, **Faugere Y**, **Fernández V**, **Fleming A**, **Garcia-Hermosa MI**, **Sotillo MG**, **Garric G**, **Gasparin F**, **Giordan C**, **Gehlen M**, **Gregoire ML**, **Guinehut S**, **Hamon M**, **Harris C**, **Hernandez F**, **Hinkler JB**, **Hoyer J**, **Karvonen J**, **Kay S**, **King R**, **Lavergne T**, **Lemieux-Dudon B**, **Lima L**, **Mao C**, **Martin MJ**, **Masina S**, **Melet A**, **Buongiorno Nardelli B**, **Nolan G**, **Pascual A**, **Pistoia J**, **Palazov A**, **Piolle JF**, **Pujol MI**, **Pequignet AC**, **Peneva E**, **Pérez Gómez B**, **Petit de la Villeon L**, **Pinardi N**, **Pisano A**, **Pouliquen S**, **Reid R**, **Remy E**, **Santoleri R**, **Siddorn J**, **She J**, **Staneva J**, **Stoffelen A**, **Tonani M**, **Vandenbulcke L**, **von Schuckmann K**, **Volpe G**, **Wettre C and Zacharioudaki A** (2019) From observation to information and users: The Copernicus marine service perspective. *Frontiers in Marine Science 6*, 234. https://doi.org/10.3389/fmars.2019.00234

Lellouche JM, **Greiner E**, **Le Galloudec O**, **Regnier C**, **Benkiran M**, **Testut CE**, **Bourdalle-Badie R**, **Drevillon M**, **Garric G and Drillet Y** (2018) The Mercator Ocean global high-resolution monitoring and forecasting system. *New Frontiers in Operational Oceanography 1*, 563–592.

Lellouche JM, **Greiner E**, **Ruggiero G**, **Bourdallé-Badie R**, **Testut CE**, **Le Galloudec O**, **Benkiran M and Garric G** (2023) Evolution of the Copernicus Marine Service global ocean analysis and forecasting high-resolution system: potential benefit for a wide range of users. *Proceeding Eurogoos*, 242–251. Available at http://hdl.handle.net/10793/1883

Lellouche JM, **Le Galloudec O**, **Drévillon M**, **Régnier C**, **Greiner E**, **Garric G**, **Ferry N**, **Desportes C**, **Testut CE**, **Bricaud C**, **Bourdallé-Badie R**, **Tranchant B**, **Benkiran M**, **Drillet Y**, **Daudin A and De Nicola C** (2013) Evaluation of global monitoring and forecasting systems at Mercator Océan. *Ocean Science 9*(1), 57–81. https://doi.org/10.5194/os-9-57-2013

Lermusiaux PF, **Chiu CS**, **Gawarkiewicz GG**, **Abbot P**, **Robinson AR**, **Miller RN**, **Haley PJ**, **Leslie WG**, **Majumdar SJ**, **Pang A and Lekien F** (2006) Quantifying uncertainties in ocean predictions. *Oceanography 19*(1), 90–103. https://doi.org/10.5670/oceanog.2006.93

Li L, **Carver R**, **Lopez-Gomez I**, **Sha F and Anderson J** (2024) Generative emulation of weather forecast ensembles with diffusion models. *Science Advances 10*(13), eadk4489. https://doi.org/10.1126/sciadv.adk4489

Liu G, **Bracco A and Brajard J** (2023) Systematic bias correction in ocean mesoscale forecasting using machine learning. *Journal of Advances in Modeling Earth Systems 15*(11). https://doi.org/10.1029/2022MS003426

Madec G and the NEMO team (2015) NEMO ocean engine (version 3.6 stable). Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace No. 27. ISSN 1288-1619.

Martin SA, **Manucharyan GE and Klein P** (2023) Synthesizing Sea surface temperature and satellite altimetry observations using deep learning improves the accuracy and resolution of gridded sea surface height anomalies. *Journal of Advances in Modeling Earth Systems 15*(5). https://doi.org/10.1029/2022MS003589

Martin SA, **Manucharyan GE and Klein P** (2024) Deep learning improves global satellite observations of ocean eddy dynamics. *Geophysical Research Letters 51*(17). https://doi.org/10.1029/2024GL110059

Ronneberger O, **Fischer P and Brox T** (2015) U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer International Publishing, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Shi X, **Chen Z**, **Wang H**, **Yeung DY**, **Wong WK and Wc W** (2015) Convolutional LSTM network: A machine learning approach for precipitation now casting. *Advances in Neural Information Processing Systems 28*, pp 802–810.

Taburet G, **Sanchez-Roman A**, **Ballarotta M**, **Pujol MI**, **Legeais JF**, **Fournier F**, **Faugere Y and Dibarboure G** (2019) DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Science 15*(5), 1207–1224. https://doi.org/10.5194/os-15-1207-2019

Vaswani A, **Shazeer N**, **Parmar N**, **Uszkoreit J**, **Jones L**, **Gomez AN**, **Kaiser Ł and Polosukhin I** (2017) Attention is all you need. *Advances in Neural Information Processing Systems 30*.

Wang X, **Wang R**, **Hu N**, **Wang P**, **Huo P**, **Wang G**, **Wang H**, **Wang S**, **Zhu J and Xu J** (2024) *Xihe: A data-driven model for global ocean eddy-resolving forecasting.* Preprint, arXiv:2402.02995.

Xiong W, **Xiang Y**, **Wu H**, **Zhou S**, **Sun Y**, **Ma M and Huang X** (2023) *Ai-GOMS: Large AI-driven global ocean modeling system.* Preprint, arXiv:2308.03152

Cite this article: Botvynko D, Haslée P, Gaultier L, Le Sommer J, de Boyer Montégut C, el Aouni A, Chapron B and Fablet R (2026). Neural ocean forecasting from sparse satellite-derived observations: A case study for SSH dynamics and altimetry data. *Environmental Data Science*, 5: e43. doi:10.1017/eds.2026.10063