

ORCAst: Operational High-Resolution Current Forecasts

PIERRE GARCIA,^{a,c} INÈS LARROCHE,^{a,d} AMÉLIE PESNEC,^a HANNAH BULL,^a THÉO ARCHAMBAULT,^a
EVANGELOS MOSCHOS,^a ALEXANDRE STEGNER,^a ANASTASE CHARANTONIS,^b AND DOMINIQUE BÉREZIAT^c

^a *Amphitrite, Paris, France*

^b *Inria, Sorbonne Université, Paris, France*

^c *Sorbonne Université, CNRS, Paris, France*

^d *Univ. Brest, CNRS, Ifremer, LOPS, Brest, France*

(Manuscript received 15 January 2025, in final form 21 July 2025, accepted 5 September 2025)

ABSTRACT: We present ORCAst, a multistage, multiarm network for Operational High-Resolution Current Forecasts over 1 week. Producing real-time nowcasts and forecasts of ocean surface currents is a challenging problem due to indirect or incomplete information from satellite remote sensing data. Entirely trained on real satellite data and in situ measurements from drifters, our model learns to forecast global ocean surface currents using various sources of ground-truth observations in a multistage learning procedure. Our multiarm encoder–decoder model architecture allows us to first predict sea surface height and geostrophic currents from larger quantities of nadir and Surface Water and Ocean Topography (SWOT) altimetry data, before learning to predict ocean surface currents from much more sparse in situ measurements from drifters. Training our model on specific regions improves performance. Our model achieves stronger nowcast and forecast performance in predicting ocean surface currents than various state-of-the-art methods.

SIGNIFICANCE STATEMENT: Our study introduces a novel neural network designed to produce high-resolution, real-time 7-day forecasts of ocean surface currents. Accurate forecasting of ocean currents is important for ship routing, climate studies, and tracking of pollutants. This is a challenging problem, due to sparse in situ observations, indirect or incomplete satellite observations, and complex ocean dynamics. We develop a multistage training procedure to learn ocean currents by progressively refining forecasts with data of increasing quality. Our method only uses real satellite and in situ data, including the latest generation of satellite altimetry from the recently launched SWOT satellite, and does not depend on numerical simulations.

KEYWORDS: Currents; Ocean dynamics; Altimetry; Satellite observations; Operational forecasting; Deep learning

1. Introduction

Ocean currents are an important component of Earth's climate system, influencing weather patterns, marine ecosystems, and global heat distribution. Nevertheless, accurate reconstructions and forecasts of ocean surface current fields is a challenging task. There are two main types of observational data for estimating ocean surface currents: in situ measurements, for example, from drifters, and satellite observations, for example, from satellite altimetry and from multispectral optical satellites. These observational data can be assimilated into numerical models or used in data-driven methods to reconstruct and forecast ocean currents.

In situ measurements, such as those from drifter buoys transported by the currents, allow direct measurements of ocean surface currents (Ifremer 2024; Elipot et al. 2022). However, drifters are sparse and unevenly distributed across the global ocean. Drifters are commonly used for evaluating the performance of models for estimating ocean currents (Kugusheva et al. 2024; Martin et al. 2024; Le Guillou et al.

2023b). In our work, we use drifters both as ground-truth data to fine-tune our model and to evaluate our method. Other sources of in situ data, for example, from instruments onboard ships, are generally not publicly available, but limited examples of ship routes can be used to evaluate the performance of ocean current nowcasts and forecasts (Kugusheva et al. 2024).

Satellite altimetry measures sea surface height (SSH), which is related to surface currents through the geostrophic balance approximation, i.e., an equilibrium between the Coriolis and pressure forces. Far from the equator, geostrophic currents are the main components of the total ocean surface current. Since their first launch in 1978, nadir-pointing altimeters allow low-resolution estimation of SSH through sparse along-track measurements (Tapley et al. 1982). A widespread method of interpolating these sparse measures of SSH is the Data Unification and Altimeter Combination System (DUACS) (Taburet et al. 2019). This method is a global optimal interpolation, a linear interpolation method with optimally chosen coefficients of covariance matrices with respect to expert knowledge and historical data. However, due to the limited number of along-track measurements, the global maps of SSH provided by this optimal interpolation have a coarse resolution from 1/4° to 1/8° in specific regional areas. In February 2022, a new satellite called Surface Water and Ocean Topography (SWOT) was launched, providing high-resolution SSH data, below the kilometer, with a 120-km swath, significantly increasing the potential of satellite

Pierre Garcia, Inès Larroche, Amélie Pesnec, and Hannah Bull contributed equally to this work.

Corresponding author: Pierre Garcia, pierre.garcia@amphitrite.fr

DOI: 10.1175/AIES-D-25-0002.1 e250002

© 2025 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Brought to you by IFREMER/BILIOTHEQUE LA | Unauthenticated | Downloaded 10/28/25 02:46 PM UTC

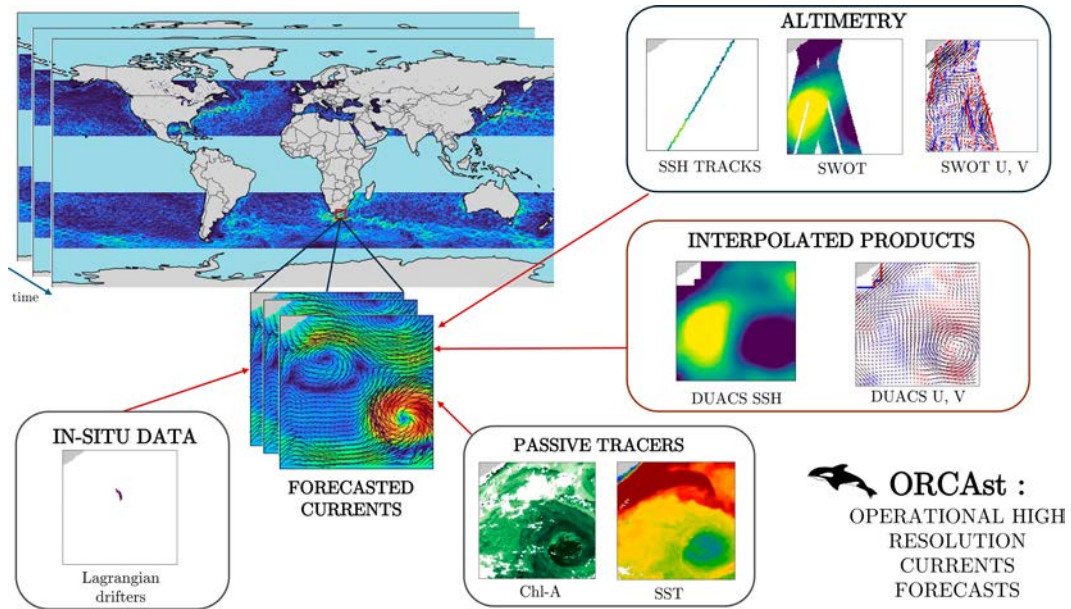


FIG. 1. ORCAst is an operational neural network model that achieves state-of-the-art forecasts of ocean surface currents in extratropical latitudes. It is trained exclusively on observational data, including drifters, SWOT and nadir altimetry, SST, and CHL measurements, in a multistage training process.

altimetry in measuring small-scale structures (Morrow et al. 2018).

In addition to SSH, ocean currents also have a physical signature on other remote sensing variables such as sea surface temperature (SST) and chlorophyll a (CHL). SST and CHL are passive tracers or scalar quantities transported by the currents via advection dynamics. They are measured by multi-spectral optical satellites, which allow for wide data coverage and high-resolution images (300 m–1 km) when no clouds are present (Emery et al. 1989). Numerous studies demonstrate the importance of including multiple physical variables in reconstructing reliable surface currents (Thiria et al. 2023; Fablet et al. 2024; Martin et al. 2023; Archambault et al. 2024b; Kugusheva et al. 2024; Ciani et al. 2024).

There are two main methods of forecasting ocean surface currents using available drifter observations or observations from satellites: numerical models based on physical equations and data-driven approaches. Ocean global circulation models (OGCMs) simulate ocean circulation using differential equations describing fluid dynamics (Tonani et al. 2015). Starting from an estimation of the ocean state, OGCMs can forecast key physical variables. Initial conditions of the model can be estimated by assimilating observational data from satellites or in situ sensors. For example, the Mercator global forecasting system (Drévilion et al. 2008; Mercator Océan International 2024) produces a global 10-day forecast of the ocean, by using the NEMO 3.6 model (Gurvan et al. 2017) and by assimilating satellite observations of SSH, SST, sea ice concentration, and in situ measurements. Physical simulations produce physically realistic fields at the cost of complex hyperparameter tuning and computationally expensive inference, although some methods can be used to reduce the computational

burden of data assimilation (Le Guillou et al. 2023a; Ubelmann et al. 2022).

Data-driven approaches, such as deep learning methods, use large quantities of data to learn the evolution of the physical system, rather than modeling the system dynamics through explicit equations. Deep neural networks have demonstrated high accuracy both in forecasting atmospheric evolution (Lam et al. 2023; Chen et al. 2023) and in forecasting ocean evolution (Wang et al. 2024; Aouni et al. 2024). However, the aforementioned approaches are trained on reanalysis data from numerical models and not purely on observational data. More closely related to our work, various methods produce forecasts or time series of SSH and ocean surface currents using either numerical simulations or observational data for training or a mix of both. Filoche et al. (2022) introduce a loss computed on nadir SSH observations only and use it to perform a variational interpolation with a deep image prior. Archambault et al. (2023) extend this idea to leverage SST information. Martin et al. (2023) present a self-supervised neural network to perform delayed-time (DT) SSH interpolation from partial SSH nadir-pointing and SST satellite observations. Using a related approach, Archambault et al. (2024a) show that multi-stage training, i.e., using different datasets successively in the learning phase, increases SSH reconstruction quality. However, all these methods produce a delayed-time series of SSH fields and must be adapted to forecast ocean surface currents.

In this work, we introduce ORCAst, a data-driven method for Operational High-Resolution Current Forecasts, illustrated in Fig. 1. The novel features of this model are the following:

- ORCAst is an operational neural network model to forecast SSH and ocean surface currents, without being limited

to the geostrophic approximation and achieving state-of-the-art performance.

- ORCAst is trained purely on observational data, including drifters, SWOT, and nadir altimetry, as well as SST and CHL observations.
- ORCAst learns high-resolution ocean surface currents through a three-phase training procedure, which progressively refines forecasts with ground-truth data of increasing quality.

The article is structured as follows: In [section 2](#), we describe the in situ and satellite data used in this study. [Section 3](#) provides a detailed explanation of the architecture of our model ORCAst and our evaluation strategy. In [section 4](#), we present the experimental results demonstrating the operational benefits of ORCAst. Finally, in [section 5](#), we discuss the conclusions and offer perspectives for future work.

2. Data

Ocean surface currents can be estimated from direct and indirect measurements from different sources of in situ and satellite remote sensing data. In this section, we provide details on the different data sources used to train or evaluate our model: in situ measurements from drifters and from instruments onboard ships ([section 2a](#)) and SSH measurements from nadir-pointing altimeters and SWOT, as well as derived geostrophic currents ([section 2b](#)), in addition to visible and infrared satellite images of chlorophyll and sea surface temperature ([section 2c](#)). We also describe our data preprocessing pipeline ([section 2d](#)).

a. In situ measurements

We use two types of in situ measurements: sparse observations from drifters and estimations of ocean currents from instruments onboard ships.

Lagrangian drifters are buoys equipped with a drogue and a GPS and are advected by ocean currents. Our dataset contains hourly measurements of global drifters ([Ifremer 2024](#); [Elipot et al. 2022, 2016](#)). To smooth noise, for instance, due to surface winds and waves, we compute moving averages of observations over periods of 24 h. Drifters are sparsely and non-uniformly distributed, which can bias learning and evaluation. [Figure 2](#) illustrates the spatial distribution of drifters in the Mediterranean, Gulf Stream, and Agulhas regions in 2023.

Onboard instruments from ships also estimate the effect of currents, by measuring both speed through water (STW) and speed over ground (SOG). STW can be measured with instruments such as Wavex or acoustic Doppler current profiler (ADCP), and SOG can easily be measured by regular GPS localization. The difference between these measurements, $SOG - STW$, is an indicator of the effect of currents felt by ships along their trajectory. These data are limited to specific examples of ship voyages and thus are only used for evaluation purposes.

b. Satellite altimeter measurements

Far from the equator, the ocean circulation is primarily driven by its geostrophic component, which can be derived

from SSH. Two main types of satellite measure the SSH: nadir-pointing altimeters and the Ka-band radar interferometer (KaRIn) on board the SWOT satellite. Numerous interpolation methods exist to produce complete estimations of SSH and geostrophic currents such as [Taburet et al. \(2019\)](#) DUACS and [Scott et al. \(2024\)](#) NeurOST.

Nadir-pointing altimeters such as those used in various satellite missions like *Sentinel-6A*, *Jason-3*, and *CryoSat-2*, measure SSH along the satellite's track with a spatial resolution of approximately 7 km. The data are processed through the DUACS protocol, which ensures consistency across multiple altimeter missions by standardizing them to a reference mission ([Copernicus Marine Service 2023](#)).

KaRIn altimetry from SWOT provides exciting opportunities for the precise reconstruction of the SSH. SWOT provides 120-km-wide swaths of high-resolution observations ([NASA/JPL and CNES 2024](#)). This precision allows observation of mesoscale/submesoscale SSH patterns. However, only one satellite is presently in orbit, and its revisit time period is 21 days, which leaves important areas unobserved. We have access to around 1 year of SWOT observations, compared to decades of nadir-pointing altimeter data. [Figure 3](#) shows examples of not only fine-scale eddies visible from SWOT but also internal gravitational waves close to the equator. This illustrates not only SWOT's potential in estimating ocean surface currents but also its limitations, particularly in equatorial latitudes.

DUACS is a level 4 (L4) product, providing a low-resolution (0.125°) mapping of SSH through optimal interpolation ([Taburet et al. 2019](#)) of nadir altimetry. The geostrophic currents can then be derived from this mapping and are available at [Copernicus Marine Service \(2024\)](#) and [Copernicus Climate Service \(2024\)](#). For evaluation and stage-1 training, we use a delayed-time DUACS product, which uses 6 days of future observations from nadir tracks.

NeurOST SSH is a state-of-the-art method for predicting global SSH at 0.1° resolution. [Martin et al. \(2024\)](#) train a neural network to predict nadir-pointing along-track SSH from SST and SSH inputs, using only observations in a self-supervised manner. This product is produced in a delayed time, integrating 14 days of future observations.¹

c. Satellite imagery

SST and CHL concentrations are passive tracers of ocean currents, as water motion advects these scalar variables. Small-scale structures in ocean surface currents are visible using high-resolution satellite imagery of SST and CHL. However, both infrared and visible waves are blocked by clouds, and so information provided from these sensors is partial. Although methods exist for interpolating SST and CHL, we use L3 and

¹ The authors of NeurOST have informed us that there were some issues in the production of these data, potentially reducing quality, and that there will be a new release in the near future. We were unaware of this issue at the time of analysis and have used the NeurOST data as provided in November 2024.

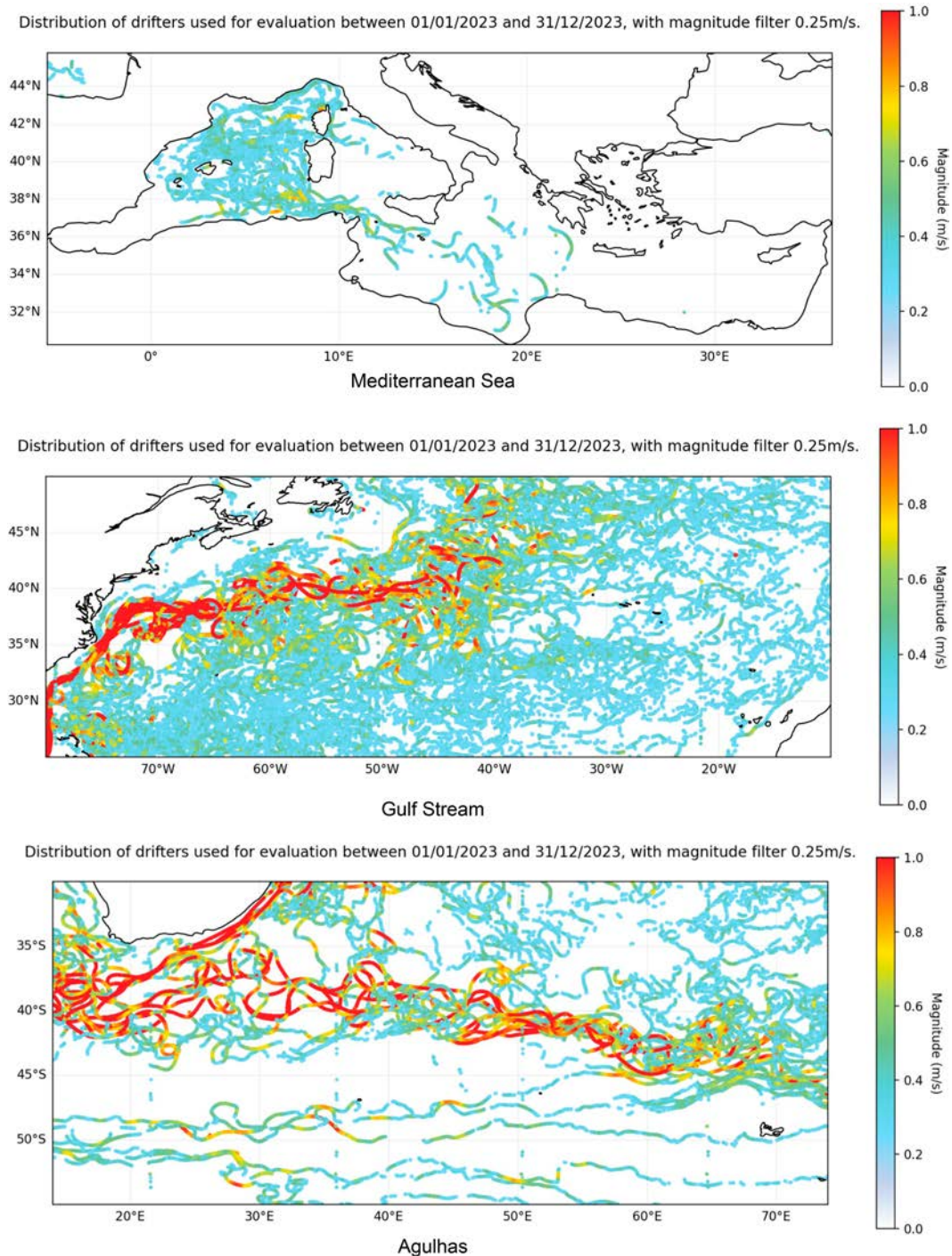


FIG. 2. In situ observations of currents from drifters in three key regions in 2023, used for evaluation.

L2 products and allow our model to learn to combine these different sources of information from past days.

SST is measured using infrared sensors at a high spatial resolution. We use the daily L3 products (GLO 2024; Jonasson et al. 2024; MET Norway 2023), with a resolution of 0.02° .

CHL is measured using data from the Ocean and Land Color Instrument (OLCI) aboard *Sentinel-3* satellites, which

provides ocean color geophysical products. The chlorophyll product is provided using the Ocean Colour for MERIS (OC4Me) algorithm (O'Reilly et al. 1998). This algorithm calculates the concentration of algal pigments by analyzing water-leaving reflectances across 21 spectral bands. The chlorophyll concentrations are provided in log10-scaled values. We use data available at 1.2-km resolution (EUMETSAT for Copernicus 2024).

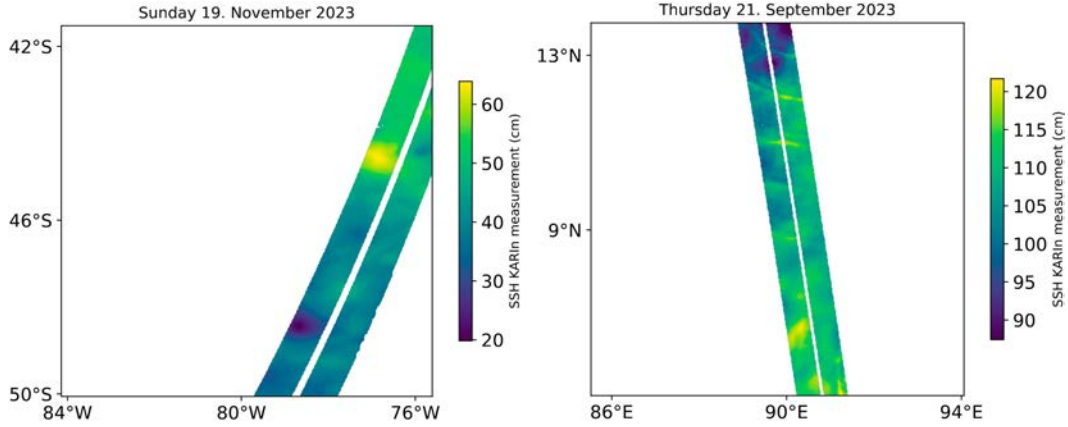


FIG. 3. SWOT SSH observation, consisting of two 50-km-wide bands, separated by a 20-km gap covered by traditional nadir altimetry instruments. (left) Visible cyclonic and anticyclonic eddies at high resolution. (right) Visible internal gravitational waves. In particular, close to the equator, it is more difficult to estimate ocean surface currents from SSH as a result of the reduced Coriolis effect.

d. Data processing pipeline

Our data processing pipeline interpolates all data sources into a common grid at a resolution of $1/30^\circ$. We compute means and standard deviations of the key variables—SSH, U and V components of ocean currents, SST, and CHL—at a spatial resolution of 2° and a temporal resolution of 1 week. To estimate the means and standard deviations of SSH and ocean surface currents, we use the DUACS L4 product. We normalize all variables using these values to remove average seasonal and spatial effects.

3. Method

ORCAst is a deep learning model to forecast ocean currents. Our model takes as input T time steps of past and present satellite observations of SSH and optical satellite data and predicts τ future time steps of SSH along with the U and V components of ocean surface currents. We learn the parameters θ of the neural network f_θ , shown in Eq. (1):

$$f_\theta(X^{0:T}) = \hat{Y}_{\text{SSH}}^{T+1:T+\tau}, \hat{Y}_{U,V}^{T+1:T+\tau}, \quad (1)$$

where X designates past and present observations and \hat{Y} designates the forecasted fields. The inputs and outputs are gridded regional crops of dimension $h \times w$ and at identical spatial resolutions. In practice, we input $T = 11$ days of past and present satellite observations and output $\tau = 7$ future days of satellite observations. Inputs X and outputs \hat{Y} are daily images of size 128×128 at $1/30^\circ$ resolution.

In section 3a, we present our multiarm encoder–decoder model architecture, and in section 3b, we describe our three-stage training procedure. Section 3c provides the geographic zones used for training and evaluation, and section 3d outlines our evaluation metrics and our baseline comparisons.

a. Model architecture

Our model architecture, illustrated in Fig. 4, is based on simple video prediction (SimVP), a flexible encoder–decoder

architecture initially designed for video prediction (Gao et al. 2022). A similar architecture is used to reconstruct a time series of SSH from nadir altimetry and SST inputs by Martin et al. (2024). The ORCAst model has the following key features:

- one 2D encoder per input variable, allowing the model to best exploit each variable and time step with fewer parameters;
- one 2D decoder per output variable, allowing us to remove decoder arms during training [e.g., in stage 3 (S3) of our training procedure, described in section 4a], and also reducing the number of parameters;
- we replace the deep inception module from Gao et al. (2022) with a gated spatiotemporal attention (GSTa)-based module for better efficiency; and
- we add a spatiotemporal positional encoding module before the temporal translator, which allows the model to learn regional and seasonal phenomena.

1) 2D UNIVARIATE ENCODERS

The network encodes each variable and time step separately, using different encoders (f_{θ_i} in Fig. 4) and dividing the spatial dimensions by 4. The same encoder is applied to each time step, but different encoders encode each physical variable, enabling the network to learn the different characteristics of each input data source. The encoders first apply a 4×4 convolution, followed by a group normalization (GroupNorm) from Wu and He (2020) and a Gaussian error linear unit (GELU) activation function. Then, a strided convolution divides spatial dimensions by 2. This module is then repeated to produce the final latent space.

2) SPATIAL–TEMPORAL EMBEDDING

After the encoder, we apply a positional encoding, i.e., a learnable distribution shift of the latent image informing the network about the spatiotemporal coordinates of the patch (see appendix B). The temporal dimension is then placed in

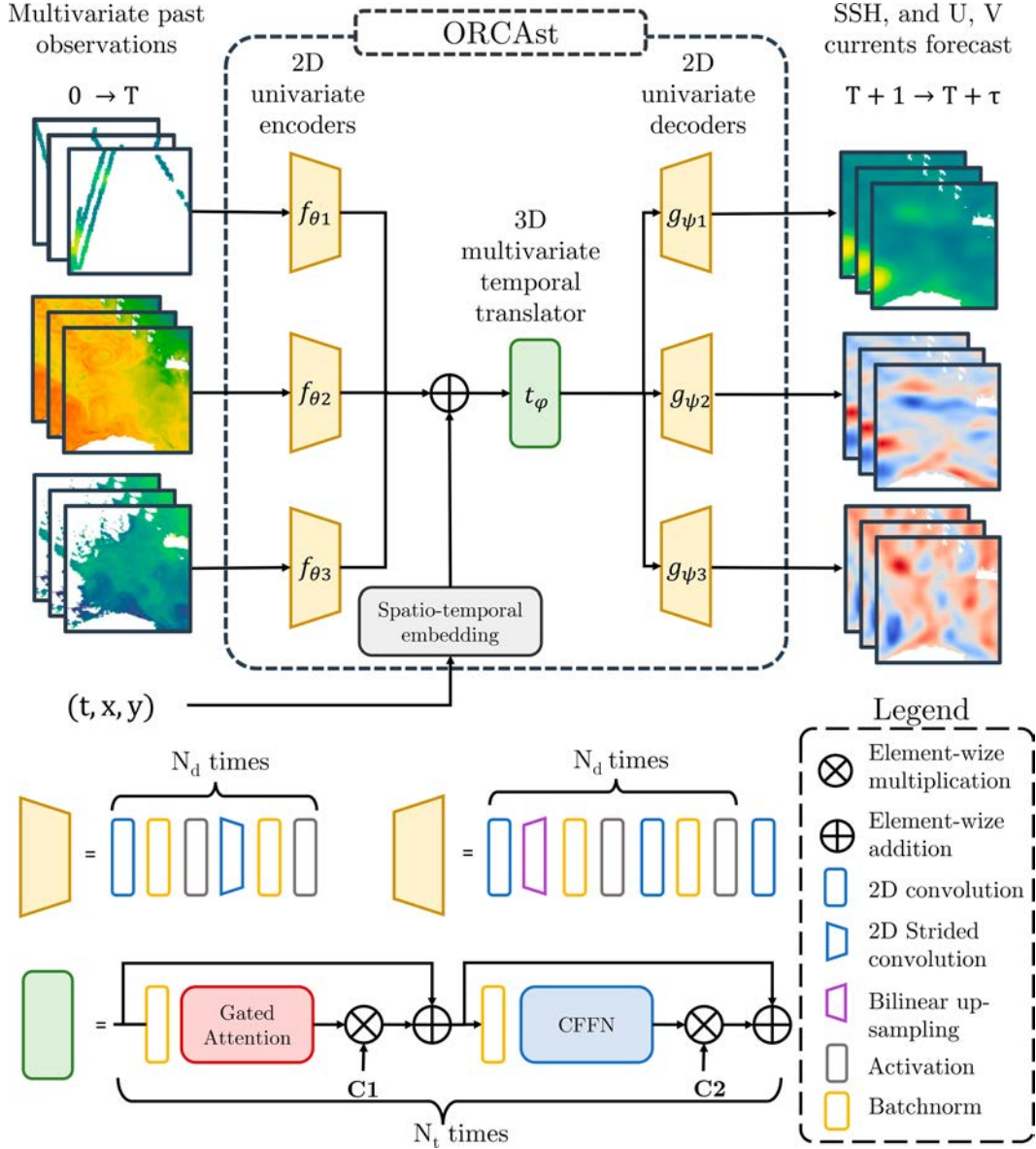


FIG. 4. Overview of ORCAst. Our models input temporal sequences of multivariate satellite observations (here, SSH, SST, and CHL) from $t = 0$ to $t = T$. Each variable is input to a different spatial encoder, which extracts information from each time step separately. A spatiotemporal positional embedding is applied to inform the network of geographical coordinates and seasonal information. A spatiotemporal translator t_ϕ is used to produce the forecast encodings. Finally, each time step is decoded independently with three univariate decoders g_ψ , forecasting SSH, and U and V components separately.

the variable dimension, meaning that the model can subsequently access temporal dependencies through the variable dimension.

3) 3D MULTIVARIATE TEMPORAL TRANSLATOR

The GSTa module (t_ϕ in 4) is composed of a sequence of gated attention subblocks. As the temporal dimension is merged with variables, GSTa is able to learn the dynamics of the time series by multivariable convolutions. A block of the GSTa model consists of a gated attention module, followed by a convolutional feed-forward network (CFFN) from Wang et al. (2022), used

with residual connections and learned rescaling, schematized by t_ϕ in Fig. 4. This module contains dilated convolutions and an attention gating process, allowing the model to learn long-range spatial interactions. For further details about the GSTa module, we refer the reader to Yu et al. (2022).

4) 2D UNIVARIATE DECODER

After the multivariate temporal translator, we apply three 2D univariate decoders to decode SSH, the zonal current component U , and the meridional current component V . Similarly

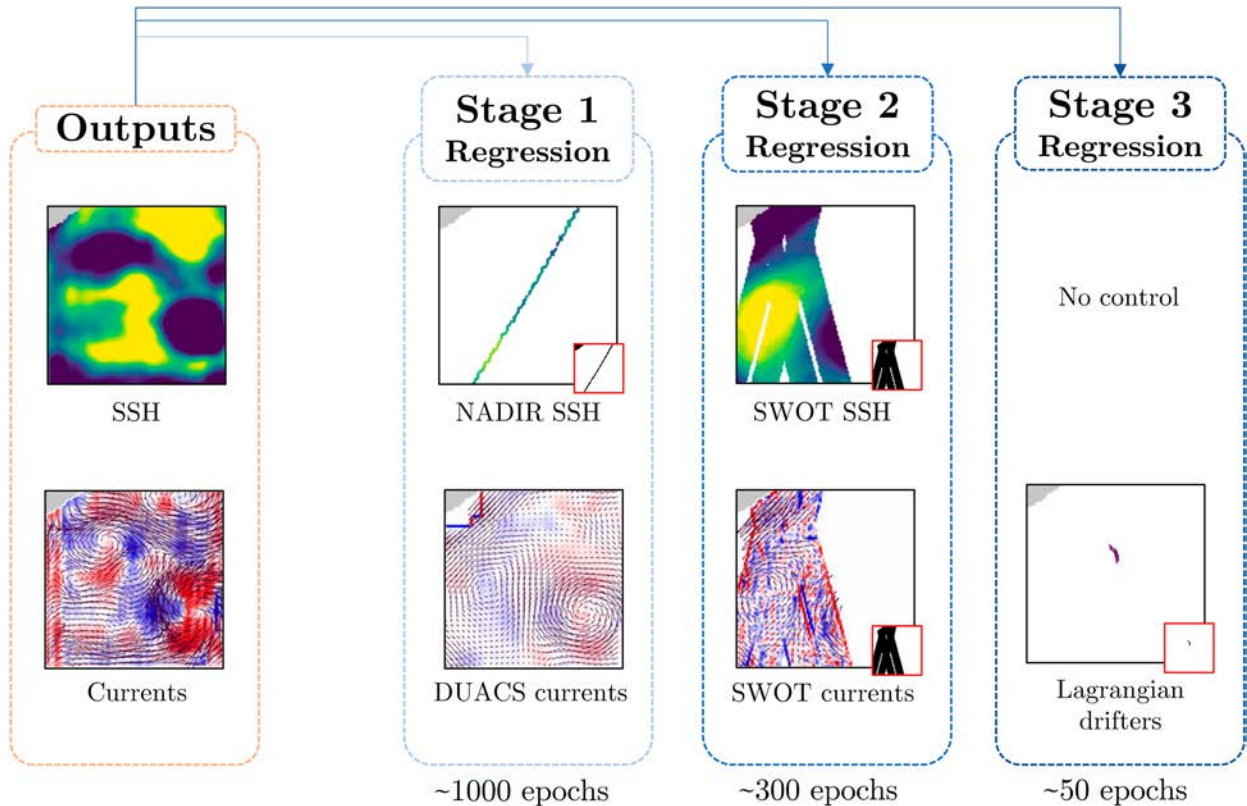


FIG. 5. ORCAst training stages. (from left to right) An output example of SSH and U and V components of ocean surface currents, represented as a vector field with colors denoting vorticity. S1 shows nadir and DUACS regression targets, where the small red box illustrates the mask of pixels taken into account when computing the loss. S2 shows SWOT regression targets, with associated masks on the loss function in the small red box. Note that in S3, we freeze the SSH decoder and train the model to learn only U and V .

to the encoders, the decoders g_{ϕ_i} apply the same operation to each time step but have different weights for each variable. They are composed of 4×4 convolutions followed by bilinear upsampling layers. These operations are repeated twice to obtain the original image dimensions.

b. Multistage training

Training a model that leverages the diverse range of oceanographic observations—such as drifters, satellite altimetry, and satellite imagery—is challenging due to the differing temporal coverage and resolutions of these data sources. One might consider developing a model to forecast ocean currents and SSH by directly comparing predictions to all available measurements, using drifter velocities as targets for currents and combining observations from nadir and SWOT altimetry for SSH forecasting. However, this approach presents several significant challenges. First, observations from different sources are not always consistent; for example, inconsistencies have been identified in some SWOT and nadir crossover points, as discussed in [appendix B](#). Second, the temporal coverage varies greatly between datasets: SWOT provides only about a year of data, while nadir altimetry has accumulated observations over several decades ([Taburet et al. 2019](#)). Third, the density of observations differs across sources, with drifter data being far sparser than satellite measurements,

making it difficult to appropriately weight each data source during model training.

Considering these difficulties, we train our model using a three-stage training strategy, illustrated in [Fig. 5](#). The goal of this strategy is to learn a first approximation of SSH and ocean currents using nadir-pointing SSH and derived geostrophic currents from interpolated SSH fields as ground truth, and to progressively refine the learning using scarcer, but more precise or direct observations (SWOT and drifters). The inputs of the neural network are the same at each stage: nadir SSH and SST. We also include experiments on using additional CHL or SWOT observation as inputs. Our strategy contains three stages of training:

- 1) Our model estimates along-track nadir SSH and DUACS delayed-time geostrophic currents. As the nadir-pointing observations are sparse, we first perform a masking operation before computing the loss, so that the network can be trained with incomplete fields. For further details about this methodology, we refer the reader to [Filoche et al. \(2022\)](#), [Archambault et al. \(2023\)](#), [Martin et al. \(2023\)](#), and [Archambault et al. \(2024b\)](#). This corresponds to stage 1 (S1) of [Fig. 5](#). In [appendix C](#) we experiment with using another DL model's output to replace DUACS targets, but we keep DUACS results as the model does not really benefit from these new targets.

TABLE 1. Geographical definitions of oceanic regions by latitude and longitude.

Region	Latitude range	Longitude range
Global (extratropical)	−60° to −20° and 20°–60°	−180° to 180°
Mediterranean Sea	30°–46°	−6° to 36°
Gulf Stream	20°–45°	−99° to −34°
Agulhas	−55° to −30°	14°–74°

- 2) Our model estimates high-resolution SWOT data of SSH and associated geostrophic currents. As SWOT produces large bands of observations, it is possible to compute spatial gradients and thus retrieve the geostrophy directly from observations without interpolation methods. However, SWOT provides partial spatial coverage, so we apply the masking operation to both SSH and surface current outputs of ORCAst before computing the loss. This corresponds to stage 2 (S2) of Fig. 5.
- 3) The final stage involves fine-tuning only the output currents, while freezing the SSH decoder. In this stage, the regression of ORCAst currents fits to sparse but direct drifter measurements of U and V , thereby allowing the model to learn from the total currents rather than a geostrophic approximation. This corresponds to stage 3 of Fig. 5.

At each stage, we train the model using mean-square-error (MSE) loss. In order for the model to learn to reconstruct structures and fast-moving ocean currents, we weight the MSE loss function proportionally to the magnitude of ocean currents in the target observations, following Kugusheva et al. (2024). We use an Adam optimizer (Kingma and Ba 2015) with the learning rate set to 10^{-3} and weight decay to 10^{-3} in stage 1 and resume training by reducing the learning rate by a factor of 10 in stages 2 and 3. We set each epoch to 1000 randomly chosen patches. During the three stages, the model is trained using 1000, 200, and 50 epochs, respectively. When training in the smaller area of the Mediterranean Sea (described in section 3c), we train for 500, 200, and 25 epochs, respectively, over the three stages. The number of epochs was chosen via a validation set. We merge patches using Gaussian kernel-weighted averaging, as detailed in appendix A of Callaham et al. (2019).

When training the model on a single A100 graphics processing unit (GPU), running 100 epochs with 1000 samples each

requires about 2 h. The first stage of training therefore lasts roughly a day, the second stage lasts a few hours, and the third stage lasts about an hour. At inference time on the same GPU, generating current forecasts over the entire global area with a spatial stride of 64 pixels and a patch size of 128×128 pixels typically completes in around 5 min.

Stages 1 and 2 of our analysis are trained using geostrophic currents, which are the dominant components of ocean surface currents far from the equator. Tropical latitudes, where the Coriolis effect is weaker and other fast ageostrophic dynamics such as internal gravitation waves (Fig. 3) or equatorial Rossby and Kelvin waves become more significant, are excluded from our analysis. See Fig. A1 in appendix A for a further illustration of the importance of latitude.

c. Geographic zones and time periods

The study area is divided into several subregions, defined in Table 1. We exclude equatorial regions as we train using geostrophic currents in stages 1 and 2, which are only a reasonable approximation of the total currents far from the equator. To compare the performance of regional-specific models and global models, we train our model on three regions: the Mediterranean Sea, the Gulf Stream, and the Agulhas regions (Table 1).

We train our model using data from 2016 to mid-December 2018 and from mid-January 2020–22 during stages 1 and 3. In stage 2, we use 6 months of available SWOT data from 2024, until July. We use 2023 to evaluate our model's performance.

d. Evaluation metrics and baselines

We evaluate our results on sparse in situ observations collected from Lagrangian drifters in the year 2023 (see Fig. 2). As we are interested in detecting eddies and other structures with fast-moving currents, we restrict our drifter evaluation set to observations above 0.25 m s^{-1} . In the following, let $\hat{\mathbf{w}}$ be the velocity vector predicted by the model at the position of the drifter and $\mathbf{w}_{\text{drifter}}$ be the velocity vector observed by the drifter. We measure prediction quality using three metrics, illustrated in Fig. 6:

- Percentage of correct angles: Let θ be the angle difference between vectors $\hat{\mathbf{w}}$ and $\mathbf{w}_{\text{drifter}}$. Then,

$$\theta = \frac{180}{\pi} \cos^{-1} \left(\frac{\hat{\mathbf{w}} \cdot \mathbf{w}_{\text{drifter}}}{\|\hat{\mathbf{w}}\| \|\mathbf{w}_{\text{drifter}}\|} \right) \in [0; 180].$$

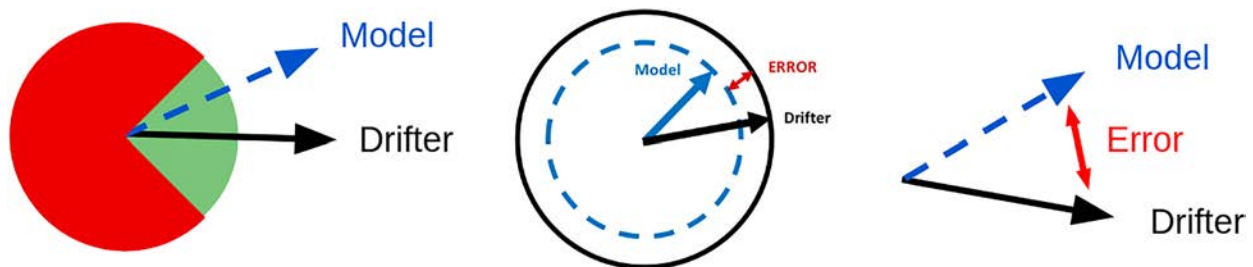


FIG. 6. Simplified illustration of our three metrics: (from left to right) angle error, magnitude error, and error vector amplitude when compared to drifter measurements.

TABLE 2. Evaluation on the extratropical global area. Our near-real-time model outperforms existing methods for forecasting ocean surface currents, including DT methods evaluated using persistence (pers.) forecasts. See [section 4d](#) for a description of our metrics. Values are rounded. We evaluate on drifter observations from 2023 with magnitude $> 0.25 \text{ m s}^{-1}$. Best results are shown in bold font.

	Correct angle %		Correct magnitude %		MEVA (cm s^{-1})	
	$T + 1$	$T + 7$	$T + 1$	$T + 7$	$T + 1$	$T + 7$
DUACS (DT pers.)	78	62	69	68	25	37
NeurOST (DT pers.)	83	65	72	67	25	31
Mercator	70	56	68	68	25	37
ORCAst	85	70	77	69	24	30

If $\theta \leq 45^\circ$, we consider the current direction as correct. We then compute the percentage of correctly predicted angles in our drifter evaluation set.

- Percentage of correct magnitude: Let ΔM be the magnitude difference between $\hat{\mathbf{w}}$ and $\mathbf{w}_{\text{drifter}}$. Then,

$$\Delta M = ||\hat{\mathbf{w}}|| - ||\mathbf{w}_{\text{drifter}}|| \in \mathbb{R}^+.$$

If $\Delta M \leq 2.5 \text{ cm s}^{-1}$, we consider the current magnitude as correct. We then compute the percentage of correctly predicted magnitudes in our drifter evaluation set.

- Mean error vector amplitude (MEVA): The error vector amplitude is the velocity difference ΔV between $\hat{\mathbf{w}}$ and $\mathbf{w}_{\text{drifter}}$, i.e.,

$$\Delta V = ||\hat{\mathbf{w}} - \mathbf{w}_{\text{drifter}}|| \in \mathbb{R}^+.$$

We compute the mean of the error vector amplitude, or the MEVA, across drifter observations in our evaluation set.

These three metrics for our predicted ocean surface current forecasts are calculated at time $T + 1$ and $T + 7$ days. Our model performance is compared to three alternative methods: DUACS, NeurOST, and Mercator. As DUACS and NeurOST do not provide a forecasting method, we compare with persistence forecasting, i.e., compare our model's forecasts at time $T + 7$ with DUACS and NeurOST predictions at time $T + 1$. Note that our method is operational, and we use only data up until time T to predict forecasts of $T + 1$ until $T + 7$. We are thus at a disadvantage when compared to 6-day delayed-time ocean surface currents from DUACS or 14-day delayed-time ocean surface currents from NeurOST. Mercator ([Drévilion et al. 2008](#); [Mercator Océan International2024](#)) provides operational near-real-time forecasts of ocean surface currents, and we compare our ocean surface current forecasts to Mercator forecasts.

4. Results

In this section, we provide quantitative and qualitative results of our ocean current forecast model, using independent in situ observations from drifters from 2023. [Section 4a](#) presents our results on the extratropical global area and the importance of our proposed three-stage training procedure. [Section 4b](#) details model performance when trained at a global scale or

trained specifically on regional areas. In [section 4c](#), we conduct ablation studies experimenting with CHL and SST inputs, as well as SWOT data as inputs. [Section 4d](#) illustrates the qualitative performance of our model, compared to baselines. Finally, [section 4e](#) evaluates our model on real ship data, demonstrating a potential application of forecasting ocean currents in the maritime industry.

a. Global forecast

[Table 2](#) presents the performance of our model, trained and evaluated over the extratropical global area, using the metrics and baselines outlined in [section 4d](#). Our method consistently outperforms the baselines in both next-day ($T + 1$) and 7-day ($T + 7$) forecasts. For example, at $T + 1$, our model achieves 85% of correct angles, compared to 78% for DUACS, 83% for NeurOST, and 70% for Mercator. Similarly, for $T + 7$ forecasts, our model reaches 70% of correct angles, outperforming DUACS (62%), NeurOST (65%), and Mercator (56%). These results highlight that despite being an operational, near-real-time forecasting model, ORCAst surpasses the delayed-time methods DUACS (6-day delay) and NeurOST (14-day delay) while also significantly improving over the near-real-time numerical model Mercator. This demonstrates the ability of our deep learning-based approach to effectively capture physical relationships from observational data.

The contribution of the three training stages to the performance of ORCAst is shown in [Table 3](#), with each stage progressively enhancing the model's accuracy. During stage 2, incorporating SWOT data allows the model to leverage high-resolution SSH observations, improving the percentage of correct angle predictions from 79% in stage 1 to 83% in stage 2. Fine-tuning in stage 3 using in situ drifter measurements further refines the model's ability to capture ocean surface currents beyond the geostrophic component. Stage 3 boosts the percentage of correct angles to 85%. Similarly, the 7-day ($T + 7$) forecast accuracy improves steadily, with the percentage of correct angles increasing from 64% in stage 1 to 68% in stage 2 and reaching 70% in stage 3. The model also experiences a substantial improvement in magnitude performance across the stages. For next-day ($T + 1$) forecasts, the percentage of correct magnitudes increases significantly from 59% in stage 1 to 73% in stage 2 and finally to 77% in stage 3. These results emphasize the critical role of each stage in enabling the model to effectively utilize diverse observational data for more accurate ocean current predictions.

TABLE 3. Importance of three-stage training. Each training stage brings additional improvement in the considered metrics. See [section 4d](#) for a description of our metrics. Values are rounded. We evaluate on drifter observations from 2023 with magnitude $> 0.25 \text{ m s}^{-1}$. Best results are shown in bold font.

	Correct angle %		Correct magnitude %		MEVA (cm s^{-1})	
	$T + 1$	$T + 7$	$T + 1$	$T + 7$	$T + 1$	$T + 7$
ORCAst stage 1	79	64	59	57	30	50
ORCAst stage 2	83	68	73	68	27	34
ORCAst stage 3	85	70	77	69	24	30

The importance of stage 2 in learning fine-scale structures is shown in [Fig. 7](#). During stage 1, the model is trained to predict large-scale ocean structures using nadir altimetry data and is not penalized for failing to detect structures below the effective resolution of nadir altimetry ($\sim 100 \text{ km}$). In stage 2, the model is further refined to predict smaller-scale structures, visible in the SST and CHL signatures, by training on high-resolution SWOT altimetry data with a lower effective resolution ($\sim 15 \text{ km}$). We refer to [appendix B](#) for an ablation study on using NeurOST instead of DUACS in stage 1.

b. Regional forecasts

This section explores the differences between training models on a large, global scale and using region-specific models. Results in [Tables 4–6](#) show that regional models consistently outperform the global model in the Mediterranean Sea, Gulf Stream region, and Agulhas region. Regional ORCAst models benefit from focusing on the unique physical characteristics of their specific areas, while the global ORCAst model must generalize across a wide variety of ocean dynamics. Although our global model incorporates geographical information through positional encoding, this alone is insufficient to

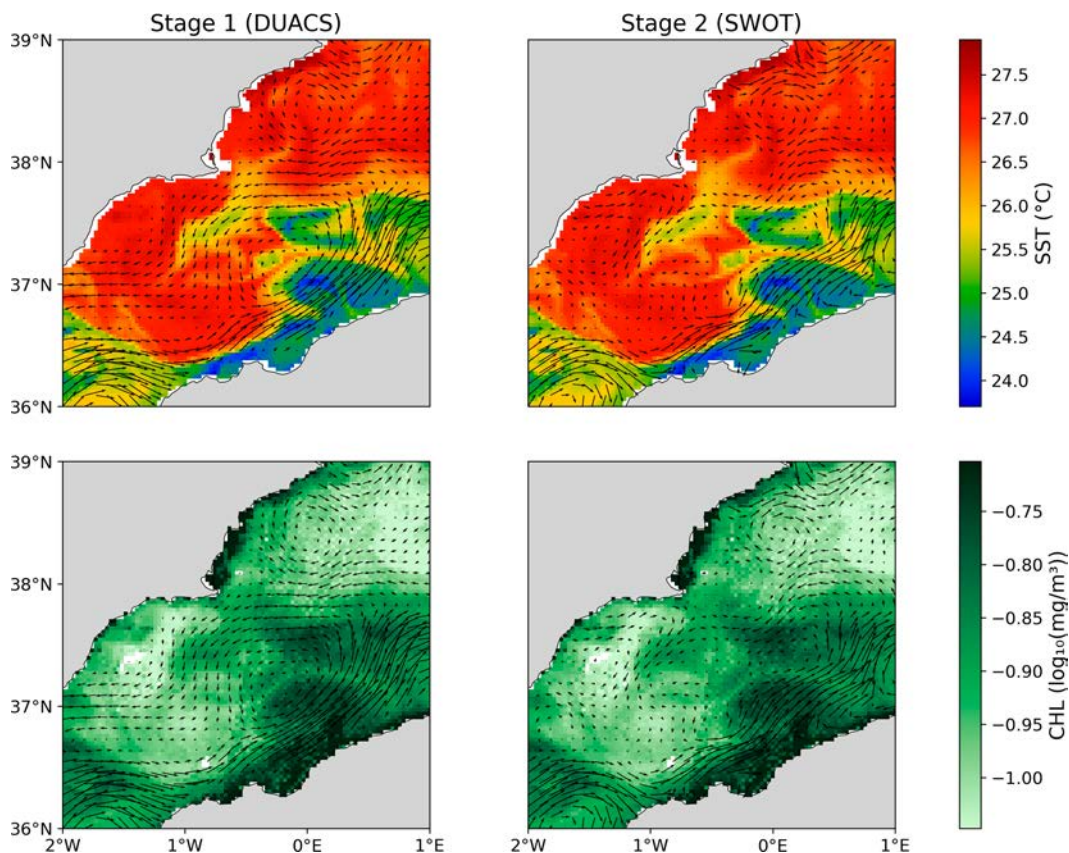


FIG. 7. Example of performance of ORCAst in the Mediterranean Sea on 31 Aug 2023. (top) SST and (bottom) CHL are shown in the background. Black arrows denote predicted ocean surface currents after (left) S1 and (right) S2 of the regional model trained in the Mediterranean Sea (see [Table 4](#)). In S1, the model is trained to predict large-scale structures using nadir altimetry data. In S2, the model is trained with fine-scale SWOT altimetry data, which allows it to be better rewarded for accurately predicting small-scale structures seen in the SST signature.

TABLE 4. Evaluation in the Mediterranean Sea. Our model trained specifically on the Mediterranean Sea outperforms ORCAst trained on the extratropical global area and also achieves higher accuracy than alternative methods. See [section 4d](#) for a description of our metrics. Values are rounded. We evaluate on drifter observations from 2023 with magnitude $> 0.25 \text{ m s}^{-1}$. Best results are shown in bold font.

	Correct angle %		Correct magnitude %		MEVA (cm s^{-1})	
	$T + 1$	$T + 7$	$T + 1$	$T + 7$	$T + 1$	$T + 7$
DUACS (DT pers.)	58	55	51	51	29	31
NeurOST (DT pers.)	74	63	65	62	25	28
Mercator	51	43	64	60	32	35
ORCAst global	77	58	78	59	26	30
ORCAst regional	85	70	88	83	21	25

fully resolve the nuanced differences across regions at the same capacity, as the model is still required to generalize across a much wider range of physical and environmental conditions, increasing the difficulty of the learning task.

For instance, the Mediterranean Sea is an enclosed basin with lower energy levels and distinct physical characteristics compared to the open ocean. The Mediterranean Sea also features smaller-scale structures and a reduced Rossby radius, which partly explains why the regional Mediterranean model achieves a significant improvement over the globally trained model. As shown in [Table 4](#), the regional model achieves 85% correct angle predictions at time $T + 1$, compared to 77% for the global model, and 88% of correct magnitudes at time $T + 1$, compared to 78% for the global model. For 7-day lead times, improvements are even more significant. At $T + 7$, the Mediterranean model achieves 70% correct angles, compared to 58% for the global model, and 83% correct magnitudes, compared to 59% for the global model. This substantial improvement in both angle and magnitude accuracy underscores the advantage of focusing on the region's unique dynamics.

In contrast, the Gulf Stream and Agulhas regions are highly energetic regions, presenting a unique challenge for forecasting. The regional models outperform the global model, particularly at 7-day lead times ($T + 7$), showcasing the ability of regional models to capture local dynamics ([Tables 5 and 6](#)). In the Gulf Stream region, the regional model achieves 88% correct angle predictions, outperforming the global model at 85% and surpassing the performance of delayed-time methods such as DUACS (83%) and NeurOST (86%), as well as near-real-time Mercator (77%) ([Table 5](#)). Furthermore, at $T + 7$, the regional model predicts angles with 73% accuracy, compared to

66% for the global model. Similarly, in the Agulhas region, the regional model achieves an angle accuracy of 93% at time $T + 1$, compared to 92% for the global model ([Table 6](#)). At $T + 7$, the regional model reaches 78% accuracy in angle prediction, surpassing the global model's 73% and indicating its superior ability to capture the region's ocean dynamics over extended time frames. These results emphasize the advantage of training on region-specific data for accurate, long-term forecasts in highly energetic ocean regions. We refer to [appendix C](#) for metrics on eddy kinetic energy (EKE) in these regions.

[Figure 8](#) illustrates how ORCAst learns to accurately predict the evolution of a dynamic structure in the Agulhas region. The Mercator model incorrectly localizes the structure but predicts a westward evolution. In contrast, ORCAst not only correctly identifies the location of the structure but also accurately forecasts its position in 7 days. This ability to both localize and predict the evolution of such features highlights the strength of ORCAst in learning complex ocean dynamics.

c. Including more input data

The flexibility of ORCAst allows us to include various inputs. In this section, we present experiments where we evaluate ORCAst using CHL or SWOT data as inputs, in addition to SST and nadir altimetry. [Table 7](#) compares model performance using CHL as additional inputs, and [Table 8](#) compares model performance using SWOT as additional inputs.

[Table 7](#) suggests that there is limited potential of CHL to provide complementary high-resolution information on ocean currents, when used in combination with high-resolution SST data. Scores for the model both with and without CHL are similar.

TABLE 5. Evaluation in the Gulf Stream region. Our model trained specifically on the Gulf Stream region outperforms ORCAst trained on the extratropical global area and also achieves higher accuracy than alternative methods. See [section 4d](#) for a description of our metrics. Values are rounded. We evaluate on drifter observations from 2023 with magnitude $> 0.25 \text{ m s}^{-1}$. Best results are shown in bold font.

	Correct angle %		Correct magnitude %		MEVA (cm s^{-1})	
	$T + 1$	$T + 7$	$T + 1$	$T + 7$	$T + 1$	$T + 7$
DUACS (DT pers.)	83	61	71	65	29	42
NeurOST (DT pers.)	86	63	72	65	25	37
Mercator	66	52	65	64	38	41
ORCAst global	85	66	74	66	27	34
ORCAst regional	88	73	77	70	24	31

TABLE 6. Evaluation in the Agulhas region. Our model trained specifically on the Agulhas region outperforms ORCAst trained on the extratropical global area and also achieves higher accuracy than alternative methods. See [section 4d](#) for a description of our metrics. Values are rounded. We evaluate on drifter observations from 2023 with magnitude $> 0.25 \text{ m s}^{-1}$. Best results are shown in bold font.

	Correct angle %		Correct magnitude %		MEVA (cm s^{-1})	
	$T + 1$	$T + 7$	$T + 1$	$T + 7$	$T + 1$	$T + 7$
DUACS (DT pers.)	87	67	71	67	29	35
NeurOST (DT pers.)	91	70	75	67	26	35
Mercator	77	62	64	66	37	38
ORCAst global	92	73	77	70	25	33
ORCAst regional	93	78	80	73	23	30

We compare ORCAst models using nadir-pointing altimetry only or nadir and SWOT as SSH inputs, in addition to SST. As SWOT data are only available from mid-2023 onward, we train our model using SWOT inputs during stage 2 and evaluate our model using August–December 2023. We keep the training procedure of stage 1 as before, using only SST and nadir altimetry as inputs. Results of this model at the end of stage 2 are provided in [Table 8](#). Using SWOT data as inputs seemingly does not help the model to better predict ocean surface currents. However, since there is only around 1 year of

available SWOT data, the training and test sets are small, making it difficult to draw strong conclusions from this experiment. SWOT observations reveal unobserved mesoscale and submesoscale structure; however, the long revisit time and limited available data from the mission may partially explain the lack of performance gains when used as additional inputs.

d. Qualitative evaluation

An example of ORCAst predicted currents at time $T + 1$ in the Mediterranean Sea is shown in [Fig. 9](#), alongside comparisons

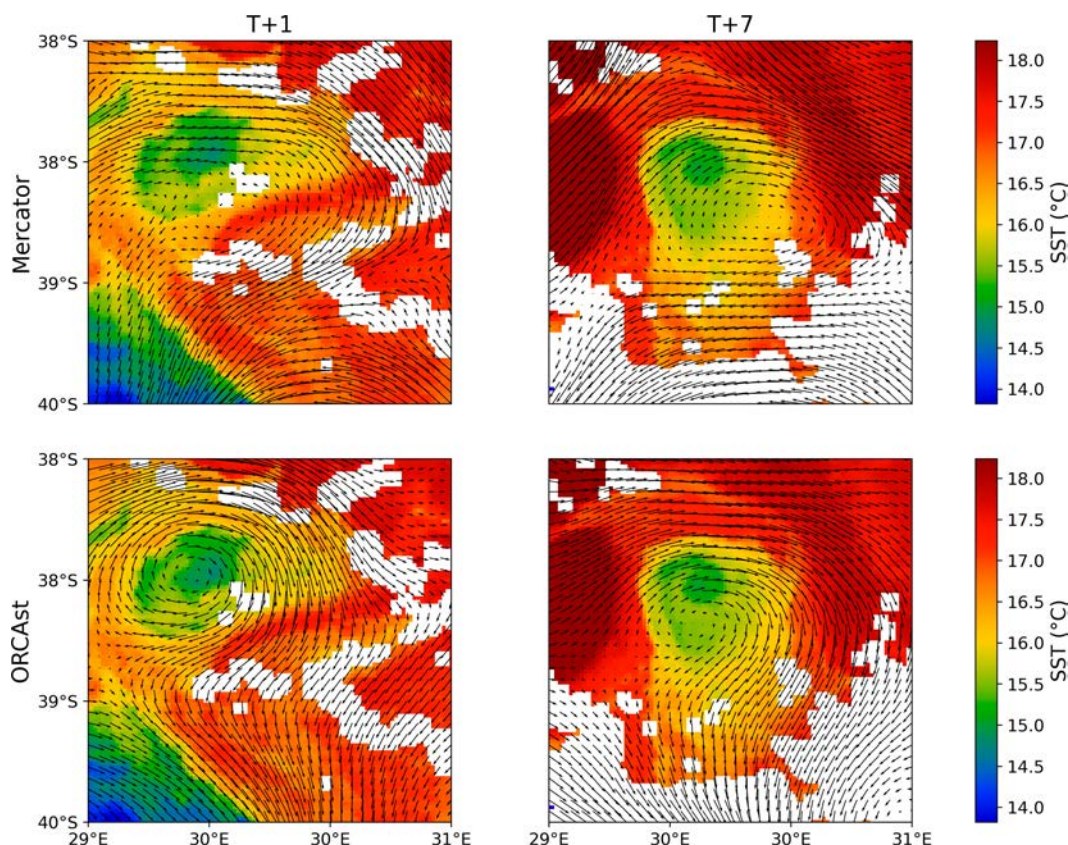


FIG. 8. Predicted evolution of a structure in the Agulhas region between 11 and 17 Oct 2023. SST is shown in the background, with black arrows denoting the ocean surface currents. The Mercator model misidentifies the structure's location but predicts an eastward movement. In contrast, ORCAst both accurately locates the structure and provides a more precise 7-day forecast of its future position.

TABLE 7. Comparison of ORCAst performances on the extratropical global area, using SST or SST + CHL as inputs, in addition to nadir SSH. Including CHL data does not bring important improvements to ORCAst currents. See section 4d for a description of our metrics. Values are rounded. We evaluate on drifter observations from 2023 with magnitude $> 0.25 \text{ m s}^{-1}$. Best results are shown in bold font.

	Correct angle %		Correct magnitude %		MEVA (cm s^{-1})	
	$T + 1$	$T + 7$	$T + 1$	$T + 7$	$T + 1$	$T + 7$
ORCAst	85	70	77	69	24	30
ORCAst + CHL	85	70	76	72	27	29

with delayed-time models from DUACS and NeurOST, as well as ocean surface currents from Mercator. SST and CHL images are shown in the background. Although DUACS and NeurOST tend to correctly localize large-scale structures, some smaller-scale structures may be missed. In the case of DUACS, this is because the method lacks high-resolution information from SST. In the case of NeurOST, this is possibly because the model learns from nadir altimetry, with a low effective resolution, and is thus not sufficiently penalized for not learning fine-scale information visible in the SST. Figure 7 illustrates the difference between training on nadir altimetry and high-resolution SWOT altimetry.

A qualitative comparison of our model outputs to drifter observations at days $T + 1$ and $T + 7$ is shown in Fig. 10. Like DUACS and NeurOST, ORCAst correctly localizes the eddy at $T + 1$. However, persistence forecasts for DUACS and NeurOST do not align well with the drifter's motion at time $T + 7$. In contrast, ORCAst correctly predicts the drifter's trajectory at time $T + 7$. Mercator fails to localize the eddy at $T + 1$, as well as its evolution at time $T + 7$.

e. Evaluation using ship data

Ships generally measure their STW and SOG, and the difference between these measurements, $\text{SOG} - \text{STW}$, serves as an indicator of the currents experienced by ships along their trajectory.

In Fig. 11, we evaluate our model using ship data from a trans-Mediterranean voyage between 21 and 26 August 2023. We compute $\text{SOG} - \text{STW}$ and compare these observations to forecasted ocean currents along the ship's trajectory, as provided by the ORCAst and Mercator models. Our findings indicate that ORCAst forecasts align more closely with the observed $\text{SOG} - \text{STW}$ values than Mercator forecasts, suggesting superior performance. Correct estimations of the currents felt along ship routes are important for maritime route

optimization, allowing ships to take routes following more favorable ocean currents and avoiding currents against the path of the ship.

Moreover, for this trans-Mediterranean voyage, we compute the Pearson correlation coefficient and the R^2 score between the current impacts predicted by DUACS, NeurOST, Mercator, and ORCAst. We use delayed-time predictions for DUACS and NeurOST, corresponding to 21 August 2023. We use operational forecasts for Mercator and ORCAst, using only data available at or before 20 August 2023. As the Mediterranean crossing takes 5 days, we use 6 days of forecasts. Table 9 shows that ORCAst performs best, even when compared to delayed-time methods.

5. Conclusions

a. Discussion

In this study, we introduce ORCAst, a deep learning framework for forecasting ocean surface currents. ORCAst produces $1/30^\circ$ high-resolution 7-day forecasts of ocean currents in global extratropical regions. Our flexible approach allows us to fuse multivariate observations of the ocean such as in situ measurements of currents, satellite altimetry, and satellite imaging of sea surface temperature and chlorophyll *a*, to improve the spatial resolution and prediction of the temporal evolution of surface currents.

For next-day predictions, we accurately predict the angle of drifters 85% of the time, surpassing delayed-time DUACS (78%), delayed-time NeurOST (83%), and Mercator (70%). For 7-day forecasts, we achieve 70% correct angle predictions, surpassing persistence forecasts of DUACS (62%) and NeurOST (65%) and numerical forecasts of the operational OGCM Mercator (56%). We also achieve superior performance on our metrics of correct magnitude and mean MEVA. The superior performance of ORCAst over delayed-time (DUACS

TABLE 8. Comparison of ORCAst performances on the extratropical global area, using nadir altimetry or SWOT and nadir altimetry as inputs, in addition to SST. Including SWOT data as inputs does not bring improvements to ORCAst currents. This model was trained as usual during S1 and then trained with SWOT inputs during S2. It is evaluated on August–December 2023, rather than the whole of 2023, due to the availability of SWOT data; thus, the first line of results is not identical to the line corresponding to S2 in Table 3.

	Correct angle %		Correct magnitude %		MEVA (cm s^{-1})	
	$T + 1$	$T + 7$	$T + 1$	$T + 7$	$T + 1$	$T + 7$
ORCAst	83	68	73	69	25	30
ORCAst + SWOT	82	67	72	68	26	31

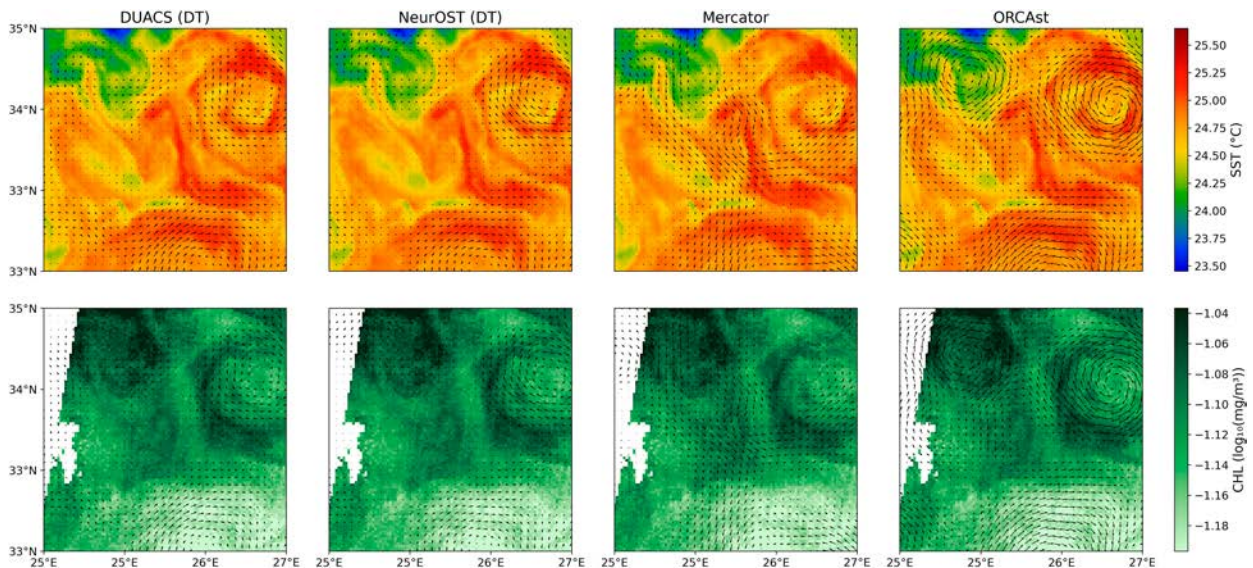


FIG. 9. Model performance in the Mediterranean Sea on 28 Jun 2023. (top) The corresponding SST images in the background and (bottom) the corresponding CHL images in the background. The field of ocean surface currents for each model is represented using black arrows. ORCAst correctly localizes both large- and fine-scale structures, visible in the SST and CHL signatures. We note that for this example, we illustrate a higher-resolution DUACS product, only available for the Mediterranean Sea ([Copernicus Marine Service 2024](#)).

and NeurOST) and near-real-time (Mercator) methods highlights the potential of deep learning in leveraging large quantities of observational data to produce accurate operational ocean current forecasts.

We show that progressively training the network on sparser, but more direct, observations of currents leads to improved performance at each training stage. We train using nadir-pointing altimetry and DUACS geostrophic currents, followed by SWOT

SSH and SWOT geostrophic currents, and finally with drifter current measurements. In particular, training on drifters allows the model to learn from observations beyond geostrophic ocean surface currents.

Regional models consistently outperform the globally trained model. The Mediterranean Sea, with its unique dynamics and fine spatial structures, showed the largest improvement when using a regional model. Our regional models

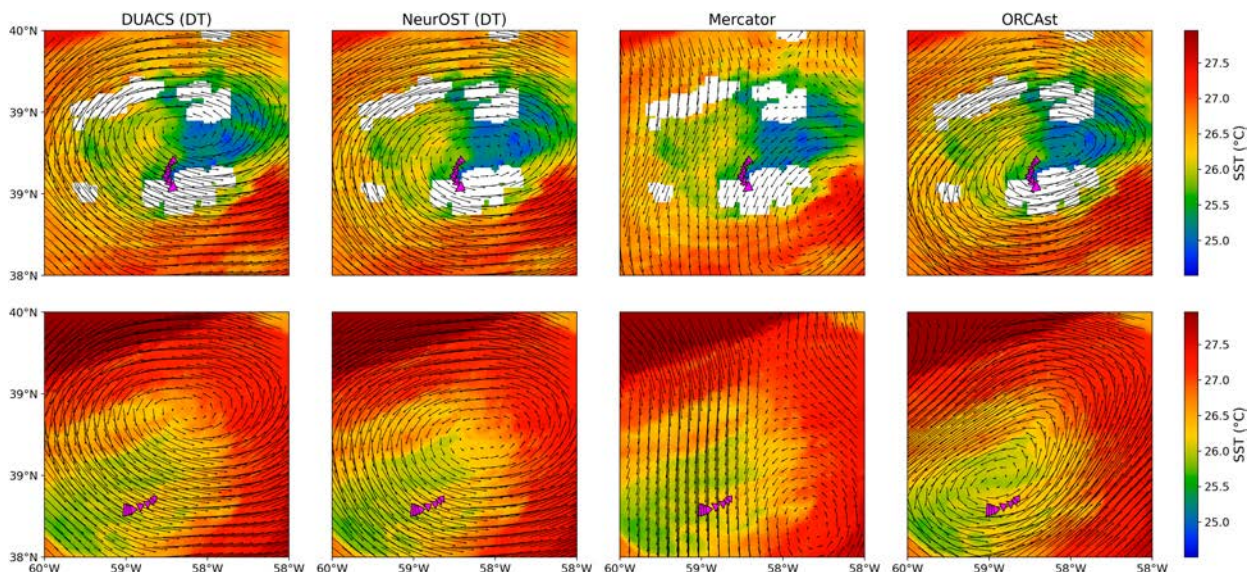


FIG. 10. Drifter (in magenta) with WMO identification number 4402878 on 12 and 18 Aug 2023. The image backgrounds show the corresponding SST, with the field of predicted ocean surface currents represented using black arrows. We compare the model performance to in situ observations from drifters at different lead times. ORCAst correctly localizes the eddy and predicts its evolution over 7 days.

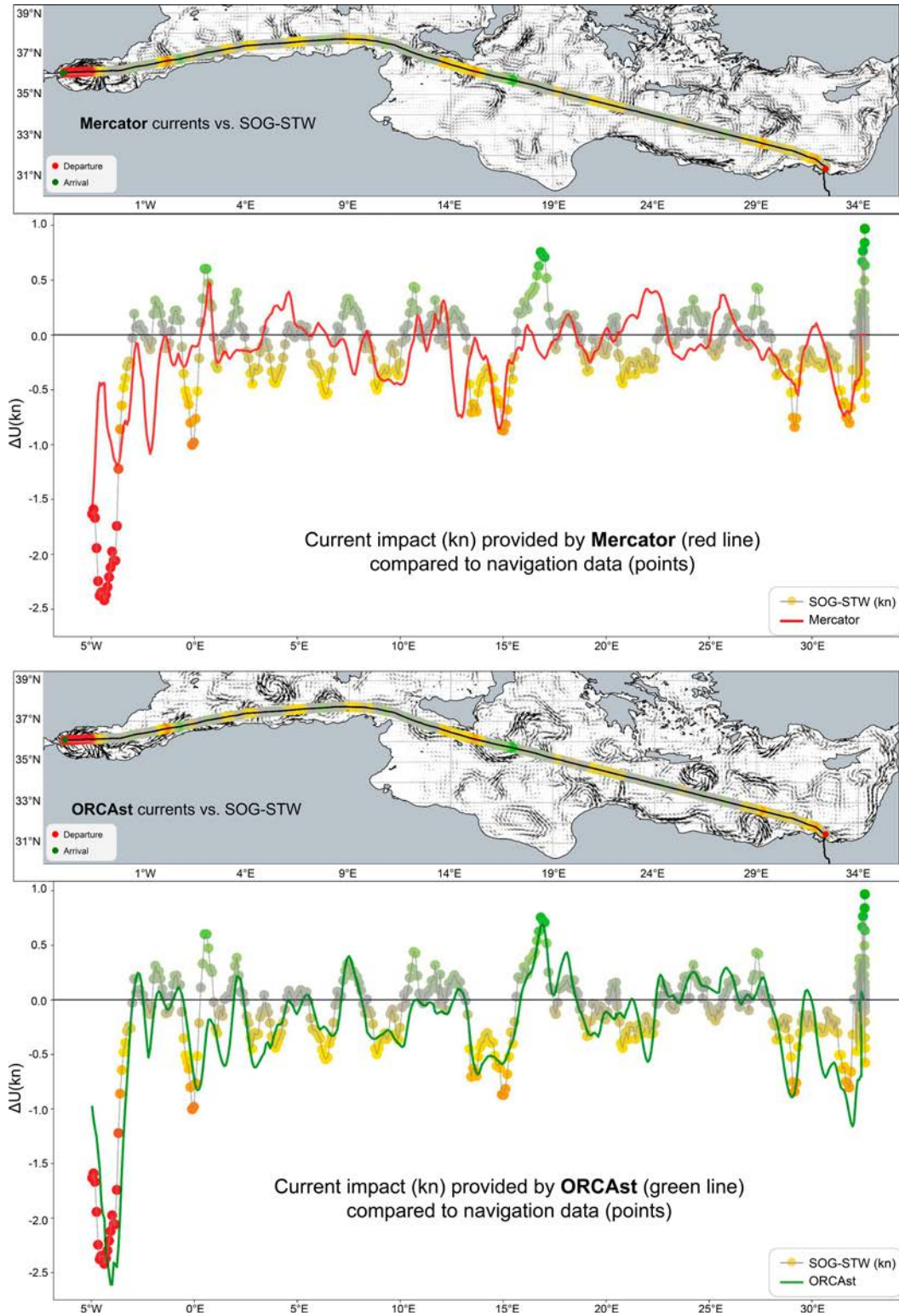


FIG. 11. Evaluation of Mercator and ORCAst on ship data for a trans-Mediterranean voyage between 21 and 26 Aug 2023. Maps show ocean currents estimated by Mercator and ORCAst, with the ship route colored proportionally to SOG-STW measurements. The graphs show SOG-STW measurements, compared with the current impact predicted by Mercator and ORCAst. The SOG-STW curve is colored proportionally to SOG-STW, in the same way as in the map. Currents predicted by ORCAst more closely follow the SOG-STW measurements than currents predicted by Mercator. See Table 9 for a quantitative evaluation of this voyage.

TABLE 9. Correlation and R^2 scores between the SOG–STW and the current impact measured by DUACS, NeurOST, Mercator, and ORCAst for the trans-Mediterranean voyage illustrated in Figure 11. For DUACS and NeurOST, we consider DT current estimations for 21 Aug 2023. For Mercator and ORCAst, we consider ocean current forecasts for 21–26 Aug 2023, made using data available at or before 20 Aug 2023.

	Pearson’s correlation (r)	R^2 score
DUACS (DT pers.)	0.73	0.50
NeurOST (DT pers.)	0.74	0.53
Mercator	0.31	−0.01
ORCAst	0.79	0.54

also outperform the global model in energetic regions such as the Gulf Stream and Agulhas regions, better capturing localized physical dynamics over longer lead times.

The inclusion of chlorophyll- a data did not yield substantial improvements when SST data were already used, suggesting limited complementary information from chlorophyll a for forecasting surface currents. Similarly, incorporating SWOT data as inputs did not significantly enhance model performance, likely due to the limited availability of SWOT observations (1 year of data). The addition of SWOT measures over the coming years should further enhance ORCAst’s learning and performance. However, as there is little overlap between the SWOT measurements in 2023 and the drifter trajectories used for validation, we cannot demonstrate yet the effectiveness of using SWOT data as inputs. We note that training our model using SWOT data as targets led to significant improvements for surface currents at meso- and submesoscale.

The qualitative assessments reinforce the quantitative metrics, demonstrating that ORCAst can provide detailed forecasts that align closely with observed patterns, outperforming baseline methods. Moreover, the evaluation against real ship data highlights the practical applicability of the model. Accurate current predictions are essential for optimizing ship routing and reducing emissions, making ORCAst a valuable tool for the maritime industry.

b. Perspectives

Throughout this study, we demonstrate the potential of fusing multivariate observations with deep learning. However, a number of issues still need to be addressed to improve the quality of our forecasted ocean currents.

First, ORCAst produces no forecast between latitudes -20° and 20° . This is because the geostrophic approximation is not valid near the equator, and the link between currents and altimetry weakens. One alternative approach would be to train a deep learning model using an assimilated numerical model as a target on the equator.

Second, to further develop the fusion of observations from various sources, we are interested in using ship data. For instance, the automatic identification system (AIS) provides information about the ship’s steering and speed over ground, which can help us deduce the surface water velocity. However, using AIS data to obtain indirect current measurements

is difficult and calls for additional studies. Other sources of onboard observations exist, such as acoustic Doppler current profiler (ADCP) observations or Wavex sensor data. These more direct measurements of currents along the ship’s route can also be used as targets for evaluation or training; however, they may also be noisy, often measuring incoherent current velocities.

Finally, from a methodological point of view, we are interested in going beyond regression models (like ORCAst). Regression models estimate the mean of a distribution, and their predictions are therefore smoothed, leading to a domain gap compared to real physical samples, with smoothing getting stronger when conditions are scarce or irrelevant. On the other hand, generative deep learning allows the network to generate multiple examples from the same set of inputs, hopefully forming a physics-matching distribution. In particular, denoising diffusion probabilistic models (DDPMs) and flow matching techniques have led to an impressive image generation breakthrough in the past years. Using these sampling models, we might be able to estimate the full conditional distribution of the currents, instead of only the mean forecast.

Acknowledgments. This research study was supported by a CIFRE thesis grant from the ANRT 2023/1590 and partially financed by the iLab2023 Grant DOS0220247/00 of the French Public Bank of Investment (BPIFrance). This work was granted access to the HPC resources of IDRIS under the allocation AD011015571 made by GENCI. The authors of this paper have affiliations with Amphitrite, a company developing maritime solutions such as optimal ship routing, using methods similar to those discussed in this study. As such, there exist potential conflicts of interest regarding the presentation and interpretation of the results. The authors declare that the study was conducted with rigor and integrity, and the interpretation of the results is based on the evidence presented in the data. Large language models (LLMs) were used to assist with coding and writing; however, the authors have made every effort to ensure the content meets ethical and academic standards.

Data availability statement. The original satellite data and in situ data from drifters presented in the study are openly available. We refer to the data section of this article for citations of all of our data sources. We unfortunately are unable to diffuse the ship data used in our analysis; however, these data are not necessary for the reproduction of our model and key results. The core model architecture, SimVP, is publicly available at <https://github.com/chengtan9907/OpenSTL>. This architecture has been applied in a similar context in the NeurOST model, whose code is also publicly accessible at <https://github.com/smartin98/NeurOST>. The specific modifications we made to this base model are detailed in section 3a of the manuscript. Due to constraints related to commercial applications, we are unable to release our full implementation. However, we provide comprehensive technical details to enable reproduction of our model and training strategy by adapting the aforementioned publicly available code.

APPENDIX A

Learned Positional Embeddings

Learned positional embeddings provide insights into the spatiotemporal conditions most relevant for ORCAst to forecast ocean currents. These embeddings are represented as 32-dimensional vectors. To visualize patterns within the embeddings, we create a 3D latitude–longitude–time grid encompassing all spatiotemporal positions used during inference. To efficiently cluster points in this large grid, we apply the minibatch k -means clustering algorithm with $k = 30$ clusters and a batch size of 8. Figure A1 plots these $k = 30$ clusters on a latitude–longitude grid for a chosen temporal dimension (first week of the year). Latitude is a strong determining factor of cluster membership, demonstrating that the positional embedding learns the strong influence of latitude in predicting ocean currents.

Furthermore, Fig. A2 shows the correlation and the distance between clusters. We see a clear separation between Southern Hemisphere (clusters 0–18) and Northern Hemisphere (clusters 19–29) and Southern Hemisphere embeddings. The Southern Hemisphere is more represented in the clustering (19 clusters versus 11 clusters) because the ocean-to-continent ratio is higher in the south, which means that the model sees more south examples and adapts its encoding accordingly. Generally, Southern Hemisphere clusters are correlated with each other and negatively correlated with Northern Hemisphere clusters. Southern Hemisphere clusters are closest to other Southern Hemisphere clusters, and Northern Hemisphere clusters are closest to other Northern Hemisphere clusters.

This is expected, as the Coriolis effect varies with latitude, reaching its maximum at the poles. In the Northern Hemisphere, the westerlies drive currents eastward, while the

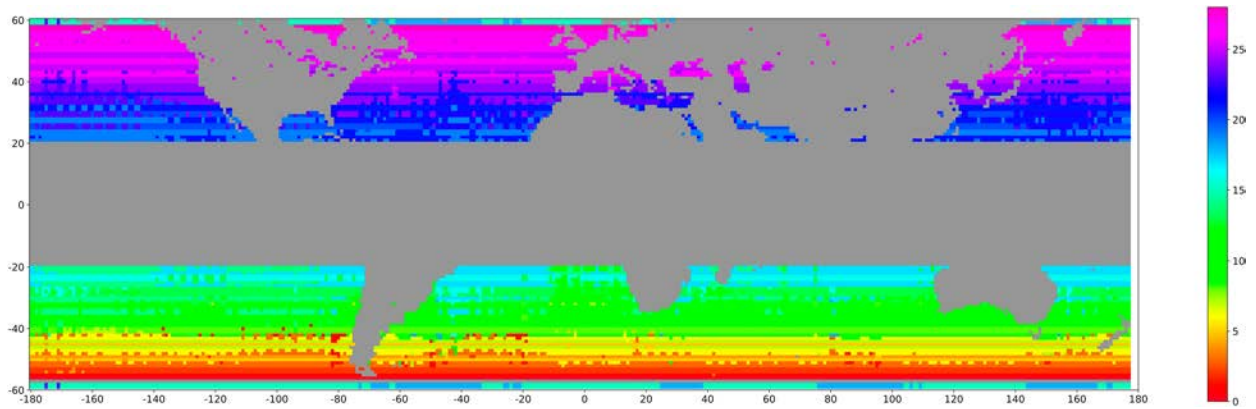


FIG. A1. Clustering of positional embeddings using k means, where $k = 30$. Each color represents a cluster learned by our model. We reorder the numbering and colors of the groups for clearer visualization purposes. Land areas and the equator are in gray.

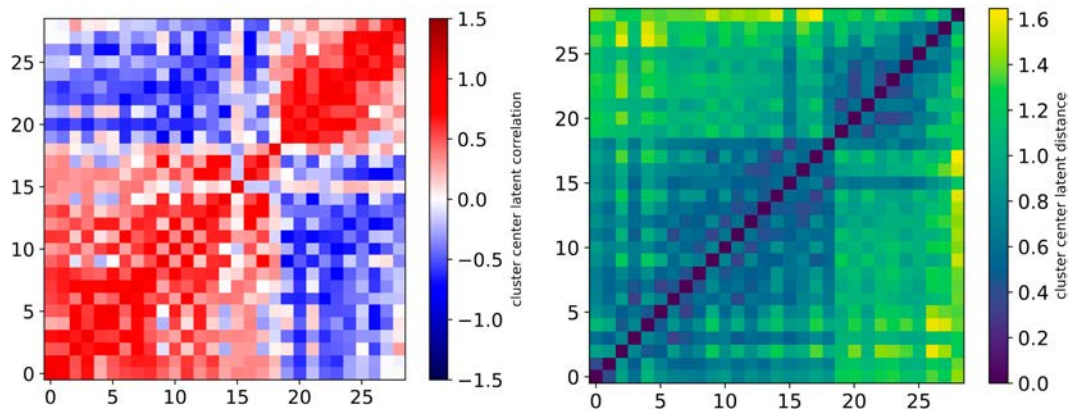


FIG. A2. (left) The correlation and (right) the distance between clusters of our positional embedding, chosen using the k -means method with $k = 30$. Geographic zones corresponding to cluster numbers are shown in Fig. A1. Southern Hemisphere clusters are numbered 0–18, and Northern Hemisphere clusters are numbered 19–29.

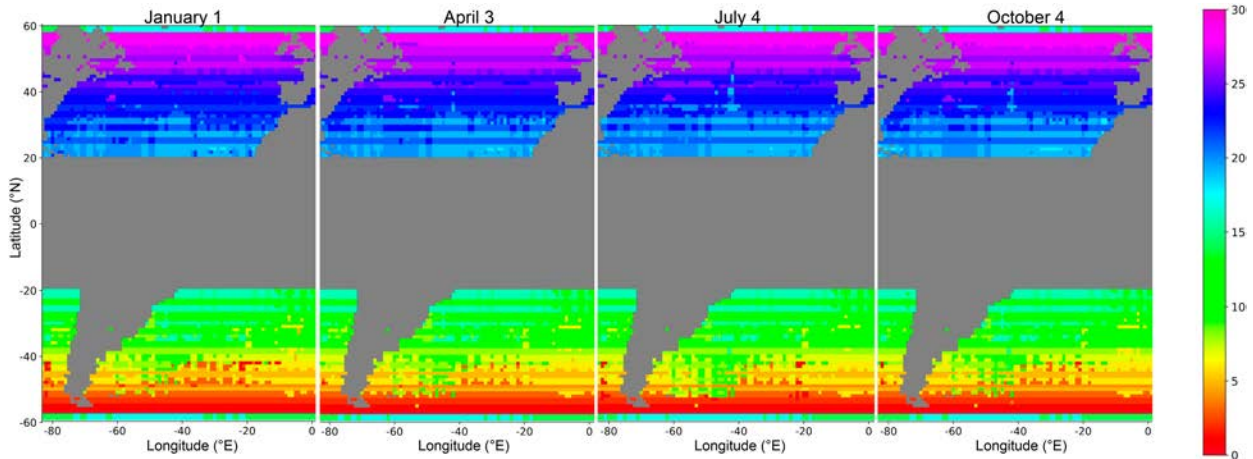


FIG. A3. Clustering of positional embeddings for longitudes 80°W–0° at four different seasons using k means, where $k = 30$. Each color represents a cluster learned by our model. We reorder the numbering and colors of the groups for clearer visualization purposes. Land areas and the equator are in gray.

Coriolis effect deflects them to the right, forming clockwise gyres. In the Southern Hemisphere, a similar mechanism creates counterclockwise gyres due to the Coriolis effect deflecting currents to the left. The variability in the Coriolis effect modulates the intensity and structure of ocean currents in both hemispheres.

Although we denormalize our dataset at a spatial resolution of 2° and at a temporal resolution of 1 week, our positional encoding of the day of the year theoretically allows the model to learn seasonal differences, which may occur at a finer scale, such as the speed of the evolution of the eddies over the 7-day forecast or the size of the eddies. Figure A3 plots $k = 30$ clusters at longitudes 80°W–0° at dates across four seasons. There is minor seasonal variation. In the Northern Hemisphere, some clusters corresponding to higher latitudes in July are the same as clusters corresponding to lower latitudes in January. In the Southern Hemisphere, some clusters corresponding to higher latitudes in July are the same as clusters corresponding to lower latitudes in January.

APPENDIX B

Bias of SWOT

We calculate the bias between absolute dynamic topography (ADT) measured by SWOT and by nadir altimetry at the same location and time, as well as the standard deviation of their differences (Fig. B1). We find that SWOT's KaRIn instrument overestimates SSH compared to nadir altimetry by 5.26 ± 3.32 cm. This bias, which likely arises from differences in the mean dynamic topography (MDT), could complicate training, as our model uses nadir SSH in stage 1 and SWOT SSH in stage 2. While correcting for this offset could help a model aiming to predict ADT, we do not apply such a correction here, as our focus is on predicting ocean surface currents, which should not be significantly affected by an MDT-related bias. Instead, to address potential distributional shifts between data sources, we adopt a three-stage training approach, using a different ground-truth dataset at each stage. By first training on nadir and then transitioning to SWOT, the model progressively learns to account for such biases and generalize across observation types.

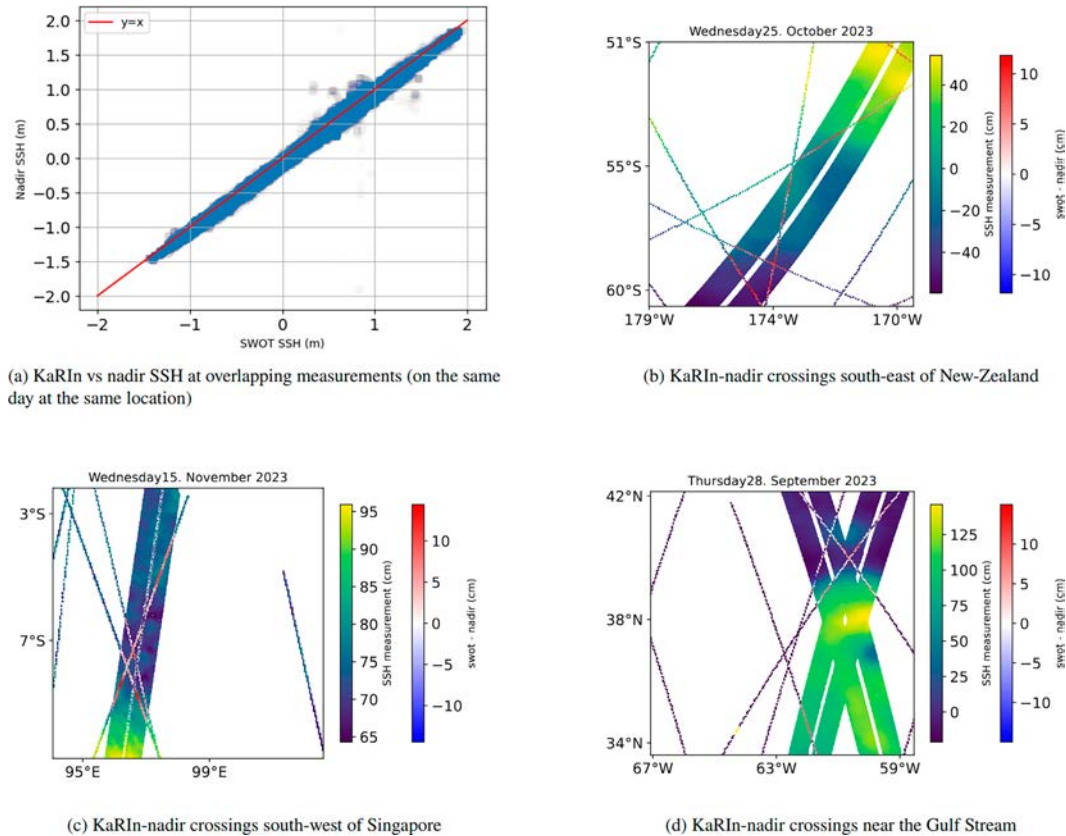


FIG. B1. We compare KaRIn and nadir altimetry across intersection points of both measures of SSH. We find that there is an underestimation of SSH from nadir compared to SSH measured by SWOT's KaRIn instrument, of around 5.26 ± 3.32 cm.

APPENDIX C

Ablation Study

NeurOST performs better than DUACS, as shown in Table 2; however, we train using DUACS as targets in

stage 1 of our learning procedure. As shown in Table C1, replacing DUACS with NeurOST improves model performance during stage 1 of training. However, the difference in performance diminishes in stages 2 and 3.

TABLE C1. Comparison of model performance using DUACS and NeurOST as targets during S1 of training in the Gulf Stream region. Replacing DUACS with NeurOST improves performance in S1 of training. However, the performance difference diminishes in S2 and S3, where the models achieve similar results. See section 4d for a description of our metrics. Values are rounded. We evaluate on drifter observations from 2023 with magnitude $> 0.25 \text{ m s}^{-1}$. Best results are shown in bold font.

	Correct angle %		Correct magnitude %		MEVA (cm s^{-1})	
	$T + 1$	$T + 7$	$T + 1$	$T + 7$	$T + 1$	$T + 7$
NeurOST (DT pers.)	86	63	72	65	25	37
S1	82	67	67	65	29	33
S1 on NeurOST	87	72	73	68	26	32
S2	86	72	76	72	25	32
S2, NeurOST for S1	87	72	76	71	25	31
S3	88	73	77	70	24	31
S3, NeurOST for S1	89	73	79	71	23	30

APPENDIX D

Eddy Kinetic Energy

Eddy kinetic energy (EKE) is commonly used to quantify the intensity of mesoscale variability, such as ocean eddies and turbulent motions. It represents the kinetic energy associated with deviations from the mean flow and is particularly useful for comparing the dynamical activity of ocean models or observational datasets. For a given point, EKE is defined as $EKE = (1/2)(u'^2 + v'^2)$, where u' and v' are the anomalies of the zonal and meridional velocities, respectively, relative to the mean flow. We compute the EKE using model outputs from 2023, with mean flows from 2022 to 2024. Figure D1 shows the EKE of DUACS, NeurOST, Mercator, and ORCAst in our three regional domains. ORCAst tends to have a higher EKE than other models.

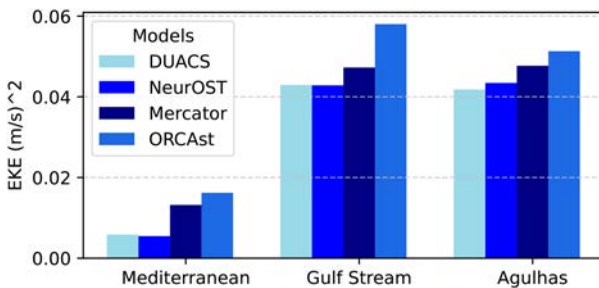


FIG. D1. EKE for our three zones of interest.

REFERENCES

- Aouni, A. E., Q. Gaudel, C. Regnier, S. Van Gennip, M. Drevillon, Y. Drillet, and J.-M. Lellouche, 2024: Glonet: Mercator's end-to-end neural forecasting system. *arXiv*, 2412.05454v2, <https://doi.org/10.48550/arXiv.2412.05454>.
- Archambault, T., A. Filoche, A. Charantonis, and D. Béréziat, 2023: Multimodal unsupervised spatio-temporal interpolation of satellite ocean altimetry maps. *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023)*, Vol. 4, Science and Technology Publications, 159–167, <https://www.scitepress.org/PublishedPapers/2023/116201/116201.pdf>.
- , —, —, and —, 2024a: Pre-training and fine-tuning attention based encoder decoder improves sea surface height multi-variate inpainting. *VISAPP 2024—19th Int. Conf. on Computer Vision Theory and Applications*, VISIGRAPP, Roma, Italy, 1–10, <https://hal.sorbonne-universite.fr/hal-04475205>.
- , —, A. A. Charantonis, D. Béréziat, and S. Thiria, 2024b: Learning sea surface height interpolation from multi-variate simulated satellite observations. *J. Adv. Model. Earth Syst.*, **16**, e2023MS004047, <https://doi.org/10.1029/2023ms004047>.
- Callahan, J. L., K. Maeda, and S. L. Brunton, 2019: Robust flow reconstruction from limited measurements via sparse representation. *Phys. Rev. Fluids*, **4**, 103907, <https://doi.org/10.1103/PhysRevFluids.4.103907>.
- Chen, L., X. Zhong, F. Zhang, Y. Cheng, Y. Xu, Y. Qi, and H. Li, 2023: FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate Atmos. Sci.*, **6**, 190, <https://doi.org/10.1038/s41612-023-00512-1>.
- Ciani, D., C. Fanelli, and B. Buongiorno Nardelli, 2024: Estimating ocean currents from the joint reconstruction of absolute dynamic topography and sea surface temperature through deep learning algorithms. *Ocean Sci.*, **21**, 199–216, <https://doi.org/10.5194/os-21-199-2025>.
- Copernicus Climate Service, 2024: Global ocean gridded L4 sea surface heights and derived variables reprocessed 1993 ongoing. Satellite Observations, NetCDF-4 Format, accessed 1 February 2024, <https://doi.org/10.48670/moi-00148>.
- Copernicus Marine Service, 2023: Global ocean along track L3 sea surface heights NRT. Satellite Observations, NetCDF-4 Format, accessed 1 February 2024, <https://doi.org/10.48670/moi-00147>.
- , 2024: European seas gridded L4 sea surface heights and derived variables NRT. Satellite Observations, NetCDF-4 Format, accessed 1 February 2024, <https://doi.org/10.48670/moi-00142>.
- Drévilion, M., and Coauthors, 2008: The GODAE/Mercator-ocean global ocean forecasting system: Results, applications and prospects. *J. Oper. Oceanogr.*, **1**, 51–57, <https://doi.org/10.1080/1755876X.2008.11020095>.
- Elipot, S., R. Lumpkin, R. C. Perez, J. M. Lilly, J. J. Early, and A. M. Sykulski, 2016: A global surface drifter data set at hourly resolution. *J. Geophys. Res. Oceans*, **121**, 2937–2966, <https://doi.org/10.1002/2016JC011716>.
- , A. Sykulski, R. Lumpkin, L. Centurioni, and M. Pazos, 2022: Hourly location, current velocity, and temperature collected from global drifter program drifters world-wide. Accessed 31 July 2023, <https://doi.org/10.25921/x46c-3620>.
- Emery, W., J. Brown, and Z. Nowak, 1989: AVHRR image navigation: Summary and review. *Photogramm. Eng. Remote Sens.*, **55**, 1175–1183.
- EUMETSAT for Copernicus, 2024: Ocean colour service: Product data format specification—OLCI level 1 & level 2 instrument products. Platform: Sentinel-3; sensor: OLCI (optical); data level: Level 2; collection ID: EO:EUM:DAT:0407; parameters: Ocean, ocean colour, level 2 data, accessed 1 February 2024, <https://data.eumetsat.int/product/EO:EUM:DAT:0407>.
- Fablet, R., B. Chapron, J. Le Sommer, and F. Sévellec, 2024: Inversion of sea surface currents from satellite-derived SST-SSH synergies with 4DVarNets. *J. Adv. Model. Earth Syst.*, **16**, e2023MS003609, <https://doi.org/10.1029/2023MS003609>.
- Filoche, A., T. Archambault, A. Charantonis, and D. Béréziat, 2022: Statistics-free interpolation of ocean observations with deep spatio-temporal prior. *ECML/PKDD Workshop on Machine Learning for Earth Observation and Prediction (MACLEAN)*, Grenoble, France, ESA, 9–19, <https://hal.sorbonne-universite.fr/hal-03765735v1>.
- Gao, Z., C. Tan, L. Wu, and S. Z. Li, 2022: Simvp: Simpler yet better video prediction. *2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, IEEE Computer Society, 3160–3170, <https://doi.org/10.1109/CVPR52688.2022.00317>.
- Gurvan, M., and Coauthors, 2017: NEMO ocean engine. Zenodo, 412 pp., <https://zenodo.org/records/3248739>.
- Ifremer, 2024: Global ocean-in-situ near-real-time observations. In-Situ observations, NetCDF-4 format, accessed 1 February 2024, <https://doi.org/10.48670/moi-00036>.
- Jonasson, O., A. Ignatov, and G. Irina, 2024: NOAA/STAR: GHRSSST NOAA/STAR ACSPO v2.81 0.02 degree L3S daily dataset from LEO satellites. NOAA/STAR, accessed

- 1 February 2024, https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:GHRSSST-L3S_LEO_PM-STAR.
- Kingma, D. P., and J. Ba, 2015: Adam: A method for stochastic optimization. *Proc. Third Int. Conf. on Learning Representations (ICLR)*, San Diego, CA, ICLR, <https://doi.org/10.48550/arXiv.1412.6980>.
- Kugusheva, A., H. Bull, E. Moschos, A. Ioannou, B. Le Vu, and A. Stegner, 2024: Ocean satellite data fusion for high-resolution surface current maps. *Remote Sens.*, **16**, 1182, <https://doi.org/10.3390/rs16071182>.
- Lam, R., and Coauthors, 2023: Learning skillful medium-range global weather forecasting. *Science*, **382**, 1416–1421, <https://doi.org/10.1126/science.adi2336>.
- Le Guillou, F., L. Gaultier, M. Ballarotta, S. Metref, C. Ubelmann, E. Cosme, and M.-H. Rio, 2023a: Regional mapping of energetic short mesoscale ocean dynamics from altimetry: Performances from real observations. *Ocean Sci.*, **19**, 1517–1527, <https://doi.org/10.5194/os-19-1517-2023>.
- , B. Maxime, S. Metref, C. Ubelman, E. Cosme, R. M.-H., 2023b: The WOC data challenges. ESA, accessed 1 February 2024, <https://2024-dc-woc-esa.readthedocs.io/en/latest/index.html>.
- Martin, S. A., G. E. Manucharyan, and P. Klein, 2023: Synthesizing sea surface temperature and satellite altimetry observations using deep learning improves the accuracy and resolution of gridded sea surface height anomalies. *J. Adv. Model. Earth Syst.*, **15**, e2022MS003589, <https://doi.org/10.1029/2022MS003589>.
- , —, and —, 2024: Deep learning improves global satellite observations of ocean eddy dynamics. *Geophys. Res. Lett.*, **51**, e2024GL110059, <https://doi.org/10.1029/2024GL110059>.
- Mercator Océan International, 2024: Global ocean physics analysis and forecast. Copernicus Marine Service, accessed 1 February 2024, <https://doi.org/10.48670/moi-00016>.
- MET Norway, 2023: Mediterranean Sea—High resolution and ultra high resolution L3S sea surface temperature. CMEMS, accessed 1 February 2024, <https://doi.org/10.48670/moi-00171>.
- Morrow, R., D. Blurmstein, and G. Dibarbour, 2018: Fine-scale altimetry and the future SWOT mission. *New Frontiers in Operational Oceanography*, GODAE Ocean View, 191–226.
- NASA/JPL and CNES, 2024: The SWOT_L3_LR_SSH product, derived from the L2 SWOT KaRIn low rate ocean data products. CNES, accessed 1 February 2024, <https://doi.org/10.24400/527896/A01-2023.018>.
- NOAA/STAR, 2024: GHRSSST NOAA/STAR ACSPO v2.81 0.02 degree L3S daily dataset from LEO satellites. PO.DAAC, accessed 1 February 2024, https://podaac.jpl.nasa.gov/dataset/L3S_LEO_DY-STAR-v2.81.
- O'Reilly, J., S. Maritorena, B. Mitchell, D. Siegel, K. Carder, S. Garver, M. Kahru, and C. McClain, 1998: Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.*, **103**, 24 937–24 953, <https://doi.org/10.1029/98JC02160>.
- Scott, A. M., E. M. Georgy, and K. Patrice, 2024: NeuroST Level 4 sea surface height and surface geostrophic currents analysis product (version 2024.0). Daily mapped neural network product using Level 3 altimetry observations and MUR Level 4 SST, PO.DAAC, accessed 1 February 2024, <https://doi.org/10.5067/NEURO-STV24>.
- Taburet, G., A. Sanchez-Roman, M. Ballarotta, M.-I. Pujol, J.-F. Legeais, F. Fournier, Y. Faugere, and G. Dibarbour, 2019: DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Sci.*, **15**, 1207–1224, <https://doi.org/10.5194/os-15-1207-2019>.
- Tapley, B. D., G. H. Born, and M. E. Parke, 1982: The SEASAT altimeter data and its accuracy assessment. *J. Geophys. Res.*, **87**, 3179–3188, <https://doi.org/10.1029/JC087iC05p03179>.
- Thiria, S., C. Sorrow, T. Archambault, A. Charantonis, D. Berezat, C. Mejia, J.-M. Molines, and M. Crépon, 2023: Downscaling of ocean fields by fusion of heterogeneous observations using deep learning algorithms. *Ocean Modell.*, **182**, 102174, <https://doi.org/10.1016/j.ocemod.2023.102174>.
- Tonani, M., and Coauthors, 2015: Status and future of global and regional ocean prediction systems. *J. Oper. Oceanogr.*, **8**, s201–s220, <https://doi.org/10.1080/1755876X.2015.1049892>.
- Ubelmann, C., L. Carrere, C. Durand, G. Dibarbour, Y. Faugère, M. Ballarotta, F. Briol, and F. Lyard, 2022: Simultaneous estimation of ocean mesoscale and coherent internal tide sea surface height signatures from the global altimetry record. *Ocean Sci.*, **18**, 469–481, <https://doi.org/10.5194/os-18-469-2022>.
- Wang, W., and Coauthors, 2022: PVT v2: Improved baselines with pyramid vision transformer. *Comput. Visual Media*, **8**, 415–424, <https://doi.org/10.1007/s41095-022-0274-8>.
- Wang, X., and Coauthors, 2024: XiHe: A data-driven model for global ocean eddy-resolving forecasting. arXiv, 2402.02995v4, <https://doi.org/10.48550/arXiv.2402.02995>.
- Wu, Y., and K. He, 2020: Group normalization. *Int. J. Comput. Vision*, **128**, 742–755, <https://doi.org/10.1007/s11263-019-01198-w>.
- Yu, W., M. Luo, P. Zhou, C. Si, Y. Zhou, X. Wang, J. Feng, and S. Yan, 2022: Metaformer is actually what you need for vision. 2022 *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, Institute of Electrical and Electronics Engineers, 10 809–10 819, <https://doi.org/10.1109/CVPR52688.2022.01055>.