Learning Sea Surface Height Interpolation from Multi-variate Simulated Satellite Observations

Théo Archambault, Arthur Filoche, Anastase Alexandre Charantonis, Dominique Béréziat, Sylvie Thiria

HAL Id: hal-04568666

https://hal.sorbonne-universite.fr/hal-04568666

Submitted on 5 May 2024

Learning Sea Surface Height Interpolation from Multi-variate Simulated Satellite Observations.

Théo Archambault[1,2], Arthur Filoche[3], Anastase Charantonis[2,4,5], Dominique Béréziat[1], Sylvie Thiria[2]

[1]Sorbonne Université, CNRS, LIP6, Paris, France
[2]Sorbonne Université, CNRS, IRD, MNHN, LOCEAN, Paris, France
[3]University of Western Australia, Perth, Australia
[4]ENSIIE, CNRS, LaMME, Evry, France
[5]Inria, Paris, France

Key Points:

- We developed a realistic simulation of satellite observations of sea surface height and temperature
- We compare deep learning supervised and unsupervised strategies to interpolate the sea surface height
- We present a hybrid learning strategy using supervised training on our simulated data and unsupervised fine-tuning on real-world data

Corresponding author: Théo Archambault, `theo.archambault@lip6.fr`

Abstract

Satellite-based remote sensing missions have revolutionized our understanding of the Ocean state and dynamics. Among them, space-borne altimetry provides valuable Sea Surface Height (SSH) measurements, used to estimate surface geostrophic currents. Due to the sensor technology employed, important gaps occur in SSH observations. Complete SSH maps are produced using linear Optimal Interpolations (OI) such as the widely-used Data Unification and Altimeter Combination System (DUACS). On the other hand, Sea Surface Temperature (SST) products have much higher data coverage and SST is physically linked to geostrophic currents through advection. We propose a new multi-variate Observing System Simulation Experiment (OSSE) emulating 20 years of SSH and SST satellite observations. We train an Attention-Based Encoder-Decoder deep learning network (ABED) on this data, comparing two settings: one with access to ground truth during training and one without. On our OSSE, we compare ABED reconstructions when trained using either supervised or unsupervised loss functions, with or without SST information. We evaluate the SSH interpolations in terms of eddy detection. We also introduce a new way to transfer the learning from simulation to observations: supervised pre-training on our OSSE followed by unsupervised fine-tuning on satellite data. Based on real SSH observations from the Ocean Data Challenge 2021, we find that this learning strategy, combined with the use of SST, decreases the root mean squared error by 24% compared to OI.

Plain Language Summary

The surface of the ocean is observed through various sensors embedded in satellites. Specifically, the height of the sea surface is a crucial variable as it can be used to estimate surface currents. It is currently measured through satellite altimeters, but the data acquisition process leaves gaps in their observations. Providing fully gridded maps of the sea surface height is thus an important interpolation problem. The widely used interpolated product has some troubles, especially when dealing with small and rapidly evolving eddies. To enhance its quality, we propose an artificial neural network, a trainable method able to estimate complete sea surface height images. The flexibility of these methods allows us to use different satellite information, such as the sea surface temperature, which has a better resolution. Usually, neural networks are trained on a dataset upon which they learn the link between input and output data. However, in a realistic geoscience scenario, the output is never known for sure. We propose a dataset that simulates this problem and explores methodologies to train these methods. Finally, we design a new way of learning SSH reconstruction using both simulated data and real satellite observation.

1 Introduction

1.1 Background

Since the first ocean remote sensing missions in the 1970s, satellite observation has become one of the most determining contributions to understanding ocean state and dynamics (S. Martin, 2014). Through the years, satellites have provided a huge amount of data of various physical natures with wide spatial coverage that complemented in situ datasets. Among these techniques, satellite altimetry is used to retrieve the Sea Surface Height (SSH) a determining variable of the ocean circulation. The SSH spatial gradient can be used to estimate geostrophic circulation, i.e. the currents issued from the equilibrium between the Coriolis force and the pressure force in the surface layer of the Ocean. SSH (also called Absolute Dynamical Topography by the altimetry community) is currently measured by nadir-pointing altimeters, meaning that they can only take measurements vertically, along their ground tracks, by calculating the return time of a radar pulse. This leads to large gaps in the observed SSH, and providing a gap-free product (L4) is

a challenging Spatio-Temporal interpolation problem. One of the most widely used L4 products in oceanography applications is provided by the Data Unification and Altimeter Combination System (DUACS) (Taburet et al., 2019). It is a linear Optimal Interpolation (OI) of the nadir along-track measurements leveraging a covariance matrix tuned on 25 years of data. However, several studies show that DUACS OI misses some of the mesoscales structures and eddies (Amores et al., 2018; Stegner et al., 2021). Improving the reconstruction of gridded altimetry products remains an open challenge.

To enhance the quality of the SSH reconstruction and sea surface current estimation, using additional physical information such as the Sea Surface Temperature (SST) has been demonstrated to be beneficial (Ciani et al., 2020; Thiria et al., 2023; S. A. Martin et al., 2023; Archambault et al., 2023; Fablet et al., 2023). SST motion is linked to ocean circulation (Isern-Fontanet et al., 2006), and therefore to SSH, as currents transport heat in an advection dynamics. SST measurements obtained through passive infrared technology offer a remarkably high spatial resolution, ranging from 1.1 to 4.4 km (Emery et al., 1989), even if intermittent clouds introduce data gaps. On the other hand, microwave sensors provide lower-resolution SST data (25 km) which can be obtained through non-raining clouds. Infrared and microwave data are then combined with in situ measurements, to produce fully gridded SST maps (Donlon et al., 2012; Chin et al., 2017). Thus, a crucial challenge lies in developing efficient reconstruction methods capable of fusing data derived from different remote sensing techniques, each presenting distinct interpolation challenges. This is essential to unlock the full potential of satellite oceanography products.

1.2 SSH interpolation with deep neural networks

In the last decade, deep learning has emerged as one of the leading methods to address image inverse problems. Neural networks demonstrated remarkable flexibility in fusing observations from various sources and modalities, exhibiting their capacity to learn complex relationships given enough training samples (McCann et al., 2017; Ongie et al., 2020). Prior works proved that it is possible to use SST to enhance SSH reconstruction with a deep-learning network, whether from a downscaling perspective (Nardelli et al., 2022; Thiria et al., 2023) or an interpolation one (Fablet et al., 2023; S. A. Martin et al., 2023). However, training neural networks usually requires fully gridded ground truth, which is unavailable in a realistic geoscience scenario. To overcome this limitation, it is possible to design a twin experiment of the satellite observing system on a numerical simulation, also called an Observing System Simulation Experiment (OSSE). Neural networks can then be trained on simulated data and applied to satellite observations. The Ocean Data Challenge 2020 (CLS/MEOM, 2020) is a 1-year OSSE providing SSH simulated observations and ground truth, aiming to compare various reconstruction methods. Among them, Fablet et al. (2021) performed a supervised deep learning of the SSH interpolation and extended their study using SST showing increased performance (Fablet et al., 2023). However, if the SSH-only network was successfully applied to real data, adapting its SST-using version is still a challenging problem. Another way to overcome the lack of ground truth is to employ loss functions allowing the neural network to learn from observations alone. Archambault et al. (2023); S. A. Martin et al. (2023) trained a neural network using only SST and SSH observations showing the potential of unsupervised learning for SSH interpolation. This last option has the advantage of not suffering from the domain gap between the simulation and the real data, but we expect unsupervised interpolations to produce less accurate reconstructions.

1.3 Contributions

First, as the previously existing Ocean Data Challenge OSSE provided only one year of data without SST, it presents clear limitations to train neural networks. We pro-

pose a new OSSE that includes 20 years of SSH and SST data, with realistic simulated observations of these variables.

Second, we compare a fixed neural architecture trained in a supervised and unsupervised way, with or without SST. The SSH interpolation is learned by an Attention-Based Encoder-Decoder (ABED) on our OSSE. Its assessment involves evaluating errors in SSH and geostrophic currents reconstruction. Additionally, a comparison of the eddy structures is conducted, both quantitatively and visually.

Third, we propose a hybrid learning strategy consisting of supervised pre-training on our OSSE and unsupervised fine-tuning on real-world observations. Specifically, we compare the same network architecture, trained in the three following manners: supervised on our OSSE and directly applied to observations, trained directly on observations, and the proposed hybrid approach.

This paper is structured as follows. In Section 2, after giving a rationale for leveraging SST information in the interpolation method, we detail our OSSE. In Section 3 we present our architecture and the training loss functions. In Section 4 we evaluate the interpolation on our OSSE, in terms of SSH reconstruction and geostrophic circulation errors. We also perform an eddy detection to demonstrate that SST-using methods retrieve more realistic ocean structures, and we compare ourselves to existing state-of-the-art methods on the Ocean Data Challenge 2020 OSSE. Finally, we compare the learning strategies on real observations. In Section 5, we discuss the limitations and perspectives of this work.

2 Multi-variate data

In the following, we provide a rationale for the SSH and SST relationship, outline the reference data source we utilized (Global Ocean physics Reanalysis (CMEMS, 2020)), and detail our OSSE's SSH and SST observations. We also present the satellite observations that will be used for training and fine-tuning.

2.1 Physical relationship between SSH and SST

One of the most important uses of SSH data is to recover oceanic currents through geostrophic approximation. It consists of supposing a static equilibrium between the surface projection of the Coriolis force and the resultant pressure forces. Far from the Equator, where the Coriolis force projection is null, it is a good approximation of the circulation. The surface geostrophic currents can be computed from the SSH h following Equation 1

$$\mathbf{w}_{\text{geo}} = \begin{pmatrix} u_{\text{geo}} \\ v_{\text{geo}} \end{pmatrix} = \begin{pmatrix} -\dfrac{g}{f}\dfrac{\partial h}{\partial y} \\ \dfrac{g}{f}\dfrac{\partial h}{\partial x} \end{pmatrix} \tag{1}$$

where u_{geo} and v_{geo} are the Eastward and Northward geostrophic currents, x and y the Eastward and Northward coordinates and where $f = 2\Omega_r \sin(\phi)$ is the Coriolis factor, Ω_r being the Earth the rotation period, ϕ the latitude and g the gravitational acceleration.

In a first approximation, the surface temperature T can be considered as a passive tracer transported by surface currents. The evolution of a scalar in a velocity field is described by the linear advection given in Equation 2.

$$\frac{\partial T}{\partial t} + \mathbf{w}.\nabla T = 0 \tag{2}$$

Combining the geostrophic and the advection Equations (1,2), we understand why a time series of SST observations should provide pertinent information for constraining the SSH

reconstruction. Several studies pointed out the interest in using SST to reconstruct SSH as Isern-Fontanet et al. (2006); González-Haro et al. (2020), which established spectral relations between SSH and SST in a Surface Quasi Geostrophic framework. However, the physical link between temperature and SSH is more complex, as other phenomena must be considered, such as diffusion, convection, circulation between water depths, atmosphere interactions, and viscosity. Satellite observations of temperature and sea surface height also suffer from instrumental errors and are, by nature, limited to observing the ocean surface. This is why neural network architectures, thanks to their flexibility, seem appropriate to learn the complex underlying link between the data.

2.2 Observing System Simulation Experiment

To effectively replicate the relationship between the two variables, we propose an Observing System Simulation Experiment (OSSE), meaning a twin experiment that accurately models the satellite observations of the Ocean. This approach is widely used in the geosciences community as it provides a way to test reconstruction methods and errors (Gaultier et al., 2016; Amores et al., 2018; Stegner et al., 2021). With this mindset, SSH and SST variables of a high-resolution simulation are considered as the ground truth ocean state upon which we simulate satellite measurements. The coherence of the relation between SSH and SST is ensured by the physical model, while with our OSSE we produce enough pairs of ground truth/observation to train a neural network.

In this paper, we denote $\mathbf{X}^{ssh}$ and $\mathbf{X}^{sst}$ the ground truth fields of the SSH and SST and $\mathbf{Y}^{ssh}$ and $\mathbf{Y}^{sst}$, the simulated observations. Hereafter, we detail the reference dataset of our OSSE and the observation operators of the two variables.

2.2.1 Base simulation

We conduct our experiments on the Global Ocean Physics Reanalysis product (GLORYS12) (CMEMS, 2020). It provides various physical data such as SSH, SST, and oceanic currents with a spatial resolution of $1/12°$ (around 8 km). GLORYS12 is based on the NEMO 3.6 model (Madec et al., 2017) and assimilates satellite observations (SSH along-track observations and SST full domain observations) through a reduced-order Kalman filter. It is updated annually by the Copernicus European Marine Service, making it impossible to use in near real-time applications. We select a temporal subset of this simulation from 2000/03/20/ to 2019/12/29, for a total of 7194 days.

We select a portion of the Gulf Stream, between 33° to 43° North and -65° to -55° East. This area is known for its intense circulation, its water mass of very different temperatures, and is far enough from the equator that the geostrophic approximation can be applied. Comparing the surface circulation of the model with its geostrophic approximation, we find that an RMSE of 6.6 cm/s for u_{geo} and 6.1 6.1 cm/s for v_{geo}. Considering the high intensity and variations of the currents in the Gulf Stream (with 37.1 and 34.3 cm/s of standard deviation for u and v respectively), geostrophy seems to be an adequate estimation. Thus, we expect a significant synergy between SSH and SST which a neural network can learn. For computational reasons, we resample the data to images of size 128×128 with a bilinear interpolation, corresponding to a resolution of $0.078°$ by pixel (approximately 8.7 km). Doing so, the perceptive field of the network covers the entire 10° by 10° area.

2.2.2 SSH simulated observations

The nadir-pointing altimetry satellites take approximately a measurement per second, along their ground tracks. Their observations are a series of values with precise spatio-temporal coordinates that we aim to simulate. To do so, we retrieve the support of real-world satellite observations denoted $\mathbf{\Omega} = \{\Omega_i = (t_i, lat_i, lon_i), i \in [0:N]\}$ from the Coper-

nicus sea level product (CMEMS, 2021). Using $\mathbf{\Omega}$ and the ground truth data $\mathbf{X}^{ssh}$ we simulate SSH observations $\mathbf{Y}^{ssh}$ as the trilinear interpolation of the simulated field on each point of the support. We add an instrumental error $\varepsilon \sim \mathcal{N}(0, \sigma)$ with $\sigma = 1.9\,\mathrm{cm}$, which is the distribution used in the Ocean data challenge 2020 (CLS/MEOM, 2020). The SSH observations $\mathbf{Y}^{ssh}$ is defined as following:

$$\mathbf{Y}^{ssh} = \mathcal{H}^{ssh}\left(\mathbf{X}^{ssh}, \mathbf{\Omega}\right) + \varepsilon \tag{3}$$

where $\mathcal{H}^{ssh}$ is the trilinear interpolation operator of the ground truth $\mathbf{X}^{ssh}$ on the support $\mathbf{\Omega}$. An example of these simulated along-track measurements is presented on the first row of Figure 1. For the neural network input observations, we regrid these data to a daily 128×128 image. We set the pixel value with no simulated satellite observation to zero, and we average the daily measurements of SSH inside each pixel to represent the mean of the daily data from the different satellites (if any). As GLORYS12 reanalysis assimilates along-track SSH data, selecting satellite measurements at the same location as the assimilated data might introduce a bias in our observations. To overcome this issue, we desynchronize the real satellite ground tracks from the one we use to produce SSH observations by introducing a time delay (772 days) between the real L3 satellite observations and the simulation. It ensures that simulated along-track data is selected randomly, rather than specifically where the model assimilated real-world observations.
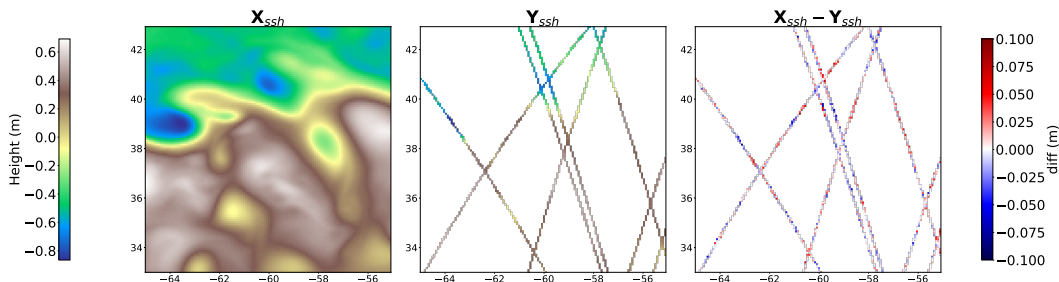

Figure 1. Images of the ground truth SSH from GLORYS12, the simulated along-track measurements, and the difference.

2.2.3 SST simulated observations

SST remote sensing is based on direct infrared imaging, leading to wider measurement swaths but making the data sensitive to cloud cover. The so-called L3 satellite products have much higher data coverage, but no observation is possible when clouds are present. To fill the gaps, the L3 products from several satellites are merged and interpolated to form the fully gridded image, using complementary microwave satellite sensors (which produce lower resolution data but are less sensitive to clouds), and in situ measurements (Donlon et al., 2012; Chin et al., 2017). This results in various resolutions in the same product, where high-resolution structures are artificially smoothed when the cloud cover (C) is too thick.

We simulate the SST observation operator $\mathcal{H}^{sst}$ as follows:

$$\mathbf{Y}^{sst} = \mathcal{H}^{sst}\left(\mathbf{X}^{sst}, C\right) = (1 - C) \odot \left(\mathbf{X}^{sst} + \varepsilon\right) + C \odot \mathcal{G}_{\sigma_t, \sigma_x} \star \left(\mathbf{X}^{sst} + \varepsilon\right) \tag{4}$$

where $\odot$ is the element-wise product, $\star$ the convolution product, and ε is a white Gaussian noise image of size 32×32 linearly upsampled to a 128×128 image. We also use

a spatio-temporal Gaussian filter, $\mathcal{G}_{\sigma_t,\sigma_x}$ with $\sigma_t = 1.23$ days and $\sigma_x \approx 16$(km) to simulate the smoothing of the interpolation performed by satellite products. To compute a realistic cloud cover C, we use 2 years of data from an NRT L3 product (CMEMS, 2023), which we periodically replicate to match the length of our dataset. We then linearly interpolate the cloud cover to our spatial resolution, and perform an average filter with a kernel size approximately equal to 43 (km). This step is essential, as applying a binary mask results in patches at the frontiers between cloud-free and cloudy regions. Our SST observations thus present a spatially and temporally correlated noise, with different resolutions depending on cloud cover. In the end, $\mathcal{H}^{sst}$ adds a noise with RMSE of $0.48\,^{\circ}\text{C}$ where the SST standard deviation of the ground truth is $4.96\,^{\circ}\text{C}$, which we present in Figure 2. This observation operator is different from real-world degradations but produces an image with an in-equal noise resolution similar to the errors present in the L4 SST products. Also, as SST presents strong annual variations that should be removed, we deseasonalize it. For each SST image, we subtract the mean image calculated for the corresponding day across the dataset. This is known to improve machine learning timeseries prediction (Ahmed et al., 2010), and in our case, it produces better reconstructions as shown in Appendix 6.3.
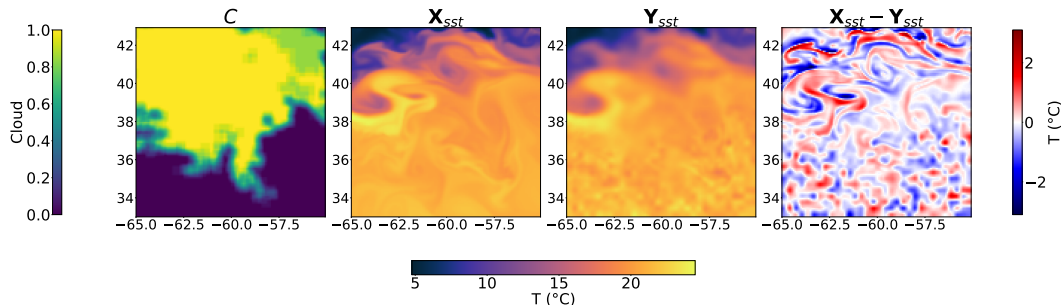

Figure 2. Images of our cloud cover, the ground truth SST from GLORYS12, the noised SST, and the difference.

2.3 Satellite observations

To constitute a dataset of real-world observations, we propose the L3 SSH product that we used to recover realistic satellite ground tracks (CMEMS, 2021). These data are the inputs used in the DUACS optimal interpolation process and are available from the years 1993 to 2023. For the L4 SST product, we use the Multiscale Ultrahigh Resolution (MUR) SST (NASA/JPL, 2019). MUR SST is produced through an optimal interpolation of infrared, microwave, and in situ measurements (Chin et al., 2017). Its resolution is very high (0.01°), so we linearly interpolate the data to our resolution (0.078°), and are available from 2002/05/31 to the present. We select satellite observations from 2002/06/01 to 2022/02/09 for a total of 7194 days which is the same number of timesteps that our OSSE. We also select the same geographical area between 33° to 43° North and -65° to -55° East. The two data are presented in Figure 3.

3 Proposed interpolation method

3.1 Learning the interpolation

The observation operator $\mathcal{H}^{ssh}$ previously described can be seen as a forward operator that we aim to inverse. In the past years, deep neural networks, especially convolutional neural networks, have proven their ability to solve ill-posed image inverse problems (McCann et al., 2017) and more specifically inpainting problems (Jam et al., 2021;
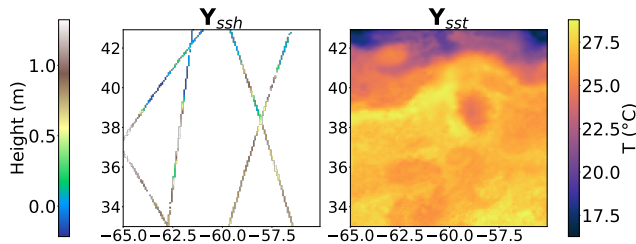
Figure 3. Images of satellite observations of the SSH and the SST, respectively.

Qin et al., 2021). A neural network f_θ is trained on a database to estimate the true state from observations $f_\theta(y) = \hat{x}$. Learning this inversion operator thus requires (y, x) pairs (supervised) or only y (unsupervised) (Ongie et al., 2020).

We chose to perform the interpolation on a temporal window of 21 days; the input is thus a tensor of 21 images of SSH, with or without SST images, and the output is the 21 corresponding days of SSH only. The neural network estimates the true state from observations, $\hat{\mathbf{X}}^{ssh} = f_\theta(\mathbf{Y})$, where $\mathbf{Y} = \mathbf{Y}^{ssh}$ for a SSH-only interpolation, and $\mathbf{Y} = (\mathbf{Y}^{ssh}, \mathbf{Y}^{sst})$ if the network uses SST. The length of the time window is discussed in Section 4.1, and training losses of the network in Section 3.3.

3.2 Architecture

Convolutional neural networks, one of the most used deep learning methods in image tasks, learn convolution operations able to identify features over space and/or time. These networks have been used for multiple tasks in geosciences, from forecasting (Che et al., 2022) to interpolation (Manucharyan et al., 2020; Fablet et al., 2021; S. A. Martin et al., 2023; Archambault et al., 2023), and from eddy detection (Moschos et al., 2020) to super-resolution (Nardelli et al., 2022; Thiria et al., 2023), to name a few. Over time, the machine learning community introduced various ways to organize these convolution operations, each one presenting distinct advantages. Residual layers learn small modifications between their input and output, making neural networks easier to train (He et al., 2016). Attention layers ponder their inputs by a factor between zero and one. This allows subsequent layers to focus on important features while neglecting irrelevant ones, which makes it well-suited to extracting information from contextual variables. It is widely used in many computer vision tasks (Guo et al., 2021) and can be transposed to geoscience applications such as (Che et al., 2022). An encoder-decoder architecture progressively compresses and decompresses the input data, identifying structures at different resolutions.

In this study, we compare different learning techniques on a fixed architecture: an attention-based encoder-decoder (ABED) presented in Figure 4. This neural network benefits from the layers described above. The overall structure of our neural network is inspired by Che et al. (2022), who introduced a residual U-Net with attention layers for rain nowcasting. We removed U-Net residual connections that were not suited for the interpolation task and changed the attention and the upsampling blocks. The encoder starts with a batch normalization and a 3D convolution (in time and space) followed by two downsampling blocks that divide spatial dimensions by 2 (see Figure 4). The decoder is composed of residual attention blocks followed by upsampling blocks.

Hereafter, we describe our attention block, which consists of two essential steps: temporal and spatial attention modules. Our approach builds upon the Convolutional Block Attention Module (CBAM) principle introduced by Woo et al. (2018), which successively performs channel and spatial attention. We extend this idea by incorporating
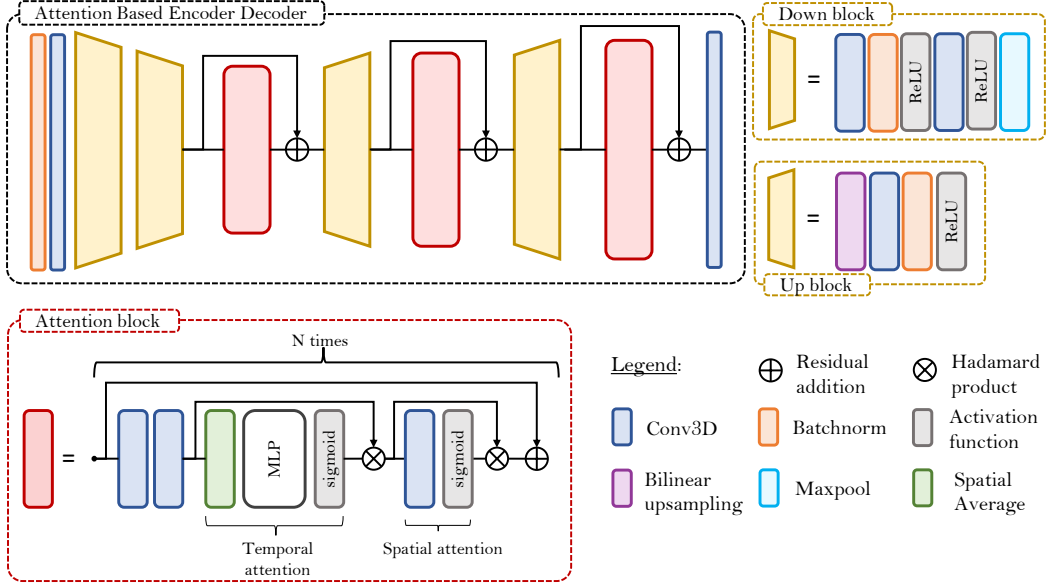
Figure 4. The architecture of the proposed Attention-Based Encoder Decoder (ABED) neural network. It is designed to take a time series of 21 images of SSH, with or without a time series of SST. The encoder divides the spatial dimensions of the images by 4 through 2 "down-block". Then, a 3D attention layer block is used to highlight relevant information in the images, followed by a residual connection. Finally, a decoding block upsamples the images, and attention and decoding blocks are repeated to get back to the initial image size.

temporal information in the channel attention mechanism. To do so, we first compute the spatial average of each channel and instant, resulting in a tensor of size $C \times T$ where C is the channel number and T is the time series length. Subsequently, we apply two one-dimensional convolutional layers with a kernel of size 1, followed by a sigmoid activation function to estimate the attention weights. This corresponds to a 2-layer perceptron shared by every time step, which is different from the CBAM, as it includes the temporal information in the channel attention. These weights are then multiplied to each timestep of every channel, enabling the network to highlight salient features and suppress irrelevant information. After performing temporal attention, we proceed with spatial attention. This step involves utilizing a 3-dimensional convolutional operation, where the kernel size's temporal length matches the time series's length. As a result, the entire time series is aggregated into a single 2D image, which serves as the basis for deriving spatial attention. A residual skip connection is then applied, and the described block is repeated 4, 2, and 1 time for the first, second, and last block, respectively. For further details about our implementation, we provide the PyTorch implementation of our network in `https://gitlab.lip6.fr/archambault/james2024`.

3.3 Loss and regularization

We propose to compare two main strategies to train the neural network. Thanks to the OSSE previously described, we have access to the ground truth, which we can use to learn the interpolation in a classic supervised fashion. However, it is also possible to train directly on observations by applying the $\mathcal{H}^{ssh}$ on the generated map $\hat{\mathbf{X}}^{ssh}$ before computing the loss (see Equations 5,6,7). Filoche et al. (2022) performed the interpolation with SSH observations only, and, using the same principle, Archambault et al. (2023)

showed that it was possible to estimate SSH images starting from SST only and constraining on SSH observations. Both these methods are fitted on one (or a small number) example and must be refitted to be applied to unseen data. Using a larger real-world satellite dataset, S. A. Martin et al. (2023) trained a neural network directly from observations by constraining it on independent satellite observations that were not given in the input. However, the lack of ground truth reference makes it harder to compare the different reconstructions, especially regarding detected eddies and structures. We propose to train neural networks using the 3 following losses:

- The MSE using ground truth:

$$\mathcal{L}_{\text{sup}}(\mathbf{X}^{ssh}, \hat{\mathbf{X}}^{ssh}) = \frac{1}{T \times H \times W} \sum_{t,x,y} \left(\mathbf{X}^{ssh}_{t,x,y} - \hat{\mathbf{X}}^{ssh}_{t,x,y} \right)^2 \tag{5}$$

- The MSE using only observations:

$$\mathcal{L}_{\text{unsup}}(\mathbf{Y}^{ssh}, \hat{\mathbf{X}}^{ssh}) = \frac{1}{N} \sum_{i} \left(\mathbf{Y}^{ssh}_i - \mathcal{H}^{ssh}(\hat{\mathbf{X}}^{ssh})_i \right)^2 \tag{6}$$

- The MSE using only observations and the regularization introduced by S. A. Martin et al. (2023):

$$\mathcal{L}_{\text{unsup\_reg}}(\mathbf{Y}^{ssh}, \hat{\mathbf{X}}^{ssh}) = \mathcal{L}_{\text{unsup}}(\mathbf{Y}^{ssh}, \hat{\mathbf{X}}^{ssh}) + \lambda_1 \frac{1}{N_1} \sum_{i} \left(\frac{\partial}{\partial s} \mathbf{Y}^{ssh}_i - \frac{\partial}{\partial s} \mathcal{H}^{ssh}(\hat{\mathbf{X}}^{ssh})_i \right)^2$$
$$+ \lambda_2 \frac{1}{N_2} \sum_{i} \left(\frac{\partial^2}{\partial s^2} \mathbf{Y}^{ssh}_i - \frac{\partial^2}{\partial s^2} \mathcal{H}^{ssh}(\hat{\mathbf{X}}^{ssh})_i \right)^2$$

$$\tag{7}$$

where $\frac{\partial}{\partial s}$ is the along-track derivation of the SSH approximated by its rate of change (see Appendix 6.1). T is the temporal length of the time series (here 21), H and W the spatial dimensions of the images (here both equals 128), and N, N_1, N_2, the number of satellite measurements of SSH, SSH first, and SSH second spatial derivative along satellite tracks, respectively. We take $\lambda_1 = \lambda_2 = 0.05$ the regularization coefficients, the same values used by S. A. Martin et al. (2023).

The losses $\mathcal{L}_{\text{unsup}}$ and $\mathcal{L}_{\text{unsup\_reg}}$ apply the observation operator $\mathcal{H}^{ssh}$, before computing the MSE, which allows the training in a framework where only observations are available. Thus, from an interpolation point of view, the inversion methods that use these losses are unsupervised as they can be trained without any ground truth image. However, if we constrain the network on the same observations that were given in input, an over-fitting of along tracks will occur with no guarantee of generalization. To avoid this problem, S. A. Martin et al. (2023) constrained their network on the observations of one satellite that were withdrawn from the input. Similarly, we remove the data of one satellite from the inputs but we calculate the loss function on all satellite observations (the ones given and the ones left aside). In doing so, the network must generalize outside the along-track measurement that was given as input. In Figure 5, we call $\mathbf{Y}^{ssh}_{\text{in}}$ the input observations and present an unsupervised inversion computational graph.

3.4 Training details

Train, validation, test split. We partitioned the OSSE dataset into three subsets: training, validation, and test data. We used the year 2017 exclusively to test our reconstruc-
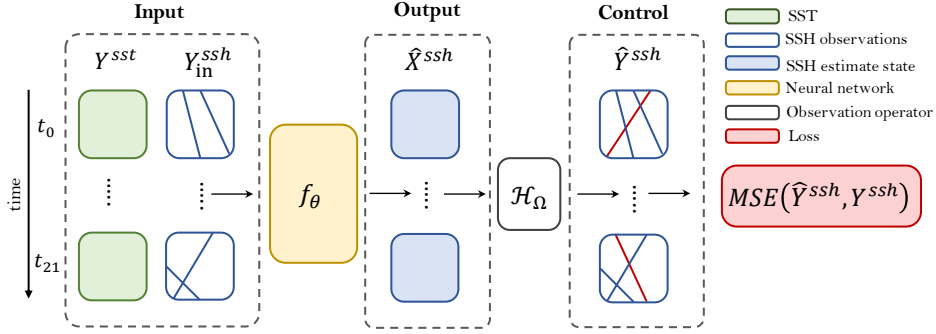
Figure 5. Computational graph of the proposed unsupervised interpolation method. The neural network input is a 21-day time series of SSH satellite observations, excluding data from a single satellite, and optionally includes SST measurements. The network estimates a time series of SSH field states, upon which the observation operator is subsequently applied in order to deduce $\hat{\mathbf{Y}}^{ssh}$. Finally, the Mean Squared Error between the $\hat{\mathbf{Y}}^{ssh}$ and $\mathbf{Y}^{ssh}$ is used to control the network.

tions (every analysis conducted in the following was performed on this data). We validate our methods on three distinct time intervals: (1) from 2002/07/14 to 2003/07/28, (2) from 2008/01/05 to 2009/01/18, and (3) from 2013/06/28 to 2014/07/13. The remaining data was used for training, but leaving a 15-day period to prevent data leakage.

Normalization. We normalize the artificial network's input and output by subtracting the mean and dividing by the standard deviation. The normalization parameters are computed only on the neural network inputs, SST, or along-track data. Specifically, we first perform this normalization for images related to SSH along-track measurements and subsequently replace any missing values with zeros. We normalize the neural network SSH outputs with the statistics computed on the input observations (so that the method remains applicable in an unsupervised setting). When training with the regularized loss of Equation 7, we also normalize the data from the first and second SSH along-tracks derivative.

Training hyperparameters. We train every method using an ADAM optimizer (Kingma & Ba, 2017) with a learning rate starting at 5.10^{-5} and a decay of 0.99. We perform an early stopping with a patience of 8 epochs. For the supervised training, the stopping criteria is the RMSE of the reconstruction on the fully gridded domain on the validation data, but in the unsupervised setting, we compute this RMSE on left-aside along-track measurements. Doing so, the stopping strategy is still compliant with a situation where no ground truth is accessible.

Ensemble. As neural network optimization is sensitive to its weight initialization, we train 3 networks for every setting. The so-called "Ensemble" estimation is the average SSH map of the 3 networks. An ensemble estimation helps stabilize performances and enhances the reconstruction (Hinton & Dean, 2015). In the following, we call "Ensemble score" the score of the previously mentioned ensemble estimation and "Mean score" the average of the score of each network taken independently.

4 Results

In Sections 4.1 and 4.2, we compare the different training methods on our OSSE to highlight the drawbacks of unsupervised learning and the advantages of SST. In Sec-

tion 4.3, we assess the similarity of our OSSE and the previously existing one, the Ocean Data Challenge 2020. In Section 4.4, we build upon the conclusions drawn in previous sections to present a transfer learning method from our OSSE to observations.

4.1 SSH reconstruction and quality of derived geostrophic currents

We compare the fields estimated by the networks trained using the 3 losses $\mathcal{L}_{\text{sup}}$, $\mathcal{L}_{\text{unsup}}$ and $\mathcal{L}_{\text{unsup\_reg}}$, with 3 different sets of input data: only SSH tracks, SSH and the noised SST (denoted nSST), and noise-free SST (denoted SST). The noise-free SST provides an upper-bound performance of the neural network in the case of a perfect physical link between SSH and SST. We give the RMSE of the SSH estimates fields on the test set in Table 1, and the RMSE on the velocity fields in Table 2. Systematically, the ensemble reconstruction has a lower RMSE than the mean performance, which is usual in machine learning, as individual member errors are compensated by others. Comparing the ensemble scores, we observe that the supervised loss function outperforms the unsupervised framework in every data scenario. Specifically, in the SSH+SST scenario, the supervised loss decreases the ensemble RMSE of $\mathcal{L}_{\text{unsup}}$ by 17%, and 9% without SST. Also, adding SST as an additional input to the network generally improves performance compared to using SSH alone. This improvement is observed across all three loss functions, as the error values decrease for SSH+nSST compared to SSH. For instance, the SSH-only ensemble RMSE is decreased by 31% and 20% for SST and nSST, respectively, with $\mathcal{L}_{\text{sup}}$. The regularization introduced by S. A. Martin et al. (2023) slightly increases reconstruction but is still close to the unregularized inversion.

Loss	SSH	SSH+nSST	SSH+SST
$\mathcal{L}_{\text{sup}}$	4.16 — 3.81	3.34 — 3.03	**2.97 — 2.63**
$\mathcal{L}_{\text{unsup}}$	4.56 — 4.20	3.84 — 3.49	3.56 — 3.16
$\mathcal{L}_{\text{unsup\_reg}}$	4.33 — 4.07	3.76 — 3.52	3.48 — 3.20

Table 1. SSH reconstruction RMSE in centimeters (mean score on the left and ensemble score on the right) of 3 ABED networks. The interpolation is trained using the 3 different losses described in Section 3.3 with the following settings: SSH-only interpolation, SSH and noised SST, and SSH and noise-free SST. All metrics are given on the central image of a 21-day time window.

We estimate the surface currents from the reconstructed SSH from Equation 1, and we compare it to the surface circulation of the model. The errors on velocity in Table 2 follow the same patterns as the RMSE on the SSH fields but with lesser differences between methods. The RMSE is not too far from the minimal error achievable through geostrophy, which is $6.57\,^{\text{cm}}/_{\text{s}}$ for u and 6.14 for v on this data.

In Figure 6, we show the daily errors of the different methods on the test year. We notice a strong temporal variability of the RMSE, with a notable increase in late summer. Specifically, in August and September, all methods are performing worse than in Winter, which can be explained by the high kinetic energy of the ocean in summer (Zhai et al., 2008; Kang et al., 2016).

An important challenge of ocean satellite products is to provide real-time estimations, as many applications cannot use products available with too much time delay. In an operational framework, products that are immediately available are called Near Real Time (NRT) whereas those that require a time delay before release are called Delayed Time (DT). While in Table 1 we presented the results obtained on the central image of the time window, we can also display their scores along the 21-day temporal window as in Figure 7. The central image is a 10-day Delayed Time reconstruction as we need im-

Loss	SSH		SSH+nSST		SSH+SST	
	u	v	u	v	u	v
$\mathcal{L}_{\text{sup}}$	12.8	13.9	11.1	12.0	**10.1**	**10.7**
$\mathcal{L}_{\text{unsup}}$	13.4	15.5	12.0	14.1	11.1	13.1
$\mathcal{L}_{\text{unsup\_reg}}$	12.8	14.3	11.7	12.9	11.0	12.0

Table 2. Eastward (u) and Northward (v) surface currents in cm/s. The currents were estimated by applying the geostrophic approximation (see Equation1) on the SSH ensemble estimation of the 3 ABED networks.
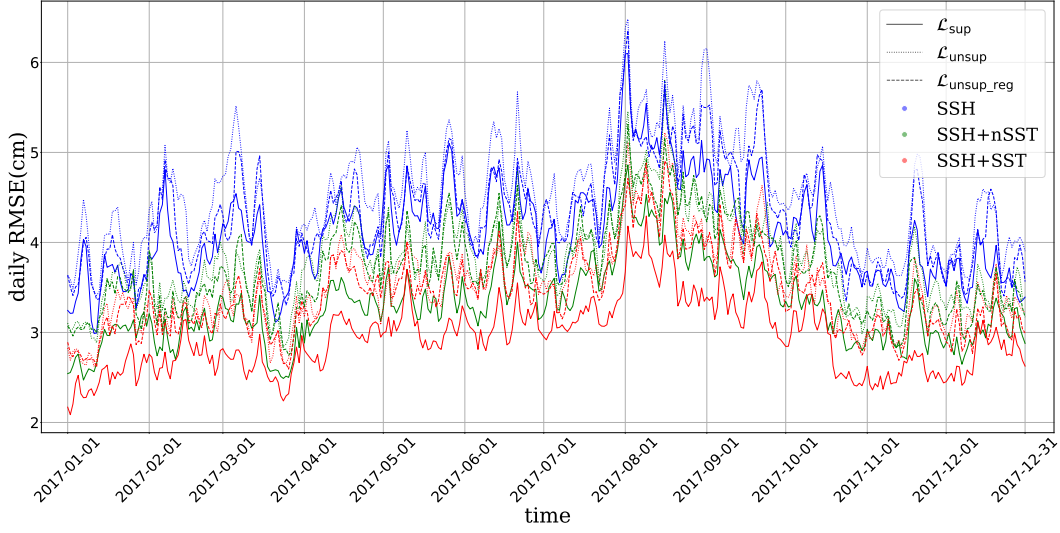
Figure 6. RMSE of the different reconstructions during the test year (2017).
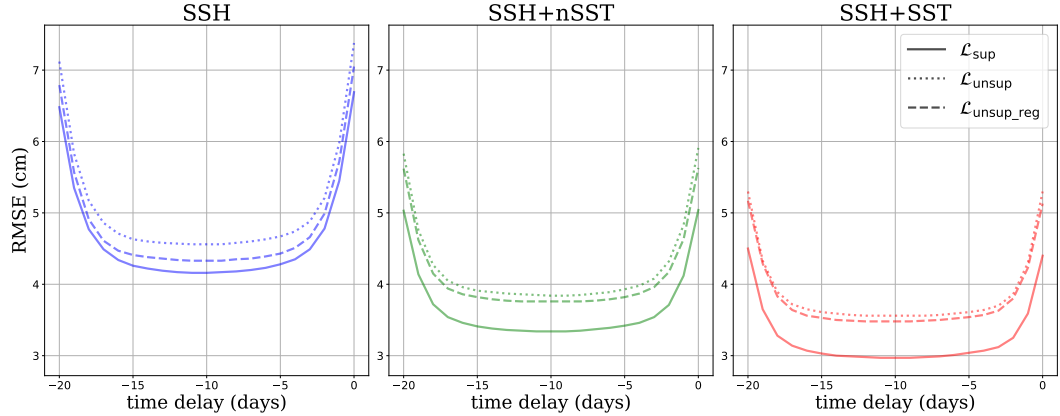
Figure 7. RMSE of the different reconstructions along the time window. The errors at a time delay of -20 correspond to an anti-causal scheme (knowing only future observations) whereas $timedelay = 0$ corresponds to a causal scheme (knowing no future observations). Knowing both past and future observations leads to the optimal reconstruction at $timedelay = -10$.

ages of observations 10 days in the future. In Figure 7 we can verify that 21 days of data

contain enough information to reconstruct the central image: for instance, 5 days from the border of the temporal window the reconstruction error is just 3% higher than the one at the center. This means that we can significantly reduce the delay (and therefore the training cost of our model) without causing severe drops in performance, which could be useful if applied in an operational framework. However, when it comes to producing NRT products (0 delay) this graph shows that we expect a significant loss of quality in the reconstruction which is usual (Amores et al., 2018; Stegner et al., 2021).

4.2 Eddy detection analysis

4.2.1 Importance of mesoscale eddies

Mesoscale eddies play an important role in ocean circulation and dynamics, and their understanding leads to diverse applications in oceanography or navigation (Chelton, Schlax, & Samelson, 2011). Previous studies underline how these structures transport heat, especially between latitudes 0° and 40° in the North Atlantic (Jayne & Marotzke, 2002), but also salinity (Amores et al., 2017), or plankton (Chelton, Gaube, et al., 2011). In practice, mesoscale eddies and structures are estimated through geostrophic currents derived from satellite altimetry. However, operational satellite products such as DUACS OI, have too coarse resolutions to resolve accurately mesoscale structures. Performing an OSSE to simulate the satellite's remote sensing, Amores et al. (2018); Stegner et al. (2021) showed that DUACS-like optimal interpolation aggregates small eddies into larger ones (i.e. with a radius greater than 100 km). These interpolations also capture a small percentage of eddies in the model simulation (around 6% in the North Atlantic) and change the eddies' distribution and properties. This is why we are interested in finding to what extent our reconstruction methods can detect small eddies in the ground truth, and how well the detected eddies are resolved and their physical properties conserved.

4.2.2 Automatic eddy detection algorithm: AMEDA

We use the Angular Momentum for Eddy Detection and tracking Algorithm (AMEDA) introduced by Vu et al. (2018) to perform the eddies detection. It is based on the Local Normalized Angular Momentum (LNAM), a dynamic metric first introduced by (Mkhinini et al., 2014), that we define hereafter:

$$\text{LNAM}(P_i) = \frac{\sum_j \overrightarrow{P_i P_j} \times \overrightarrow{V_j}}{\sum_j \overrightarrow{P_i P_j} \cdot \overrightarrow{V_j} + \sum_j |\overrightarrow{P_i P_j}||\overrightarrow{V_j}|} = \frac{L_i}{S_i + BL_i} \tag{8}$$

where P_i is the point of the grid where we compute the LNAM, P_j is a neighbor point of the grid, $\overrightarrow{P_i P_j}$ is the position vector from P_i to P_j and $\overrightarrow{V_j}$ is the velocity vector in P_j. Thus, the unnormalized angular momentum L_i is computed through a sum of cross products and is bounded by BL_i, so that if P_i is the center of an axisymmetric cyclone (resp anticyclone), $\text{LNAM}(P_i)$ will be equal to 1 (resp -1). Also, if the circulation field is hyperbolic and not an ellipsoid, S_i will reach large values, and $\text{LNAM}(P_i)$ will be close to 0. All sum is computed on a local neighborhood of P_i, which is a hyperparameter of the method (typically a square centered in P_i). In our case, we used the default parameters where the square has a length of $2\Delta x$, with Δx being the grid resolution (≈ 9 km).

AMEDA finds potential eddy centers by searching for the local extrema of the LNAM field, more precisely by taking the points P_i where $|\text{LNAM}(P_i)| > 0.7$. The characteristic contour of an eddy is then defined as the closed streamline of maximum velocity which does not include another eddy center. We perform the AMEDA algorithm on the geostrophic velocity field of our estimation and on the ground truth currents. We then look for the eddies that are both present in the ground truth and in our estimation. An eddy is said to be detected if the distance between its barycenter and the reference one is smaller than the average of the mean radius of the two characteristic contours. This definition allows

"multiple" detection (i.e., colocalization with several eddies). Therefore, we exclude eddies that include more than one candidate in the ground truth. For further details about the AMEDA algorithm, we refer the reader to Vu et al. (2018).

4.2.3 Eddy detection performances

We present the detection scores of the different reconstruction methods, with three data scenarios and three losses. We take the ensemble SSH estimation of the neural networks and perform the AMEDA algorithm on the velocity field derived through the geostrophic approximation (see Equation 1).

In Table 3 we present the F_1 score, the recall, and the precision of the methods. The recall tells us the proportion of actual positive instances that were correctly identified by the detection (a recall of 1 means that all ground truth eddies were detected). The precision gauges our trust in the detected eddies (a precision of 1 means that all eddies in the simulation were also present in the ground truth). To aggregate recall and precision, we use the F_1 score, which is the harmonic mean of recall and precision. A value of 1 means a perfect detection: all ground truth eddies were detected, and the estimation produced no false positives.

Loss		SSH			SSH+nSST			SSH+SST	
	F_1	recall	precision	F_1	recall	precision	F_1	recall	precision
$\mathcal{L}_{\text{sup}}$	0.699	0.607	0.825	0.735	0.657	0.833	**0.762**	**0.705**	0.83
$\mathcal{L}_{\text{unsup}}$	0.692	0.634	0.76	0.715	0.665	0.772	0.731	0.694	0.771
$\mathcal{L}_{\text{unsup\_reg}}$	0.684	0.581	0.831	0.704	0.608	0.835	0.713	0.622	**0.836**

Table 3. Scores of the AMEDA eddy detection performed on the Ensemble estimation of ABED interpolation. The considered scores are the precision, the recall, and the F_1 score.

Data comparison. As expected, no matter which loss we consider, the noise-free temperature detection method outperforms the two other scenarios with higher F_1 scores. Even the noisy SST provides important information for eddy reconstruction, as the SSH-only method yields lower results than the two other scenarios. We also see that for each loss, the precision scores are less impacted by the input data than the recall is. This means that the SSH-only scenario does not produce a lot more false detection than the SST methods but misses much more structures.

Loss comparison. On the other hand, the loss function used to perform the inversion substantially impacts precision and recall. The regularization of the unsupervised loss brings the detection precision to the level of the supervised method (even higher for the SSH-only and SSH+SST) but also reduces the recall of all methods compared to their unregularized version. In other words, the regularization prevents the neural network from generating false eddies and from retrieving some structures, which leads to lower F_1 scores.

Visual comparison. We plot in Figure 8 the SSH maps and eddies detected by AMEDA, and in Figure 9 the relative vorticity ξ computed from geostrophic currents (see Equation 1) as follows:

$$\xi = \frac{\partial v_{\text{geo}}}{\partial x} - \frac{\partial u_{\text{geo}}}{\partial y} \tag{9}$$

Relative vorticity is an important quantity in the analysis of surface circulation as it highlights areas of important direction change of the stream. ξ is positive in counterclockwise spin and negative in clockwise spin. In the presented figures, we normalize relative
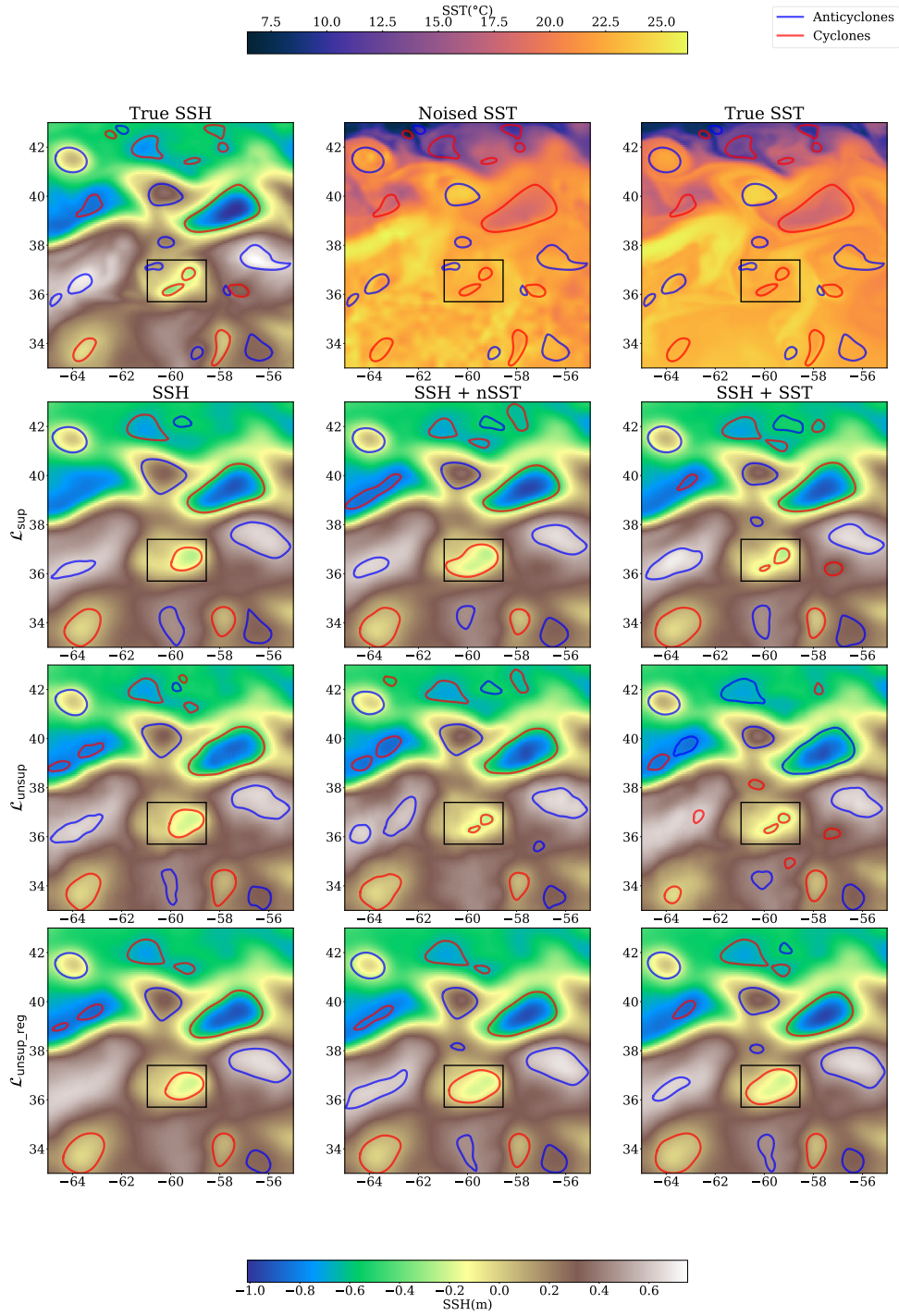
Figure 8. SSH maps and detected eddies the 1st June 2017 on our OSSE. The first line presents the True SSH, the noised SST, and the True SST, on which we plot the eddies detected on the True SSH. The second, third, and last lines present respectively the inversion using $\mathcal{L}_{\text{sup}}$, $\mathcal{L}_{\text{unsup}}$, and $\mathcal{L}_{\text{unsup\_reg}}$. The first, second, and last columns present the maps using the SSH-only, SSH+nSST, SSH+SST data, respectively. Each SSH map is the ensemble reconstruction of 3 networks with their associated eddies.
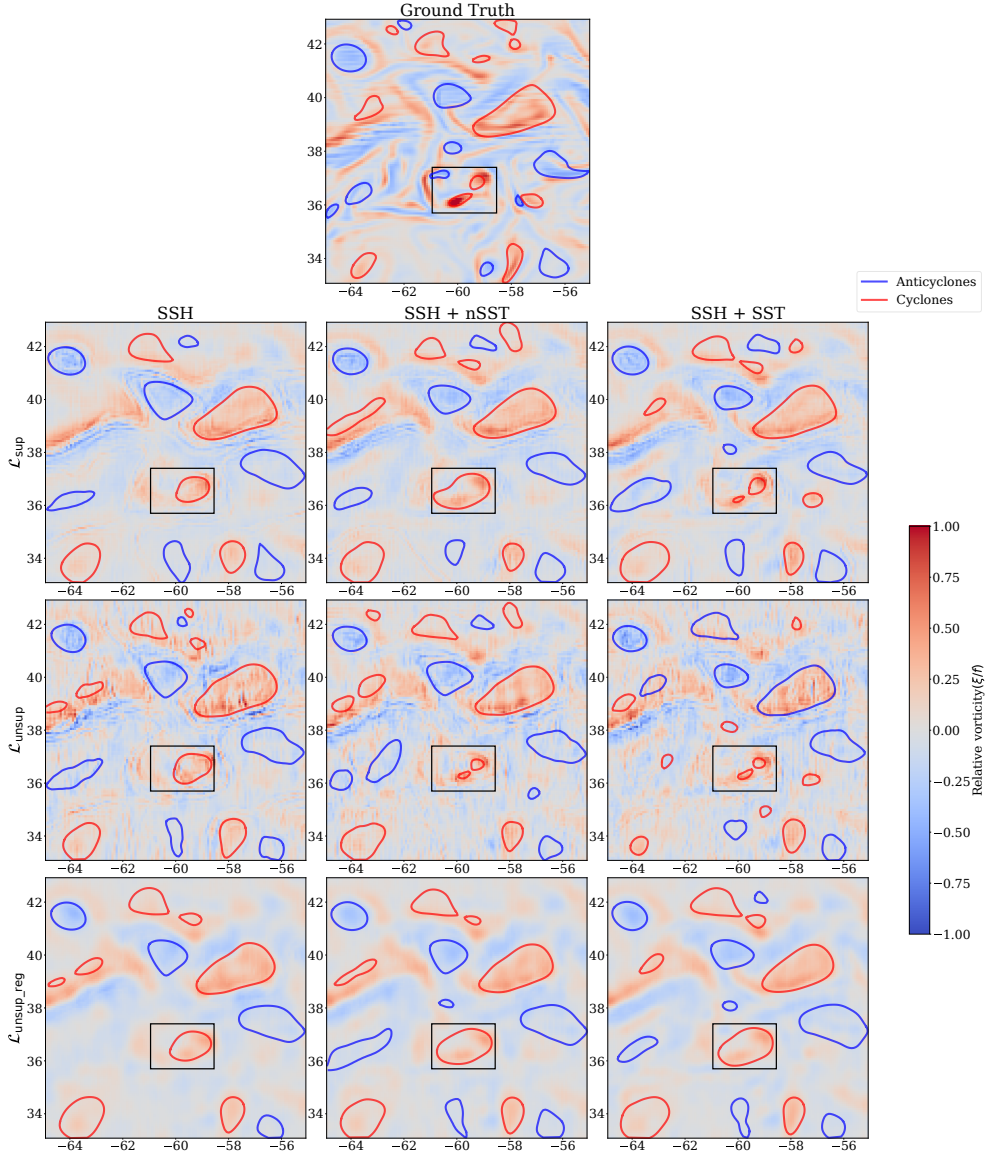
Figure 9. Relative vorticity (normalized by the Coriolis factor) and detected eddies on the 1st June 2017 in our OSSE. The first line presents the True relative vorticity. The second, third, and last lines present the neural networks trained with $\mathcal{L}_{sup}$, $\mathcal{L}_{unsup}$, and $\mathcal{L}_{unsup\_reg}$. The first, second, and last columns present the SSH-only, SSH+nSST, and SSH+SST interpolations. Each relative vorticity map is computed from the ensemble SSH estimation of the 3 networks.

vorticity fields by the Coriolis factor f. Figures 8,9 illustrate an example of the conclusions established in Table 3: the SSH-only reconstruction shows fewer eddies than the ones using SST, and aggregates small eddies into larger ones (see highlighted eddies). We also see the effect of regularization, especially in the relative vorticity fields, which are a lot smoother than the ones in the supervised and regularized inversion. This smoothing effect results in a reduced number of detected eddies, as illustrated by the two highlighted eddies that are detected separately when SST is used without regularization.

4.2.4 Physical properties of detected eddies

To further investigate the performance of the eddy detection methods, we analyze the detection outcomes based on the physical characteristics of the eddies. For instance, smaller eddies tend to have shorter lifespans, making them more challenging to detect due to their decreased likelihood of being observed by satellites. Conversely, high-speed eddies are derived from important sea surface height (SSH) variations, thus exhibiting a strong signature in the generated mapping. Figure 10 shows the detection performances as a function of some key parameters such as maximum radius, lifetime, or maximum velocity along the final closed current line.

As anticipated, using SST and nSST data contributes to the detection of eddies, as indicated by the higher F_1 scores achieved in every loss scenario. However, small and short-lived eddies are less frequently detected, resulting in lower recall scores. Specifically, only 17% of the eddies with a radius below 15 km are successfully detected in the best scenario. Nonetheless, except for the unregularized loss function, the precision scores for the detected eddies remain high, even for small and short-lived ones. This observation confirms the previously observed phenomenon where the regularization employed in the inversion process prevents the network from generating false eddy detections but also stops it from capturing a significant portion of the actual eddies. This regularization behavior is expected, as forcing a smoothness constraint on the SSH gradient field leads to denying some of the small structures.

We also want to assess the model's accuracy to estimate the eddies' physical properties. To this end, we focus on the eddies that were successfully detected by all the methods (3534 eddies out of the 7908 eddies in the ground truth) and compare the physical parameters of the estimated eddies to their values in the corresponding true eddy. We compute the RMSE and bias of the following parameters: maximum radius and velocity of the characteristic contour of the eddies. Once again, Tables 4 and 5 show that SST helps to estimate eddies radius, and velocity. Nonetheless, there is a bias of radius and velocity: the size of the eddy is statistically overestimated compared to its ground truth, while its speed is systematically underestimated. This is particularly true for the regularized unsupervised loss because of its smoothness constraint, with a velocity bias accountable for half of the RMSE.

Loss	SSH		SSH+nSST		SSH+SST	
	RMSE	bias	RMSE	bias	RMSE	bias
$\mathcal{L}_{\mathrm{sup}}$	16.7	3.6	15.7	4.2	**14.7**	3.7
$\mathcal{L}_{\mathrm{unsup}}$	16.6	**0.9**	16.3	1.3	15.5	1.3
$\mathcal{L}_{\mathrm{unsup\_reg}}$	16.6	3.5	16.5	4.1	15.7	4.5

Table 4. Eddies maximum radius RMSE and bias (km). The eddy detection is performed on geostrophic currents of the ensemble estimation and the bias is computed from the estimated radius minus ground truth radius.

Loss	SSH		SSH+nSST		SSH+SST	
	RMSE	bias	RMSE	bias	RMSE	bias
$\mathcal{L}_{\mathrm{sup}}$	14.3	-5.3	12.4	-3.1	**11.7**	**-2.0**
$\mathcal{L}_{\mathrm{unsup}}$	14.4	-5.7	13.4	-3.3	12.5	-2.9
$\mathcal{L}_{\mathrm{unsup\_reg}}$	15.2	-8.3	13.9	-7.4	13.2	-6.5

Table 5. Eddies maximum velocity RMSE and bias (cm/s).
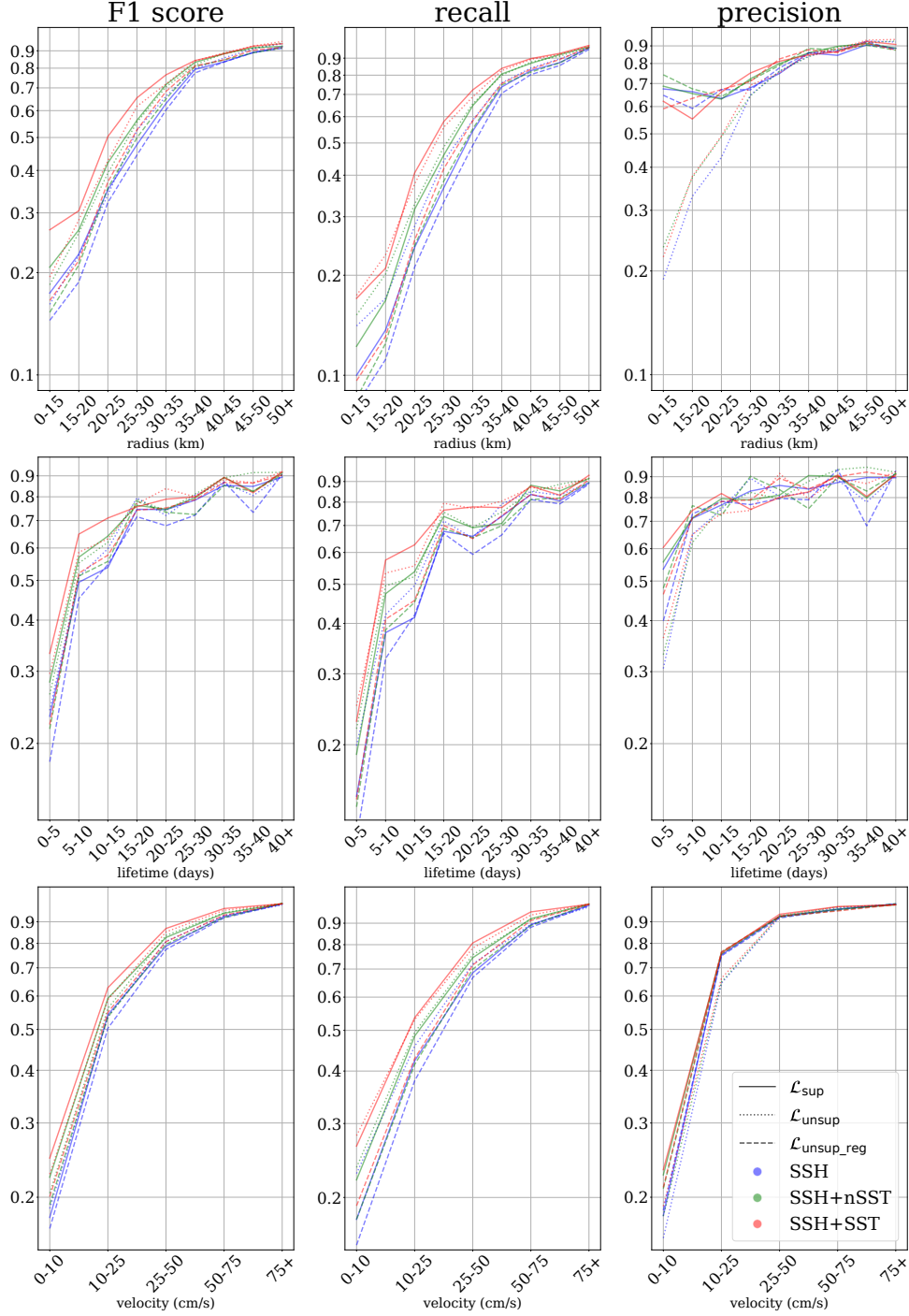
Figure 10. Detection scores of the different methods on eddies separated by radius (first row), lifetime (second row), and maximum velocity (last row). The considered scores are F_1 (first column), recall (second column), and precision (third column). The recall tells the proportion of actual positive instances that were correctly identified, the precision gauges the trust that we can put in the detected eddies, and the F_1 score aggregates these two values.

4.3 Comparison with state-of-the-art methods on a NATL60 OSSE

Comparing various SSH interpolation methods requires a common benchmark and metrics. The Ocean Data Challenge 2020 (CLS/MEOM, 2020) provides an OSSE sim-

ilar to the one described in Section 2, the state-of-the-art estimations and metrics. The included data are the ground truth SSH, nadir-pointing observations, and a simulation of the SWOT (Surface Water and Ocean Topography) observations, a new altimetry technology (Gaultier et al., 2016). In this study, we have excluded the SWOT measurements as we do not simulate them in our OSSE and focus on nadir-pointing data. The ground truth is the NATL60 simulation (Ajayi et al., 2019) which uses the same physical model (NEMO 3.6) (Madec et al., 2017) but at finer scales than GLORYS12, and without assimilation. Given that the NATL60 model also outputs SST and ocean currents fields, we retrieved and used these variables, even though they were not included in the official depository of the challenge. The state-of-the-art frameworks presented in this challenge are the following:

- DUACS: the operational linear optimal interpolation leveraging covariance matrix tuned on 25 years of data;
- DYMOST (Ubelmann et al., 2016; Ballarotta et al., 2020) and MIOST (Ardhuin et al., 2020): two variants of the linear optimal interpolation where the Gaussian covariance model is changed for a non-linear quasi-geostrophic dynamic model (for DYMOST) or by a wavelet base (MIOST);
- BFN (Le Guillou et al., 2020): a data assimilation method that performs a back and forward nudging of a quasi-geostrophic model;
- 4DVarNet (Fablet et al., 2021): a deep learning framework supervised on the Ocean Data Challenge 2020. In this configuration, it only takes SSH observations as input;
- MUSTI (Archambault et al., 2023): an unsupervised neural network fitting SSH along tracks observations starting from an SST image. The fact that this method must be fitted to new observations, limits its operational use.

This benchmark is not complete as the ConvLTSM interpolations introduced by S. A. Martin et al. (2023) were trained on real satellite observations only, and the 4DVarNet versions using SST were only computed using SSH observations from nadir pointing satellites and SWOT data (Fablet et al., 2023). Still, we are interested in evaluating the reconstructions of our networks, trained on our OSSE, on the Ocean Data Challenge 2020 to show the similarity of the two simulated observation systems. To produce our estimation, we regrid the provided data to our resolution (from 0.016° to 0.078°) using trilinear interpolation. We use the SSH simulated observations of the data challenge and the SST of the corresponding NATL60 simulation. The test period includes 42 days of simulation (between 2012/10/22 and 2012/12/02) as defined in the challenge. As such, the comparison is not fully fair since regridding and not training on the same data might bias the scores obtained. It is still a good way to evaluate the similarity of our OSSE to the Ocean Data Challenge 2020, as our approach obtains comparable performances to the state-of-the-art. Each method is then evaluated using the following metrics, and we sum up the results in Table 6:

- μ and σ_t (in cm), are respectively the RMSE of the SSH and the temporal standard deviation of this RMSE. In the data challenge, these two metrics are normalized by the root mean square of the SSH, but we prefer giving it in centimeters to be coherent with the rest of the work;
- λ_x (in degrees) and λ_t (in days) are two spectral metrics, introduced by (Le Guillou et al., 2020). We compute respectively the spatial and temporal power spectrum of the error, λ_x is then the smallest spatial wavelength where the power spectrum of the error is equal to the power spectrum of the signal and λ_t its temporal equivalent. For further information, we refer the reader to (Le Guillou et al., 2020);
- μ_u and μ_v (in cm/s) are the RMSE between the NATL60 currents and the geostrophic currents of the estimation.

Method	SST	SUP	μ	σ_t	λ_x	λ_t	μ_u	μ_v
DUACS	✗	✗	4.89	3.02	1.42	12.08	16.8	16.2
DYMOST	✗	✗	5.18	3.05	1.35	11.87	16.8	16.8
MIOST	✗	✗	4.21	2.5	1.34	10.34	14.9	14.5
BFN	✗	✗	4.7	2.73	1.23	10.64	15.1	15.3
4DVarNet	✗	✓	3.26	1.73	**0.84**	7.95	13.1	12.8
MUSTI	✓	✗	3.12	1.32	1.23	**4.14**	12.2	14.2
ABED-SSH $\mathcal{L}_{\text{sup}}$	✗	✓	3.75	2.0	1.21	8.74	13.3	13.5
ABED-SSH $\mathcal{L}_{\text{unsup}}$	✗	✗	4.06	2.19	1.32	9.29	13.7	15.1
ABED-SSH $\mathcal{L}_{\text{unsup\_reg}}$	✗	✗	4.23	2.36	1.24	9.98	13.8	14.2
ABED-SSH-SST $\mathcal{L}_{\text{sup}}$	✓	✓	**2.88**	**1.24**	0.95	4.51	**11.4**	**11.4**
ABED-SSH-SST $\mathcal{L}_{\text{unsup}}$	✓	✗	3.08	1.41	1.18	5.18	11.8	12.8
ABED-SSH-SST $\mathcal{L}_{\text{unsup\_reg}}$	✓	✗	3.39	1.65	1.18	5.7	12.4	12.3

Table 6. Comparison of the state-of-the-art reconstruction methods on the Ocean Data Challenge 2020. SST stands for whether or not the reconstruction methods are using SST, and SUP stands for whether or not the methods are supervised.

We see in the scores a predominance of neural network-based methods (MUSTI, 4DVarNet and ABED) as the importance of the SST in the reconstruction (MUSTI, and ABED). The ABED-SSH networks do not perform as well as 4DVarNet, but better than optimal interpolations (DUACS, DYMOST, MIOST) and BFN. This analysis further supports using SST data in deep learning-based methods for these inverse problems. We can expect around 2 cm of error reduction on the operational interpolation scheme DUACS with our best method (41% of reduction). We also significantly reduce the errors on currents compared to DUACS's, by $5.7 \, \text{cm/s}$ for u and $5.4 \, \text{cm/s}$ for v (35% and 34% error reduction).

4.4 Application to real satellite observations

In this section we focus on applying the developed methods to real observations with two objectives in mind: show the utility and realism of our OSSE compared to the pre-existing one, and explore transfer learning strategies. To evaluate our method on a shared benchmark, we use the Ocean Data Challenge 2021 (CLS/MEOM, 2021), which provides one year of real SSH nadir observations and evaluation metrics. All the evaluations presented in this section are computed on the along-track data from the CryoSat-2 satellite left aside in all the benchmarked methods. The comparison is done on the entire 2017 year, which is the year that we left aside from training on our OSSE to avoid data leakage. To be coherent with the area covered by all the methods, the evaluation area is smaller than the one of the OSSE (between 34° to 42° North and -65° to -55° East). These real-world measurements present instrumental errors that produce much higher RMSE scores than the ones computed on the OSSE. Also, as we do not have access to complete SSH maps, the metrics used are μ, σ_t, and λ_x (in km this time). For methods requiring SST information, we use satellite SST from (NASA/JPL, 2019) described in Section 2.3.

4.4.1 OSSE comparison

In this part, we compare the generalization to real satellite data of models trained on our OSSE with models trained on the Ocean Data Challenge 2020. As this last dataset provides one year of data it can also be used to fit neural networks, but as shown in Appendix 6.2, training on a longer dataset drastically improves reconstructions. As the ex-

isting OSSE does not provide SST data, it is possible to use NATL60 SST, but the lack of realistic noise leads to a domain gap with real data. To this day, if SSH-only neural networks have been successfully transferred to real SSH data, this is not the case for SST-aware ones. We compare ABED trained in a supervised way on our OSSE (SSH-only or using noisy SST), and on the Ocean Data Challenge 2020 (SSH-only or with NATL60 SST output). To train ABED on NATL60 data, we regrid the input and target data to our resolution, and use the data split of the challenge (CLS/MEOM, 2020); validation of the training between 2012/10/22 and 2012/12/02, and fitting on the remaining days. We use the same hyperparameters as for the training on our OSSE.

Once networks are trained on the simulation, we perform inferences on real data, excluding the tracks from the independent satellite. In Table 7, we present the mean and

Method	Training OSSE data	$\mu(cm)$	$\sigma_t(cm)$	$\lambda_x(km)$
ABED-SSH	Ocean Data Challenge 2020	8.90 — 8.50	3.18 — 3.10	148 — 143
ABED-SSH-SST	Ocean Data Challenge 2020	10.11 — 9.73	3.38 — 3.30	142 — 137
ABED-SSH	Ours	6.63 — 6.35	2.02 — 1.90	122 — 119
ABED-SSH-SST	Ours	6.28 — 6.06	1.77 — 1.73	115 — 113

Table 7. Comparison of ABED networks trained on our OSSE to the ones trained on the Ocean Data Challenge 2020. All the metrics are computed on independent real data of the Ocean Data Challenge 2021. The left scores are the mean performances on three networks and the right ones are the ensemble scores.

ensemble scores of the models on the Ocean Data Challenge 2021. As expected, ABED performs significantly better when trained on our OSSE. Specifically, ABED-SSH-SST trained on the Ocean Data Challenge leads to higher errors than its SSH-only version, which shows the domain gap between NATL60 and satellite SST. We conclude that the length of our OSSE and the addition of SST realistic noise enhanced the reconstructions of the real-world SSH.

4.4.2 Transfer OSSE learning to real-world data.

Enhancing real-world SSH reconstruction using the information of a simulation is a typical transfer learning problem, where we have access to ground truth in a source domain (OSSE) but not in a target domain (satellite data) (Pan & Yang, 2010). Given the losses described in Section 3.3 and a satellite dataset (see Section 2.3), we can consider three ways to apply our methodology to the Ocean Data Challenge 2021. We partially presented this experiment in (Archambault et al., 2024).

Observation only: Perform an unsupervised training on real-world data, with the loss function described in Equation 6. The training hyperparameters and dataset split are the same as the ones used in the OSSE study (see Section 3.4).

Simulation only: Use the networks trained on our OSSE in a supervised way directly on satellite data. As the test year of our OSSE and one of the Ocean Data Challenge 2021 are the same, we have no issues with data leaking.

Pre-training on OSSE and fine-tuning on satellite data: After the supervised pre-training on OSSE data we fine-tune the neural network on satellite data for a few epochs using the unsupervised loss. The fine-tuning is done using a small learning rate of 1.10^{-5} and a decay of 0.9. We use an early stopping with a patience of 8 epochs and we save the best model on the validation set.

We present in Table 8 the RMSE on the Ocean Data Challenge 2021 of 3 ABED networks trained with the previously mentioned methodologies. One of the first conclusions we can draw from these reconstruction scores is the interest of our OSSE in the training process. The networks fitted on the simulation perform better than their equivalent trained with observations only, except for the network trained using noise-free SST. This shows that our SST noise is realistic, as introducing SST noise during pre-training is beneficial for generalization to satellite data. Secondly, in every data scenario, the pre-trained and fine-tuned networks perform significantly better than their version trained on observation or simulation. In particular, once fine-tuned, the networks pre-trained on nSST and on SST lead to close performance, whereas without fine-tuning, the network trained on noise-free SST produces the worst reconstruction. Given an appropriate fine-tuning strategy, the features learned on noise-free SST that do not apply to satellite data are effectively modified. From this experiment, we conclude that combining supervised training on our OSSE with unsupervised re-fitting on satellite data increases performance, especially if SST is used.

Learning method / Input data	SSH	SSH+nSST	SSH+SST
Observation	7.07 — 6.75	6.63 — 6.27	—
Simulation	6.63 — 6.35	6.28 — 6.06	6.89 — 6.68
Pre-training & Fine-tuning	6.49 — 6.28	**6.02 — 5.82**	6.04 — 5.84

Table 8. Along-track SSH RMSE in centimeters (mean score on the left and ensemble score on the right) of 3 ABED networks, computed on 1 year of data provided by the Ocean Data Challenge 2021. The training strategies include observation-only training (with satellite SSH and SSH+nSST), simulation-only training (SSH, SSH+nSST, SSH+SST), and fine-tuned networks (SSH, SSH+nSST, SSH+SST). For the Fine-tuned networks, when a network is pre-trained with noise-free SST, it is still fine-tuned with noisy satellite SST.

In Table 9, we compare our method to the state-of-the-art interpolation methods provided in the context of the Ocean Data Challenge 2021. The included methods are the same as in Table 6, plus the ConvLTSM-SSH and ConvLTSM-SSH-SST (S. A. Martin et al., 2023). We give ensemble scores of the three pre-trained and fine-tuned ABED networks using only SSH, or SSH and the noised SST. The enhanced scores of ABED-SSH-SST and ConvLTSM-SSH-SST compared to their SSH-only versions emphasize the improvements brought by the SST. ABED, ConvLTSM and 4DVarNet lead to better SSH gridding than optimal interpolation-based methods (DUACS, DYMOST, MIOST) both in terms of RMSE and effective spatial resolution. We also note a significative drop in RMSE score for the BFN method compared to its OSSE reconstruction, which shows that the idealized QG model is less applicable to real-world observations.

In Figure 11, we present the SSH maps on the different reconstruction methods with their associated relative vorticity (see Equation 9). The three first methods (DYMOST, DUACS, MIOST) present smooth vorticity maps as a consequence of the optimal interpolation. All the vorticity maps from neural network-based methods: 4DVarNet, MUSTI, ConvLTSMs, and ABEDs have higher contrast and also some artifacts due to convolution operations. 4DVarNet in particular, produces very high-frequency variations on which we can see the input satellite path. We suppose this is a consequence of the U-Net's skip connections whereas the other networks have Encoder Decoder architectures, less prompt to produce high-frequency noise. For the last four methods, ConvLTSM-SSH, ConvLTSM-SSH-SST, ABED-SSH, and ABED-SSH-SST, we highlights areas where small structures are

Method	SST	Learning	$\mu(cm)$	$\sigma_t(cm)$	$\lambda_x(km)$
DUACS	✗	✗	7.66	2.66	149
DYMOST	✗	✗	6.75	2.00	131
MIOST	✗	✗	6.75	2.00	139
BFN	✗	✗	7.46	2.59	119
4DVarNet	✗	Simulation	6.56	1.84	**107**
MUSTI	✓	Observation	6.26	1.96	114
ConvLTSM-SSH	✗	Observation	6.82	1.86	114
ConvLTSM-SSH-SST	✓	Observation	6.29	**1.60**	108
ABED-SSH	✗	Pre-training & Fine-tuning	6.28	1.82	118
ABED-SSH-SST	✓	Pre-training & Fine-tuning	**5.82**	1.61	108

Table 9. Comparison of the state-of-the-art reconstruction methods on the real satellite data of the Ocean Data Challenge 2021. SST stands for whether or not the reconstruction methods are using SST. ABED-SSH and ABED-SSH-SST stands for the ensemble score of our pre-trained and fine-tuned networks.

visible in the vorticity maps of the SST-using methods but not in their SSH counterparts. The similar shape of the structures between ConvLTSM-SSH-SST and ABED-SSH-SST suggests that they are linked to the use of SST and not the deep learning method.

5 Conclusion and perspectives

5.1 Summary

In this work, we designed a new OSSE emulating 20 years of satellite observations of SSH and SST while the previously existing OSSE provided only one year of simulated SSH observations (CLS/MEOM, 2020). We were able to train an Attention-Based Encoder Decoder using 3 different loss functions (2 of them learning the reconstruction without ground truth), on three different sets of data (SSH only, SSH and noised SST, SSH, and SST). We show a systematic interpolation improvement thanks to the use of SST. Using temperature data (noisy or not), the unsupervised inversion outperforms even the supervised SSH-only neural network (3.86 cm of RMSE for the unsupervised noisy SST against 4.18 cm for the supervised SSH-only method). This shows the importance of contextual information to constrain this inverse problem, even while learning with observation only.

Using AMEDA, an automatic eddy detection algorithm, we were able to identify cyclones and anticyclones in the ground truth and compare them with the eddies detected in the geostrophic approximation of the different mappings. This allows a deeper physical interpretation than the SSH reconstruction alone. We conclude that SST aids in capturing finer structures that might be overlooked by SSH-only methods and that SST-using methods better render the key physical properties of the detected eddies, such as size, speed, or center position. Furthermore, in unsupervised reconstruction, we show that the non-regularized and regularized inversions have close detection scores, but their errors are different. The regularized inversions exhibited lower recall scores, indicating that certain eddies were not detected due to the smoothing effect of the regularization process. However, they demonstrated higher precision scores, implying increased confidence in the successfully detected eddies.

We evaluate ABED trained using the data from our OSSE on the Ocean Data Challenge 2020 and compare it with state-of-the-art interpolation techniques. We show that
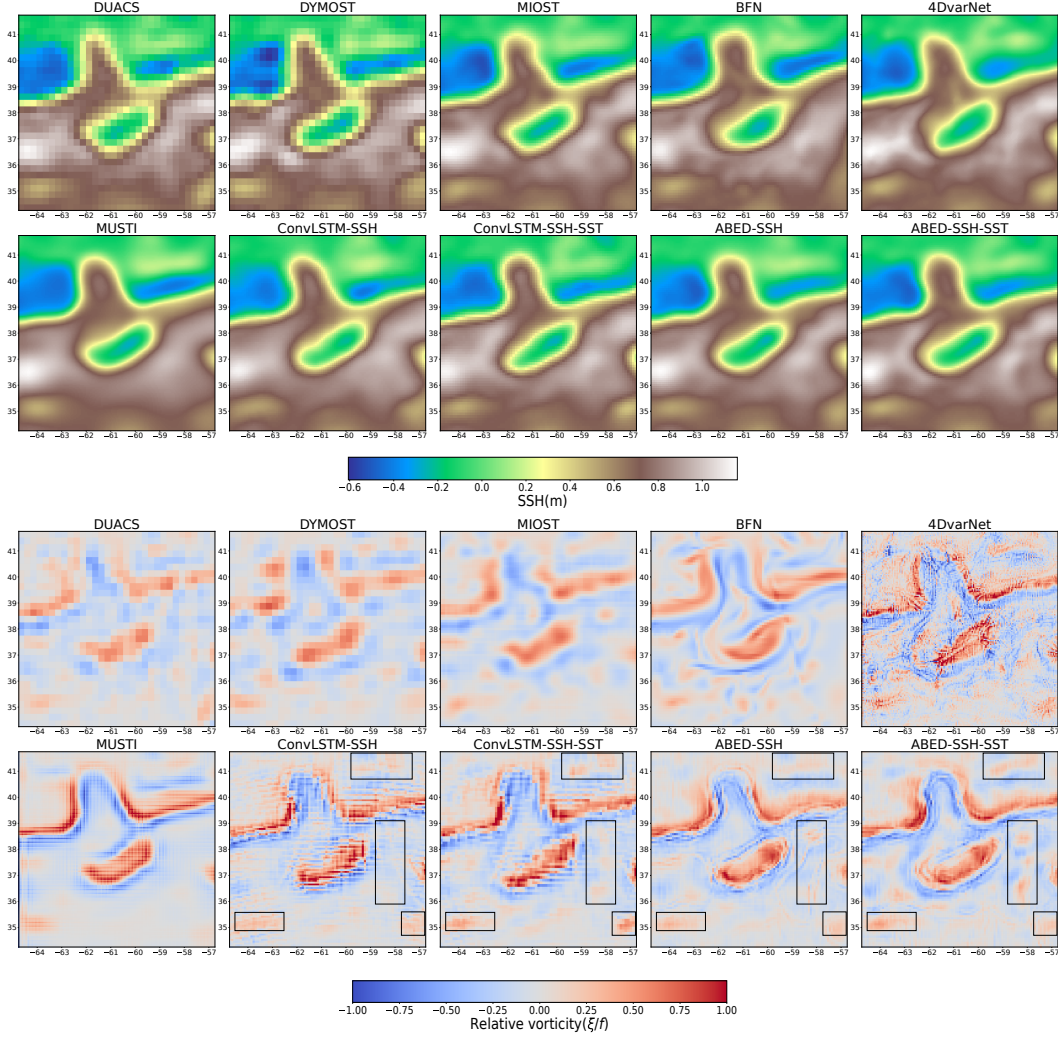
Figure 11. SSH maps and Relative Vorticity maps (normalized by the Coriolis frequency) of the methods from Table 9. The SSH maps are used to compute geostrophic currents from which we derive Relative Vorticity. Due to the different areas covered by the methods, we plot the SSH and RV on a portion of the training area; from 34.25° to 41.75° North and from -64.75° to -56.75° East. On the last four relative vorticity maps, we highlight some regions where small structures are visible in the SST-using interpolations and not visible (or less salient) in their SSH counterparts.

the utilization of SST led to a substantial improvement of 41% in terms of RMSE for SSH compared to the widely used L4 product from DUACS. Moreover, we observed significant improvements of 34% and 35% for u and v currents, respectively. These findings present promising perspectives for advancing satellite SSH gridding through the application of deep learning methodologies and the fusion of diverse physical information.

Finally, we presented a novel training strategy, using jointly OSSE and real-world satellite observations. We proposed to perform a transfer from the OSSE to the satellite domain by pre-training the neural network on the OSSE and fine-tuning it on a real-world dataset in an unsupervised way. Comparing the same network trained following three strategies: on simulation only, on observations only, or the one introduced here,

we found that using together simulation and satellite data leads to better performances. Specifically, our transfer method achieves state-of-the-art performances on the Ocean Data Challenge 2021, on which we report an RMSE improvement of 24% compared to DUACS.

5.2 Perspectives

SSH Forecast. This study focused on a delayed time interpolation of the SSH. However, near real-time and forecast data are often useful in many operational applications, such as navigation and meteorology. In future works, we would be interested in extending the output window in the future compared to the input one. In doing so, the neural network would be trained to interpolate and forecast the SSH simultaneously. We would be interested to compare a method doing the two tasks simultaneously to a method doing it successively.

Global interpolation. Furthermore, many challenges still need to be addressed to get toward a global gridded SSH product. For instance, as the geostrophic equilibrium depends on the Coriolis force surface projection and thus on the latitude considered, we may require a model to be trained on several areas with different latitudes. Also, we can wonder which strategy is more efficient between training a global model or several local models, each one specialized for a range of latitude or geographical area. Closed seas and coastal water also have very different physical interactions and might need to be reconstructed by different methods.

Using different input and output data. We have demonstrated the benefit of using multi-physical information, specifically SST, to enhance SSH reconstruction through the implementation of a flexible neural network framework. The integration of data from diverse physical sources exhibits promising outcomes, yet conventional model-based methods encounter challenges due to noise and observational difficulties associated with real-world data. In contrast, machine learning opens doors to augment these methods with diverse and abundant data sources. For instance, we employed noisy yet complete SST data in our investigation, but using L3 SST products is also possible. Furthermore, an intriguing prospect arises as to whether Level 4 (L4) and Level 3 (L3) SST products can be effectively combined, thereby potentially yielding even more precise and exhaustive information. Other physical measurements might improve the reconstruction, such as chlorophyll maps that track plankton advected by currents (Kahru et al., 2012).

Data availability statement

The GLORYS12 data (CMEMS, 2020) that we used as a reference throughout this study are freely available and distributed by the European Union-Copernicus Marine Service (`https://doi.org/10.48670/moi-00021`). The L3 altimeter (CMEMS, 2021) measurements used to retrieve along tracks coordinates and the L3 SST measurements(CMEMS, 2023) used to compute a realistic cloud cover are distributed by the same service (with doi:
`https://doi.org/10.48670/MOI-00146` and `https://doi.org/10.48670/MOI-00164` respectively).

The data of the Ocean Data Challenge 2020 OSSE (ground truth, inputs, and baselines) are available at `https://doi.org/10.24400/527896/A01-2020.002` and were developed, validated by CLS and MEOM Team from IGE (CNRS-UGA-IRD-G-INP), France and distributed by Aviso+. To this benchmark data, we add the SST and the surface current of the NATL60 simulation available here: `https://github.com/CIA-Oceanix/4dvarnet-james-uv-ssc`.

The Satellite dataset is composed of the L3 altimeter measurements previously described (CMEMS, 2021), and the Multiscale Ultra-High Resolution SST dataset(NASA/JPL, 2019), available on the following links:
`https://podaac.jpl.nasa.gov/dataset/MUR25-JPL-L4-GLOB-v04.2`.

The Ocean Data Challenge 2021 (CLS/MEOM, 2021), data and baselines are available here: `https://doi.org/10.24400/527896/a01-2021.005`. We add to this benchmark the estimated maps from (S. A. Martin et al., 2023) and (Archambault et al., 2023), taken from `https://zenodo.org/records/7730739` and
`https://gitlab.lip6.fr/archambault/visapp2023`, respectively.

The preprocessed data and the weights of our neural networks are available here: `https://doi.org/10.5281/zenodo.8380280` and our code is hosted on the following repository: `https://gitlab.lip6.fr/archambault/james2024`.

Acknowledgments

6 Annexes

6.1 Along-track spatial derivatives

We calculate the SSH's first and second spatial derivatives along the satellite ground track as described in Equation 10 and 11. Given $\mathbf{Y}^{ssh}$, the list of SSH measurements from one satellite (sorted in time), we approximate the derivative by the rate of change of the SSH:

$$\frac{\partial}{\partial s}\mathbf{Y}_i^{ssh} \simeq \frac{\mathbf{Y}_{i+1}^{ssh} - \mathbf{Y}_i^{ssh}}{\triangle s_i} \tag{10}$$

$$\frac{\partial^2}{\partial s^2}\mathbf{Y}_i^{ssh} \simeq \frac{\frac{\partial}{\partial s}\mathbf{Y}_{i+1}^{ssh} - \frac{\partial}{\partial s}\mathbf{Y}_i^{ssh}}{\triangle s_i'} \tag{11}$$

where $\mathbf{Y}_i^{ssh}$ is the i-th measurement of SSH, $\triangle s_i$ is the ground distance between the SSH measurements, and $\triangle s_i'$ is the ground distance between the two first derivative approximations. The lists of first and second spatial derivatives, $\frac{\partial}{\partial s}\mathbf{Y}^{ssh}$ and $\frac{\partial^2}{\partial s^2}\mathbf{Y}^{ssh}$, are recentered on new coordinates, corresponding to the dual coordinates of the $\mathbf{Y}^{ssh}$ and $\frac{\partial}{\partial s}\mathbf{Y}^{ssh}$, respectively. We only compute the spatial derivatives from observations coming from the same satellite and only if the measurements are taken with less than two seconds of delay. This way we estimate spatial derivatives only where the rate of change is a valid approximation of the derivation.

6.2 Impact of the OSSE temporal length on training

Our OSSE dataset is composed of 7194 days, which leads to 5504 training days once the partition between train, validation, and test sets is made. To evaluate the interest in using more data to constrain the neural network, we train ABED network in the optimal configuration (supervised and using noise-free SST). We compare the scenario where all the samples are seen during training with those where only half, a quarter, or a single year of the dataset is used. The validation and test sets remain unchanged, while the training subset is the first consecutive days from the initial training set. Table 10 presents the RMSE of the reconstructions on the test year of our OSSE. The scores of the net-

works trained with different dataset sizes clearly show better reconstruction performance when the size increases.

Size	Full	1/2	1/4	1 year
Number of training samples	5504	2752	1376	365
RMSE (cm)	2.97	3.97	4.77	7.76

Table 10. Mean RMSE score (in cm) of 3 ABED networks trained on our OSSE in a supervised manner using SSH and noise-free SST. We compare the situation where the full, half, a quarter, or one year of the dataset is used.

6.3 Impact of the SST deseasonalization on reconstruction

In the results presented in this work, we deseasonalized the SST data in the inputs of the neural networks. In Table 11, we show the RMSE of the neural network using "native" SST and the ones using deseasonalized SST. We see that this preprocessing operation decreases the RMSE in every scenario.

Loss	SSH+SST	SSH+SST (deseasonalized)
$\mathcal{L}_{\text{sup}}$	3.19 — 2.88	**2.97** — **2.63**
$\mathcal{L}_{\text{unsup}}$	3.50 — 3.09	3.56 — 3.16
$\mathcal{L}_{\text{unsup\_reg}}$	3.52 — 3.26	3.48 — 3.20

Table 11. SSH reconstruction RMSE in centimeters (mean score on the left and ensemble score on the right) of 3 ABED networks. The interpolation is trained using the 3 different losses described in Section 3.3 with the following settings: SSH + noise-free SST and SSH + deseasonalized noised-free SST.

References

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, *29*, 594-621. doi: 10.1080/07474938.2010.481556

Ajayi, A., Le Sommer, J., Chassignet, E., Molines, J.-M., Xu, X., Albert, A., & Cosme, E. (2019). Spatial and temporal variability of north atlantic eddy field at scale less than 100 km. *Earth and Space Science Open Archive*, 28. doi: 10.1002/essoar.10501076.1

Amores, A., Jordà, G., Arsouze, T., & Le Sommer, J. (2018, October). Up to what extent can we characterize ocean eddies using present-day gridded altimetric products? *Journal of Geophysical Research: Oceans*, *123*, 7220–7236. doi: 10.1029/2018JC014140

Amores, A., Melnichenko, O., & Maximenko, N. (2017, January). Coherent mesoscale eddies in the North Atlantic subtropical gyre: 3-D structure and transport with application to the salinity maximum. *Journal of Geophysical Research: Oceans*, *122*, 23-41. doi: 10.1002/2016JC012256

Archambault, T., Filoche, A., Charantonnis, A., & Béréziat, D. (2023, February). Multimodal unsupervised spatio-temporal interpolation of satellite ocean altimetry maps. In *VISAPP.* Lisboa, Portugal.

Archambault, T., Filoche, A., Charantonnis, A., & Béréziat, D. (2024, February). Pre-training and fine-tuning attention based encoder decoder improves sea surface height multi-variate inpainting. In *VISAPP.* Rome, Italy.

Ardhuin, F., Ubelmann, C., Dibarboure, G., Gaultier, L., Ponte, A., Ballarotta, M., & Faugère, Y. (2020, November). *Reconstructing ocean surface current combining altimetry and future spaceborne doppler data.* Earth and Space Science Open Archive. doi: 10.1002/ESSOAR.10505014.1

Ballarotta, M., Ubelmann, C., Rogé, M., Fournier, F., Faugère, Y., Dibarboure, G., . . . Picot, N. (2020, September). Dynamic mapping of along-track ocean altimetry: Performance from real observations. *Journal of Atmospheric and Oceanic Technology*, *37*, 1593-1601. doi: 10.1175/JTECH-D-20-0030.1

Che, H., Niu, D., Zang, Z., Cao, Y., & Chen, X. (2022). ED-DRAP: Encoder–decoder deep residual attention prediction network for radar echoes. *IEEE Geoscience and Remote Sensing Letters*, *19*. doi: 10.1109/LGRS.2022.3141498

Chelton, D. B., Gaube, P., Schlax, M. G., Early, J. J., & Samelson, R. M. (2011, October). The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll. *Science*, *334*, 328–332. doi: 10.1126/SCIENCE.1208897

Chelton, D. B., Schlax, M. G., & Samelson, R. M. (2011, October). Global observations of nonlinear mesoscale eddies. *Progress in Oceanography*, *91*, 167-216. doi: 10.1016/J.POCEAN.2011.01.002

Chin, T. M., Vazquez-Cuervo, J., & Armstrong, E. M. (2017). A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sensing of Environment*, *200*, 154–169. doi: https://doi.org/10.1016/j.rse.2017.07.029

Ciani, D., Rio, M.-H., Bruno Nardelli, B., Etienne, H., & Santoleri, R. (2020, May). Improving the altimeter-derived surface currents using sea surface temperature (SST) data: A sensitivity study to SST products. *Remote Sensing*, *12*, 1601. doi: 10.3390/RS12101601

CLS/MEOM. (2020). *SWOT data challenge NATL60 [dataset].* CNES. doi: 10.24400/527896/A01-2020.002

CLS/MEOM. (2021). *Data challenge OSE - 2021A_SSH_MAPPING_OSE [dataset].* CNES. doi: 10.24400/527896/a01-2021.005

CMEMS. (2020). *Global ocean physics reanalysis [dataset].* Mercator Ocean International. doi: 10.48670/moi-00021

CMEMS. (2021). *Global ocean along-track L3 sea surface heights reprocessed (1993-ongoing) tailored for data assimilation [dataset].* Mercator Ocean International. doi: 10.48670/MOI-00146

CMEMS. (2023). *Global oceans sea surface temperature multi-sensor L3 observations [dataset].* Mercator Ocean International. doi: 10.48670/MOI-00164

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The operational sea surface temperature and sea ice analysis (OSTIA) system. *Advanced Along Track Scanning Radiometer (AATSR), Special Issue of Remote Sensing of Environment*, *116*, 140-158. doi: 10.1016/j.rse.2010.10.017

Emery, W. J., Brown, J., & Nowak, Z. P. (1989). AVHRR image navigation-summary and review. *Photogrammetric engineering and remote sensing*, *4*, 1175–1183.

Fablet, R., Amar, M., Febvre, Q., Beauchamp, M., & Chapron, B. (2021, June). End-to-end physics-informed representation learning for satellite ocean remote sensing data: Applications to satellite altimetry and sea surface currents. *IS-PRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *5*, 295-302. doi: 10.5194/ISPRS-ANNALS-V-3-2021-295-2021

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

Filoche, A., Archambault, T., Charantonis, A., & Béréziat, D. (2022). Statistics-free interpolation of ocean observations with deep spatio-temporal prior. In *ECML/PKDD Workshop on Machine Learning for Earth Observation and Prediction (MACLEAN)*.

Gaultier, L., Ubelmann, C., & Fu, L. (2016). The challenge of using future SWOT data for oceanic field reconstruction. *Journal of Atmospheric and Oceanic Technology*, *33*, 119-126. doi: 10.1175/JTECH-D-15-0160.1

González-Haro, C., Isern-Fontanet, J., Tandeo, P., & Garello, R. (2020). Ocean surface currents reconstruction: Spectral characterization of the transfer function between SST and SSH. *Journal of Geophysical Research: Oceans*, *125*(10), e2019JC015958. (e2019JC015958 10.1029/2019JC015958) doi: https://doi.org/10.1029/2019JC015958

Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., ... Hu, S.-M. (2021, November). Attention mechanisms in computer vision: A survey. *Computational Visual Media*, *8*, 331-368. doi: 10.1007/s41095-022-0271-y

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Conference on computer vision and pattern recognition (CVPR)* (pp. 770–778). doi: 10.1109/CVPR.2016.90

Hinton, G., & Dean, J. (2015). Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*.

Isern-Fontanet, J., Chapron, B., Lapeyre, G., & Klein, P. (2006). Potential use of microwave sea surface temperatures for the estimation of ocean currents. *Geophys. Res. Lett*, *33*, 24608. doi: 10.1029/2006GL027801

Jam, J., Kendrick, C., Walker, K., Drouard, V., Hsu, J., & Yap, M. (2021, February). A comprehensive review of past and present image inpainting methods. *Computer Vision and Image Understanding*, *203*. doi: 10.1016/J.CVIU.2020.103147

Jayne, S., & Marotzke, J. (2002, December). The oceanic eddy heat transport. *Journal of Physical Oceanography*, *32*, 3328-3345. doi: 10.1175/1520-0485(2002)032⟨3328:TOEHT⟩2.0.CO;2

Kahru, M., Di Lorenzo, E., Manzano-Sarabia, M., & Mitchell, B. G. (2012, March). Spatial and temporal statistics of sea surface temperature and chlorophyll fronts in the California Current. *Journal of Plankton Research*, *34*(9), 749-760. doi: 10.1093/plankt/fbs010

Kang, D., Curchitser, E. N., & Rosati, A. (2016). Seasonal variability of the Gulf Stream kinetic energy. *Journal of Physical Oceanography*, *46*(4), 1189 - 1207. doi: https://doi.org/10.1175/JPO-D-15-0235.1

Kingma, D. P., & Ba, J. (2017). *Adam: A method for stochastic optimization.*

Le Guillou, F., Metref, S., Cosme, E., Ubelmann, C., Ballarotta, M., Verron, J., & Le Sommer, J. (2020, October). *Mapping altimetry in the forthcoming SWOT era by back-and-forth nudging a one-layer quasi-geostrophic model.* Earth and Space Science Open Archive. doi: 10.1002/ESSOAR.10504575.1

Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D., ... others (2017). *Nemo ocean engine.*

Manucharyan, G., Siegelman, L., & Klein, P. (2020, January). A deep learning approach to spatiotemporal sea surface height interpolation and estimation of deep currents in geostrophic ocean turbulence. *Journal of Advances in Modeling Earth Systems*, *13*. doi: 10.1029/2019MS001965

Martin, S. (2014). *An introduction to ocean remote sensing* (2nd ed.). Cambridge University Press. doi: 10.1017/CBO9781139094368

Martin, S. A., Manucharyan, G. E., & 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). doi: 10.1029/2022MS003589

McCann, M., Jin, K., & Unser, M. (2017, November). Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, *34*, 85–95. doi: 10.1109/MSP.2017.2739299

Mkhinini, N., Coimbra, A. L. S., Stegner, A., Arsouze, T., Taupier-Letage, I., & Béranger, K. (2014, December). Long-lived mesoscale eddies in the eastern Mediterranean Sea: Analysis of 20 years of AVISO geostrophic velocities. *Journal of Geophysical Research: Oceans*, *119*, 8603–8626. doi: 10.1002/2014JC010176

Moschos, E., Schwander, O., Stegner, A., & Gallinari, P. (2020). Deep-SST-Eddies: A deep learning framework to detect oceanic eddies in sea surface temperature images. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (p. 4307-4311). doi: 10.1109/ICASSP40776.2020.9053909

Nardelli, B., Cavaliere, D., Charles, E., & Ciani, D. (2022, February). Super-resolving ocean dynamics from space with computer vision algorithms. *Remote Sensing*, *14*, 1159. doi: 10.3390/RS14051159

NASA/JPL. (2019). *GHRSST level 4 mur 0.25deg global foundation sea surface temperature analysis (v4.2) [dataset].* Mercator Ocean International. doi: 10.5067/GHM25-4FJ42

Ongie, O., Jalal, A., Metzler, C., Baraniuk, R., Dimakis, A., & Willett, R. (2020, May). Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, *1*, 39–56.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. doi: 10.1109/TKDE.2009.191

Qin, Z., Zeng, Q., Zong, Y., & Xu, F. (2021, September). Image inpainting based on deep learning: A review. *Displays*, *69*, 102028. doi: 10.1016/J.DISPLA.2021.102028

Stegner, A., Le Vu, B., Dumas, F., Ghannami, M., Nicolle, A., Durand, C., & Faugere, Y. (2021, September). Cyclone-anticyclone asymmetry of eddy detection on gridded altimetry product in the mediterranean sea. *Journal of Geophysical Research: Oceans*, *126*. doi: 10.1029/2021JC017475

Taburet, G., Sanchez-Roman, A., Ballarotta, M., Pujol, M.-I., Legeais, J.-F., Fournier, F., ... Dibarboure, G. (2019). DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Sci*, *15*, 1207-1224. Retrieved from `https://doi.org/10.5194/os-15-1207-2019` doi: 10.5194/os-15-1207-2019

Thiria, S., Sorror, C., Archambault, T., Charantonis, A., Béréziat, D., Mejia, C., ... Crepon, M. (2023). Downscaling of ocean fields by fusion of heterogeneous observations using deep learning algorithms. *Ocean Modeling*, *182*.

Ubelmann, C., Cornuelle, B., & Fu, L. (2016, August). Dynamic mapping of along-track ocean altimetry: Method and performance from observing system simulation experiments. *Journal of Atmospheric and Oceanic Technology*, *33*, 1691-1699. doi: 10.1175/JTECH-D-15-0163.1

Vu, B. L., Stegner, A., & Arsouze, T. (2018, April). Angular momentum eddy detection and tracking algorithm (AMEDA) and its application to coastal eddy formation. *Journal of Atmospheric and Oceanic Technology*, *35*, 739-762. doi: 10.1175/JTECH-D-17-0010.1

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *Computer Vision and Pattern Recognition*.

Zhai, X., Greatbatch, R. J., & Kohlmann, J.-D. (2008). On the seasonal variability of eddy kinetic energy in the Gulf Stream region. *Geophysical Research Letters*, *35*(24). doi: 10.1029/2008GL036412