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

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

[1]LIP6, Sorbonne University, 4 place Jussieu, Paris, France
[2]LOCEAN, Sorbonne University, 4 place Jussieu, Paris, France
[3]ENSIIE, Evry, France, LaMME
[4]INRIA Paris, Paris, France

Key Points:

- This is the preprint of a paper submitted to JAMES in September 2023
- 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, and able to use temperature data
- We find temperature, enhances sea surface height reconstruction, as well as the estimation of the surface currents and mesoscale eddies.

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, spaceborne altimetry provides valuable measurements of Sea Surface Height (SSH), which is used to estimate surface geostrophic currents. However, due to the sensor technology employed, important gaps occur in SSH observations. Complete SSH maps are produced by the altimetry community using linear Optimal Interpolations (OI) such as the widely-used Data Unification and Altimeter Combination System (DUACS). However, OI is known for producing overly smooth fields and thus misses some mesostructures and eddies. 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 design a realistic twin experiment to emulate the satellite observations of SSH and SST to evaluate interpolation methods. We introduce a deep learning network able to use SST information, and a trainable in two settings: one where we have no access to ground truth during training and one where it is accessible. Our investigation involves a comparative analysis of the aforementioned network when trained using either supervised or unsupervised loss functions. We assess the quality of SSH reconstructions and further evaluate the network's performance in terms of eddy detection and physical properties. We find that it is possible, even in an unsupervised setting to use SST to improve reconstruction performance compared to SST-agnostic interpolations. We compare our reconstructions to DUACS's and report a decrease of 41% in terms of root mean squared error.

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 very important 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 the quality of the height map, we propose to use 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 is acquired with a much 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, so we propose a methodology to train these methods using only the input information. We show the feasibility of these approaches, as well as the improvements brought by the temperature information.

1 Introduction

Since the first ocean remote sensing missions in the 1970s, satellite observation of the ocean 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 measures of various physical natures with wide spatial coverage that completed in situ datasets. Among these techniques, satellite altimetry is used to retrieve the Sea Surface Height (SSH) a determining variable of the ocean circulation. Indeed, SSH spatial gradient can be used to estimate geostrophic currents, i.e. the currents necessary for the Coriolis force to balance 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 important 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 the Data Unification and Altimeter Combination System (DUACS) (Taburet et al., 2019) which performs a linear Optimal Interpolation (OI) of the nadir along-track measures leveraging a covariance matrix tuned on 25 years of data. However several studies show that DUACS reconstruction misses some of the mesoscales structures and eddies (Amores et al., 2018; Stegner et al., 2021). As such, improving the reconstruction of a gridded altimetry product is still an open challenge.

In order 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 heat is transported by currents in an advection dynamic. 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 cloud coverage also introduces data gaps. 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, thereby unlocking the full potential of satellite oceanography products.

In the last decade, deep learning has emerged as one of the leading methods in computer vision, particularly to address image inverse problems. Neural networks have demonstrated remarkable flexibility in fusing observations from various sources and modalities, exhibiting their capacity to learn complex relationships given a sufficient number of training samples (McCann et al., 2017; Ongie et al., 2020). Prior work 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 such methods often requires the fully gridded ground truth to be trained, which is not possible in a realistic geoscientific scenario. To overcome this limitation two solutions were proposed: employing loss functions that do not rely on ground truth data or conducting a twin experiment on a simulation mimicking the inverse problem we try to solve (also called an Observing System Simulation Experiment). This last option has the advantage of allowing supervised training but suffers from the domain gap that might occur between the simulation and the real world. Notably, Fablet et al. (2021) performed an efficient supervised SSH interpolation on one year of OSSE data and extended their study using SST showing increased performance (Fablet et al., 2023). On the other hand, Archambault et al. (2023); S. A. Martin et al. (2023) trained a neural network using only observations.

However, as both these studies focused on real-world data, no fully gridded ground truth reference was available for an evaluation and interpretation of the results. In this work, we design a new OSSE framework including 20 years of SSH and SST simulated observations and their associated ground truth. As the previously existing OSSE (CLS/MEOM, 2020) provided only one year of data and no SST realistic instrumental error, this new dataset is closer to a realistic multi-variate observation of the ocean. Moreover, we present a novel Attention-Based Encoder-Decoder (ABED) framework to perform spatiotemporal interpolation of SSH fields. This network leverages along-track SSH measurements and, optionally, incorporates SST contextual data. In order to assess the feasibility of training ABED in a realistic setting, where no gridded ground truth is accessible, we propose to train it using solely along-track measures and compare it with its classically supervised version.

This paper is structured as follows, in Section 2 after giving a rationale for the inclusion of SST information in the interpolation method we detail our OSSE. In Section 3 we present our architecture and the training losses. In Section 4 we evaluate the inter-

polation in terms of SSH reconstruction, and oceanic 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 a different OSSE. In Section 5, we discuss the limitations and perspectives of this work.

2 Multi-variate data simulation

In the following, we provide a rationale for the SSH and SST connection, outline the reference data source we utilized (Global Ocean physics Reanalysis (CMEMS, 2020)), and detail our OSSE's SSH and SST observations.

2.1 Physical relationship between SSH and SST

One of the most important uses of SSH data is to recover oceanic currents through surface quasi-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 Coriolis force projection is null, it is a good estimation of the circulation. The surface geostrophic currents can be computed from the SSH h following:

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

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 rotation period of the earth, ϕ the latitude and g the gravitational acceleration.

In a first approximation, the surface temperature T can be considered as a passive tracker transported by surface currents. The evolution of a scalar in a static velocity field can be described by the linear advection equation.

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

Combining the geostrophic and the advection Equations (1,2), we see why a time series of SST observations should provide pertinent information for constraining the SSH reconstruction. However, the actual physical link between temperature and sea-surface height is more complex, as other phenomena must be considered, such as diffusion, convection, circulation between water depths, and viscosity. The satellite observations of both temperature and sea surface height also suffer from instrumental errors and are by nature limited to observing the surface of the ocean. 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

In order 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 ways to test reconstruction methods and errors (Amores et al., 2018; Stegner et al., 2021; Gaultier et al., 2016). 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 measures. 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. We detail hereafter the reference dataset of our OSSE and the observation operators of the two variables.

2.2.1 Base simulation

We conduct our experiences 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 observations from satellites (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 Mars 20, 2000 to December 29, 2019, for a total of 7194 days.

We select a portion of the Golf 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 Golf 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 can be learned by a neural network. 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 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 spatiotemporal 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 Copernicus 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 $\sigma = 1.9$ 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 measures of SSH inside each pixel so that it represents the mean of the daily measures from the different satellites (if any). As the GLORYS12 simulation assimilates SSH alongracks measurements, we introduce a delay between the L3 satellite observations and the simulation. Doing so, we ensure that simulated along-tracks are taken randomly and not specifically where the model assimilated real world observations.

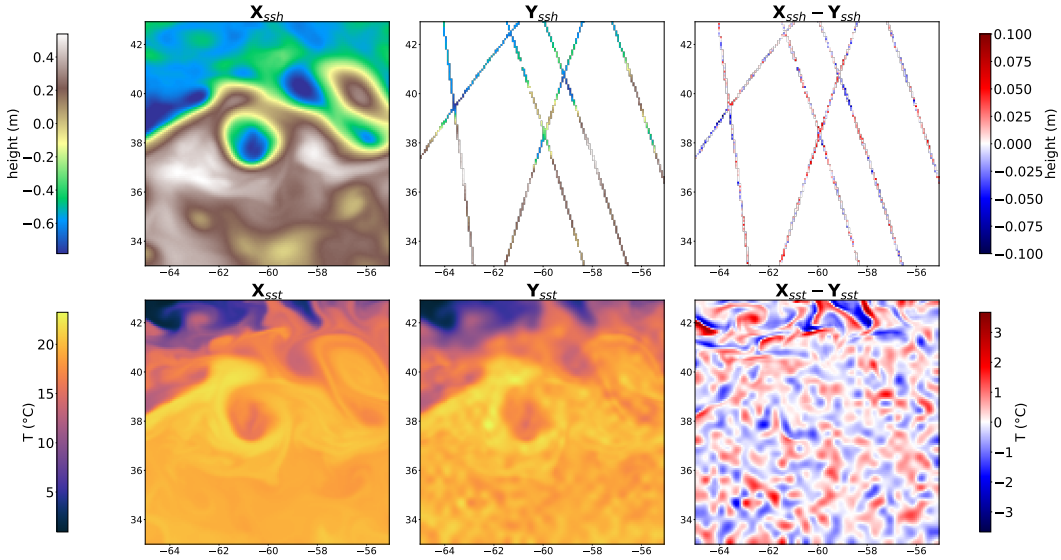

Figure 1. Images of the Ground Truth variables from GLORYS and the simulated satellite observations. The first row is the SSH variable, where the second row is the SST. The first column is the ground truth from GLORYS simulation, the second column is their associated satellite observation and the last is the difference.

2.2.3 SST observations

SST remote sensing is based on direct infrared measurements, leading to wider measurement swaths but making the measurement sensible to cloud cover. The so-called L3 satellite products have much higher data coverage, but no observation is possible when the cloud is too thick. To fill the gaps, the L3 products from several satellites are merged and interpolated to form the fully gridded image. This results in various resolutions in the same product, where high-resolution structures are artificially smoothed when the cloud coverage is too thick. To simulate this process, we use the mask of NRT L3 product (CMEMS, 2023) to retrieve a realistic cloud cover mask C (between 0 and 1) which we grid to the target resolution. The SST observation operator $\mathcal{H}^{sst}$ can then be written as:

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

where $\odot$ is the element-wise product, $\star$ the convolution product, ε is a white Gaussian noise image of size 32×32 linearly upsampled to a 128×128 image. We also use a spatial Gaussian filter ($\mathcal{G}_{0,\sigma}$) with $\sigma = 16$(km)) to simulate the smoothing of the interpolation performed by satellite products. Our SST observations thus present a spatially and temporally correlated noise, with different resolutions depending on cloud coverage. In the end, $\mathcal{H}^{sst}$ adds a noise with a standard deviation of 0.5 °C where the SST standard deviation of the ground truth is 4.96 °C. 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.

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, and 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; 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 time 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. An overview of the inputs and outputs of our method is provided in Figure 3. 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 $\mathbf{Y} = \left(\mathbf{Y}^{ssh}, \mathbf{Y}^{sst}\right)$ if the network uses SST. The length of the time window will be discussed in Section 4.1, and training losses of the network in Section 3.3.

3.2 Architecture

We propose an attention-based encoder-decoder (ABED) presented in Figure 2 to perform the interpolation over the time window. The attention mechanism allows to emphasize important features while neglecting irrelevant ones which makes it well-suited to extract information from contextual variables. It is widely used in many computer vision tasks (Guo et al., 2021) and can be transposed to geoscience applications. The overall structure of our neural network is inspired by the one used by Che et al. (2022) that introduced a residual Unet with attention layers for rain nowcasting.

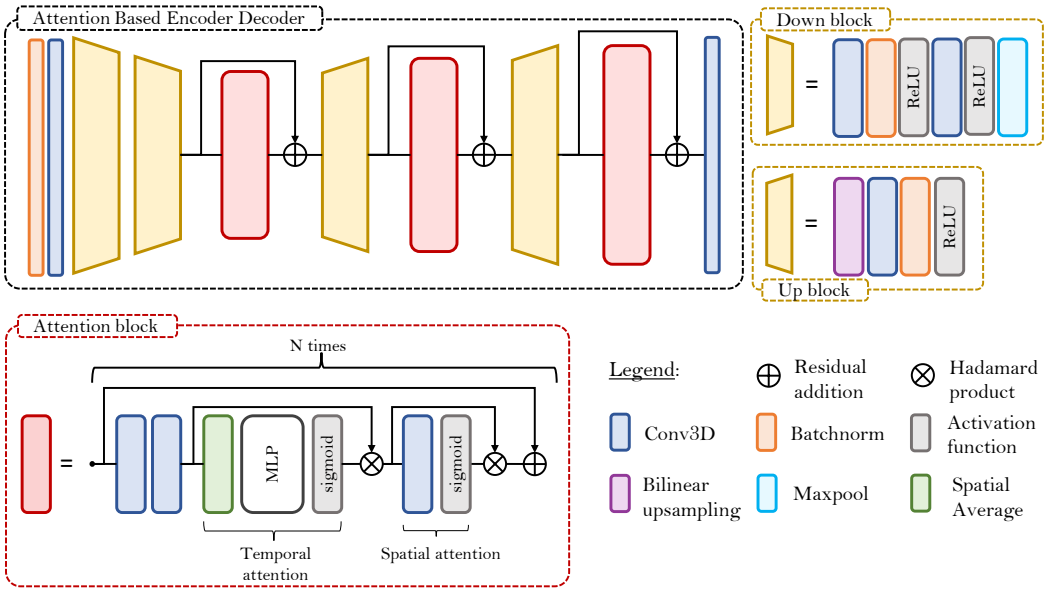

Figure 2. The architecture of the proposed Attention-Based Encoder Decoder (ABED) neural network. It is designed to take a time series of 21 images of SSH, with or without a time series of SST. The encoder divides the spatial dimensions of the images by 4 through 2 "down-block". Then the decoder uses an attention block to highlight relevant information in the images and progressively upscales it.

The encoder starts with a batch normalization and a 3D convolution (in time and in the two spatial dimensions) followed by two downsampling blocks that divide spatial dimensions by 2 (see Figure 2). The decoder is composed of residual attention blocks followed by upsampling blocks.

We describe hereafter our attention block which consists of two essential steps: temporal attention and spatial attention. 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 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 temporal length of the kernel size matches the length of the time series. 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/james_2023`.

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 observation operator $\mathcal{H}^{ssh}$ 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 overfit SSH images starting from SST only and constraining on SSH observations. Both these methods are fitted on one (or a small number) of examples and must therefore be refitted in order 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 :

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

- The MSE using only observations:

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

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

$$\begin{aligned} \mathcal{L}_{trili\_reg}(\mathbf{Y}^{ssh}, \hat{\mathbf{X}}^{ssh}) =& \mathcal{L}_{trili}(\mathbf{Y}^{ssh}, \hat{\mathbf{X}}^{ssh}) + \lambda_1 \frac{1}{N_1} \sum_{i} \left(\frac{\partial}{\partial s} \mathbf{Y}^{ssh}_{i} - \frac{\partial}{\partial s} \mathcal{H}^{ssh}(\hat{\mathbf{X}}^{ssh})_i \right)^2 \\ &+ \lambda_2 \frac{1}{N_2} \sum_{i} \left(\frac{\partial^2}{\partial s^2} \mathbf{Y}^{ssh}_{i} - \frac{\partial^2}{\partial s^2} \mathcal{H}^{ssh}(\hat{\mathbf{X}}^{ssh})_i \right)^2 \end{aligned} \tag{7}$$

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 measures of SSH, and SSH first and second along-track spatial derivative respectively. We take $\lambda_1 = \lambda_2 = 0.05$ the regularization coefficients, the same values used by S. A. Martin et al. (2023).

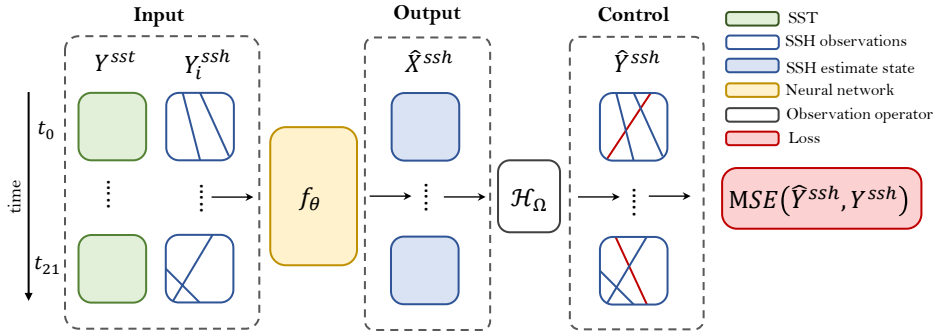

Figure 3. Computational graph of the proposed unsupervised interpolation method. The input of the neural network is a 21-day time series of SSH satellite observations, excluding data from a single satellite, and optionally includes SST measurements. The network estimates a time series of SSH fields states, 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.

The losses $\mathcal{L}_{trili}$ and $\mathcal{L}_{trili\_reg}$ 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, we remove the measure from one satellite from the input of the network but calculate the loss function on all satellite observations. Doing so, the network must generalize outside the along-track measure that was given as input. In Figure 3 we call $\mathbf{Y}_i^{ssh}$ the input observations and present an unsupervised inversion computational graph.

3.4 Training details

Train, validation, test split. We partitioned the dataset into three subsets: training, validation, and test data. We used the year 2017 exclusively for testing our reconstructions (every analysis conducted in the following was performed on this data). We validate our methods on three distinct time intervals: (1) from July 14, 2002, to July 28, 2003, (2) from January 5, 2008, to January 18, 2009, and (3) from June 28, 2013, to July 13, 2014. The remaining data was used for training, with the exception of a 15-day period set aside to prevent data leakage.

Normalization. We normalize both the input and output of the artificial network. This involves subtracting the mean and dividing by the standard deviation, which are both computed on the entire training dataset. Specifically, for images related to SSH measurements along tracks, we first perform this normalization and subsequently replace any missing values with zeros. We normalize the neural network SSH outputs with the statistics computed on the input observations (in order that the method remains applicable in an unsupervised setting).

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 measures. Doing so, the stopping strategy is still compliant with a situation where no ground truth is accessible.

Ensemble. As neural network optimization is sensible 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. Performing an ensemble estimation helps to stabilize performances, and even enhance 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 the following, we present the scores of the different reconstruction methods on the test set. In contrast to the training and validation method, we assess the quality of the reconstruction on the gridded ground truth. We compare the fields estimated by the 3 losses $\mathcal{L}$, $\mathcal{L}_{tri}$ and $\mathcal{L}_{tri\_reg}$ on 3 different sets of input data: one with only SSH tracks, one with SSH and the noised SST denoted nSST, and one with the noise-free SST of the GLORYS assimilation. We train interpolation methods on noise-free SST to provide an upper-bound performance of the neural network in the case of a perfect physical link between SSH and SST.

4.1 SSH reconstruction and quality of derived geostrophic currents

We give the RMSE of the SSH estimates fields in Table 1, and the RMSE on the velocity fields in Table 2. As expected, the supervised loss function outperforms the unsupervised framework in every data scenario. Specifically, in the SSH+SST scenario, the supervised loss decreases the RMSE of $\mathcal{L}_{tri}$ by 24%, and 8% without SST. Also, adding SST as an additional input to the network generally leads to improved 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 RMSE is decreased by 30% and 23% for SST and nSST respectively with $\mathcal{L}$. The regularization introduced by (S. A. Martin et al., 2023) slightly increases reconstruction but is still close to the unregularized inversion.

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 $6.57\,{}^{\mathrm{cm}}/_{\mathrm{s}}$ for u and 6.14 for v on this data.

In Figure 4, 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 over late Summer. Specifically, in August and September, all methods are performing worse than in Winter which can be explained by the high energy of the Ocean at this period.

An important challenge of ocean satellite products is providing 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-days temporal window

Loss	SSH	SSH+nSST	SSH+SST
$\mathcal{L}$ (supervised)	4.18 — 3.85	3.23 — 2.93	**2.92** — **2.59**
$\mathcal{L}_{tri}$	4.52 — 4.16	3.86 — 3.51	3.62 — 3.24
$\mathcal{L}_{tri\_reg}$	4.38 — 4.13	3.73 — 3.48	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.

Loss	SSH		SSH+nSST		SSH+SST	
	u	v	u	v	u	v
$\mathcal{L}^*$	13.0	14.1	10.9	11.7	**10.1**	**10.6**
$\mathcal{L}_{tri}$	13.3	15.7	12.1	14.2	11.3	13.4
$\mathcal{L}_{tri\_reg}$	12.9	14.3	11.8	12.9	11.1	12.1

*supervised

Table 2. Eastward (u) and northward (v) surface currents in cm/s. The currents were estimated by applying the geostrophy approximation (see Equation1) on the SSH ensemble estimation of the 3 ABED networks.

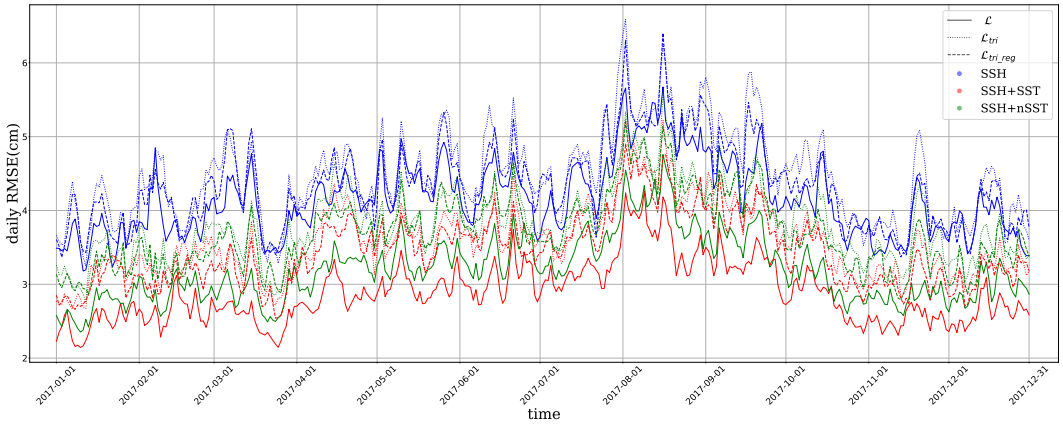

Figure 4. RMSE of the different reconstructions during the test year (2017)

as in Figure 5. The central image is a 10-day Delayed Time reconstruction as we need images of observations 10 days in the future. In Figure 5 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). To accurately produce an NRT image and even forecast, different training methods should

be tested such as centering the target time window in the future compared to observations.

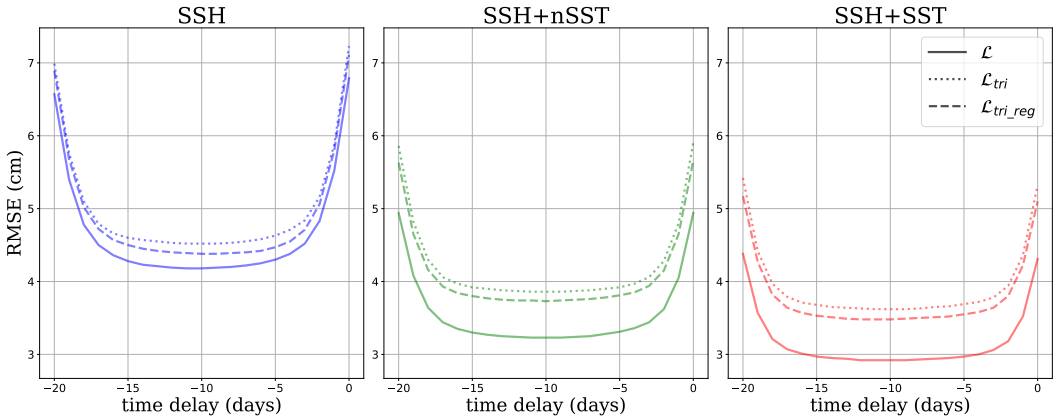

Figure 5. RMSE of the different reconstructions along the time window. The errors at a time delay of -20 correspond to an anti-causal scheme (knowing only future observations) whereas $timedelay = 0$ corresponds to a causal scheme (knowing no future observations). Knowing both past and future observations leads to the optimal reconstruction at $timedelay = -10$.

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, 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 present 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 are able to 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:

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

where P_i is the point of the grid where we compute the LNAM, P_j is a neighbor point of the grid, $\overrightarrow{P_iP_j}$ 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), $LNAM(P_i)$ 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 $LNAM(P_i)$ 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).

AMEDA finds potential eddy centers by searching for the local extrema of the $LNAM$ field. The shapes of the eddies are then defined by following closed current streamlines (either taking the last closed streamline, or the maximum velocity one). We perform the AMEDA algorithm on the geostrophic velocity field of our estimation and on the ground truth currents. An eddy is said to be detected if its ground truth barycenter is inside the closed streamline of its estimation.

4.2.3 Eddy detection performances

We present hereafter 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 measures the trust that we can put in the detected eddies (a precision of 1 means that all eddies in the simulation were also present in the ground truth). To aggregate the recall and the 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 positive.

Loss	SSH			SSH+nSST			SSH+SST		
	F_1	recall	precision	F_1	recall	precision	F_1	recall	precision
$\mathcal{L}$ (supervised)	0.719	0.617	0.86	0.765	0.685	0.866	**0.785**	**0.728**	0.852
$\mathcal{L}_{tri}$	0.704	0.647	0.771	0.727	0.672	0.79	0.739	0.692	0.793
$\mathcal{L}_{tri\_reg}$	0.714	0.609	0.863	0.725	0.623	0.865	0.742	0.644	**0.877**

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 $F1$ score.

Data comparison. As expected, no matter which loss we consider, the detection method using noise-free temperature 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 has a substantial impact on 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, adding a smoothness constraint on the SSH gradient field prevents the neural network from generating false eddies, but also prevents it from retrieving some structures.

4.2.4 Physical properties of detected eddies

In order 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 6 shows the detection performances as a function of some key parameters such as the 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 trilinear 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 accuracy of the model to estimate the physical properties of the eddies. To this end, we focus on the eddies that were successfully detected by all the methods (4881 eddies out of the 7908 eddies in the ground truth). We compare the physical parameters of the estimated eddies to their values in the corresponding true eddy. To do so, we compute the RMSE and bias of the following parameters: the maximum radius and velocity, and the average distance between the centers of estimate and true eddies. The error RMSE tells us if the eddies are well resolved, whereas the bias (estimate parameter minus ground truth parameter) tells us if the interpolation method has a global tendency to overestimate or underestimate some characteristics of the eddies.

Once again, Tables 4, 5 and 6 show that SST helps to estimate eddies radius, velocity, and position. 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. It could be interesting to know if the estimated eddies could be unbiased without decreasing the other performances.

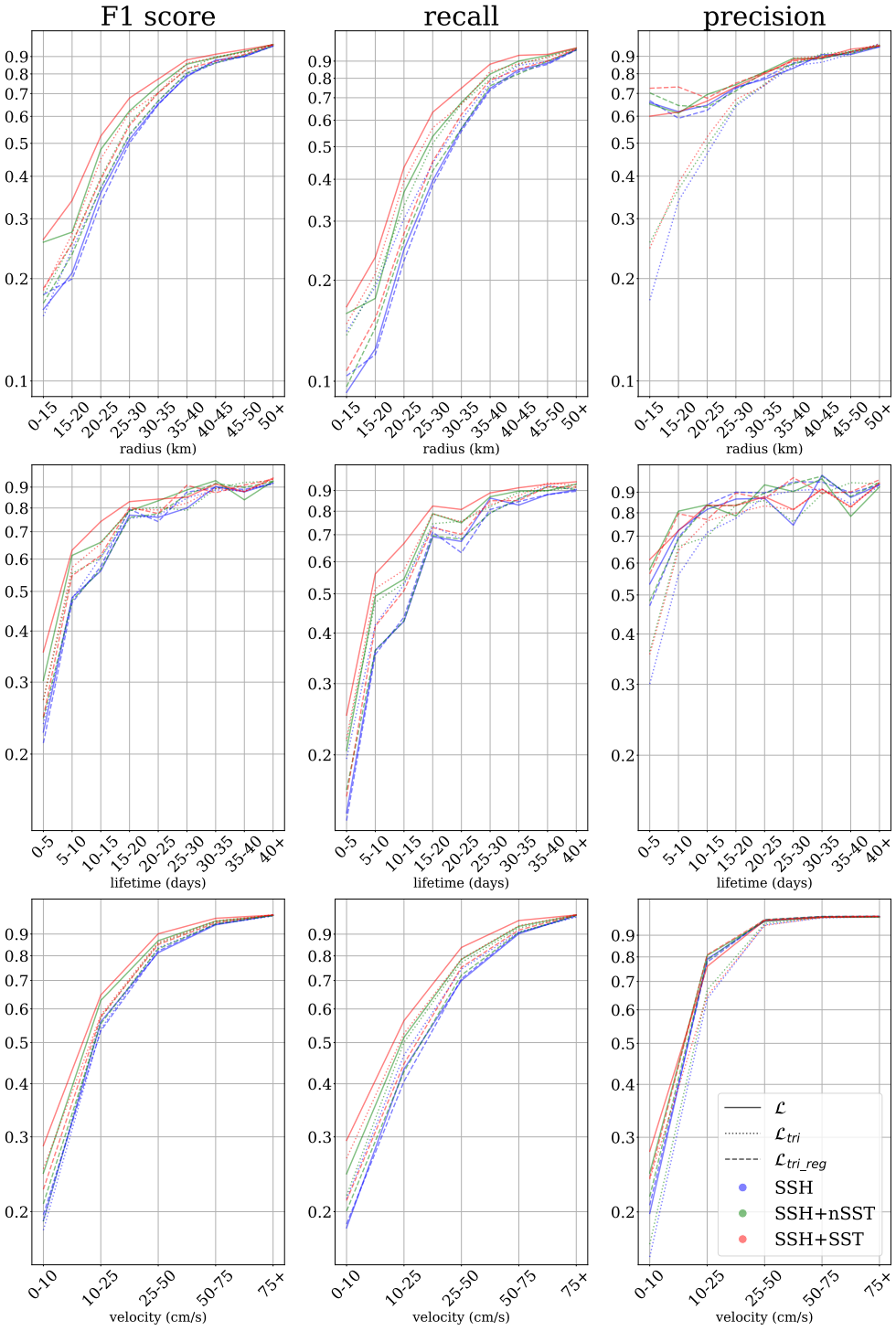

Figure 6. Detection scores of the different methods on eddies separated by radius (first row), lifetime (second row), and maximum velocity (last row). The considered scores are F_1 (first column), recall (second column), and precision (third column). The recall tells the proportion of actual positive instances that were correctly identified, the precision measures the trust that we can put in the detected eddies, and the F_1 score aggregates these two values.

Loss	SSH		SSH+nSST		SSH+SST	
	RMSE	bias	RMSE	bias	RMSE	bias
$\mathcal{L}$ (supervised)	21.8	3.8	20.5	4.5	**19.0**	3.8
$\mathcal{L}_{tri}$	21.5	**1.6**	20.4	2.2	20.8	1.7
$\mathcal{L}_{tri\_reg}$	22.4	2.9	21.7	3.3	21.2	3.9

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
$\mathcal{L}$ (supervised)	12.9	-6.2	11.0	-3.8	**10.2**	**-2.4**
$\mathcal{L}_{tri}$	12.7	-5.3	11.9	-4.5	12.2	-3.8
$\mathcal{L}_{tri\_reg}$	13.8	-8.0	13.0	-6.9	12.1	-5.8

Table 5. Eddies maximum velocity RMSE and bias (cm/s)

Loss	SSH	SSH+nSST	SSH+SST
	RMSE	RMSE	RMSE
$\mathcal{L}$ (supervised)	23.2	21.7	**20.3**
$\mathcal{L}_{tri}$	24.5	23.5	22.9
$\mathcal{L}_{tri\_reg}$	23.4	23.2	22.2

Table 6. Eddies center position RMSE (km)

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

We are interested in comparing our estimations to state-of-the-art methods for gridding SSH maps. To this end, the Ocean Data Challenge 2020 (CLS/MEOM, 2020) provides a similar OSSE to the one we used, as well as the interpolations of several methods. The studied area is the same, and the included data are SSH, SST, surface currents and the simulated along tracks measures. However, the ground truth used 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 GLORYS, and without assimilation. Also, this simulation was run for only one year, which makes it difficult to train neural networks, this is why we designed our own OSSE. The state-of-the-art framework presented in this challenge is 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 surface 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 surface quasi-geostrophic model.

- 4DVarNet (Fablet et al., 2021): introduced 4DVarNet, a supervised deep learning framework. In this configuration, it only takes SSH observations as input.
- MUSTI (Archambault et al., 2023): an unsupervised neural network that overfits SSH along tracks observations starting from an SST image. But this method must be refitted to new observations.

To produce our own estimation, we regrid the provided data to our resolution and perform the interpolation on this dataset without any finetuning or retraining. We evaluate all methods on 42 days of simulation, (between October 22nd and December 2nd 2012) which was the test set defined by the challenge. Each method is then evaluated using the following metrics, and we sum up the results in Table 7:

- μ and σ_t (in cm), are respectively the RMSE of the SSH and the temporal standard deviation of this RMSE.
- 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 geostrophic currents of the ground truth and the one 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
SSH $\mathcal{L}$	✗	✓	3.75	2.0	1.21	8.74	13.3	13.5
SSH $\mathcal{L}_{tri}$	✗	✗	4.06	2.19	1.32	9.29	13.7	15.1
SSH $\mathcal{L}_{tri\_reg}$	✗	✗	4.23	2.36	1.24	9.98	13.8	14.2
SSH+SST $\mathcal{L}$	✓	✓	**2.88**	**1.24**	0.95	4.51	**11.4**	**11.4**
SSH+SST $\mathcal{L}_{tri}$	✓	✗	3.08	1.41	1.18	5.18	11.8	12.8
SSH+SST $\mathcal{L}_{tri\_reg}$	✓	✗	3.39	1.65	1.18	5.7	12.4	12.3

Table 7. Comparison of the state-of-the-art reconstruction methods on a 43-day OSSE. SST stands for whether or not the reconstruction methods are using SST, and SUP stands for whether or not the methods are supervised.

We clearly see in thes scores a predominance of neural network-based methods (MUSTI, 4DVarNet and ours) as the importance of the SST in the reconstruction (MUSTI, and ours). This analysis highlights the interest in using deep learning-based methods for these inverse problems, as 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 cm/s for u and 5.4 cm/s for v (35% and 34% error reduction).

5 Conclusion and perspectives

5.1 Summary

Throughout this study, we show promising results for a neural interpolation of SSH tracks, even while training without fully gridded data. Leveraging an Observing System Simulation Experiment, we trained an attention-based auto-encoder neural network, with 3 different loss functions (2 of them learning the reconstruction without ground truth), and using 3 sets of data (SSH only, SSH and noised SST, SSH, and SST). We show a systematic improvement of the interpolation thanks to the use of SST as well for the SSH itself, but also for the reconstruction of currents and the detection of eddies. 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 the 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, but also that the key physical properties of the detected eddies such as size, speed, or center position are better rendered by SST-using methods. Furthermore, when it comes to 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 regularization process. However, they demonstrated higher precision scores, implying increased confidence in the eddies that were successfully detected.

We conducted an evaluation of our model's performance by comparing it with state-of-the-art interpolation techniques produced during the Ocean Data Challenge 2020 which provided diverse input data, ground truth measurements, and mappings derived from different methods. Remarkably, our approach exhibits superior performance even without retraining the neural network on this novel dataset, thanks to the integration of SST information. Notably, the utilization of SST led to a substantial enhancement of 41% in terms of RMSE for SSH when 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.

5.2 Perspectives

Transfer to real-world data. In this work, we performed SSH interpolation on an OSSE, which allows us to evaluate our methods on an idealized simulation of satellite tracks. The OSSE has the advantage of being more interpretable than real-world data, as the ground truth is accessible for performance evaluation, but is less realistic, especially in the simulation of SST noise which is hard to model. While simulations provide valuable insights, their direct application to real-world observations can be challenging due to inherent differences between the two domains. However, by employing transfer learning, we expect to leverage the knowledge gained from simulations to enhance reconstruction algorithms for real-world data. In further work, we are interested in comparing strategies able to learn directly from observations such as (Archambault et al., 2023; S. A. Martin et al., 2023), and a network that would benefit from supervised pre-training on simulation, and transfer on real-world observations. For instance, a model supervised on a simulation can be adapted to the new domain using one of the unsuper-

vised losses used in this study. This will require accurate modeling of the SST input noise and an adapted transfer strategy in order for the pre-training to be efficient.

Global interpolation. Furthermore, to get toward a global gridded SSH product, many challenges still need to be addressed. 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 should 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, in our investigation, we employed noisy yet complete SST data, 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 measures might improve the reconstruction, such as chlorophyll maps that track plankton advected by currents (Kahru et al., 2012).

Data availability statement

The GLORYS 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) measures used to retrieve along tracks coordinates and the L3 SST measures(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+.

The preprocessed data and the weights of our neural networks are available here: `https://doi.org/10.5281/zenodo.8380281` and our code is hosted on the following repository: `https://gitlab.lip6.fr/archambault/james_2023`

Acknowledgments

This research was supported by the grant for T.Archambault PhD from Sorbonne Université. The authors acknowledge the AMPHITRITE team, Alexandre Stegner, Briac Le Vu, and Evangelos Moschos, for their useful advice and assistance with the OSSE design and the automatic eddy detection algorithm AMEDA.

6 Annexes

6.1 Along-track spatial derivative

To calculate the first and second spatial derivatives of the SSH along the ground tracks which we use to regularize the network (see Section 3.3). Given $\mathbf{Y}^{ssh}$, we approx-

imate the derivative by the rate of change of the SSH:

$$\frac{\partial}{\partial s}\mathbf{Y}_i^{ssh} \simeq \frac{\mathbf{Y}_{i+1}^{ssh} - \mathbf{Y}_i^{ssh}}{\triangle s} \tag{9}$$

$$\frac{\partial^2}{\partial s^2}\mathbf{Y}_i^{ssh} \simeq \frac{\frac{\partial}{\partial s}\mathbf{Y}_{i+1}^{ssh} - \frac{\partial}{\partial s}\mathbf{Y}_i^{ssh}}{\triangle s} \tag{10}$$

where $\triangle s$ is the ground distance between the measures. We only compute the spatial derivatives from observations coming from the same satellite and only if the measures 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 Detection plot

In the following, we present the Ensemble interpolation of the methods for three days. To select which day to look at, we computed the daily error of the SSH method and the SSH+SST error. The first day is chosen so that the gap between these two errors is maximal (the SST method performs a lot better than the SSH-only method). The second day corresponds to the median error gap, and the last to the minimal error gap. With the reconstructions of every method and the corresponding SSH ground truth, we also provide the SST and the noised SST. To highlight the eddy signatures in SST data we plot the ground truth eddy computed on SSH on SST images as well.

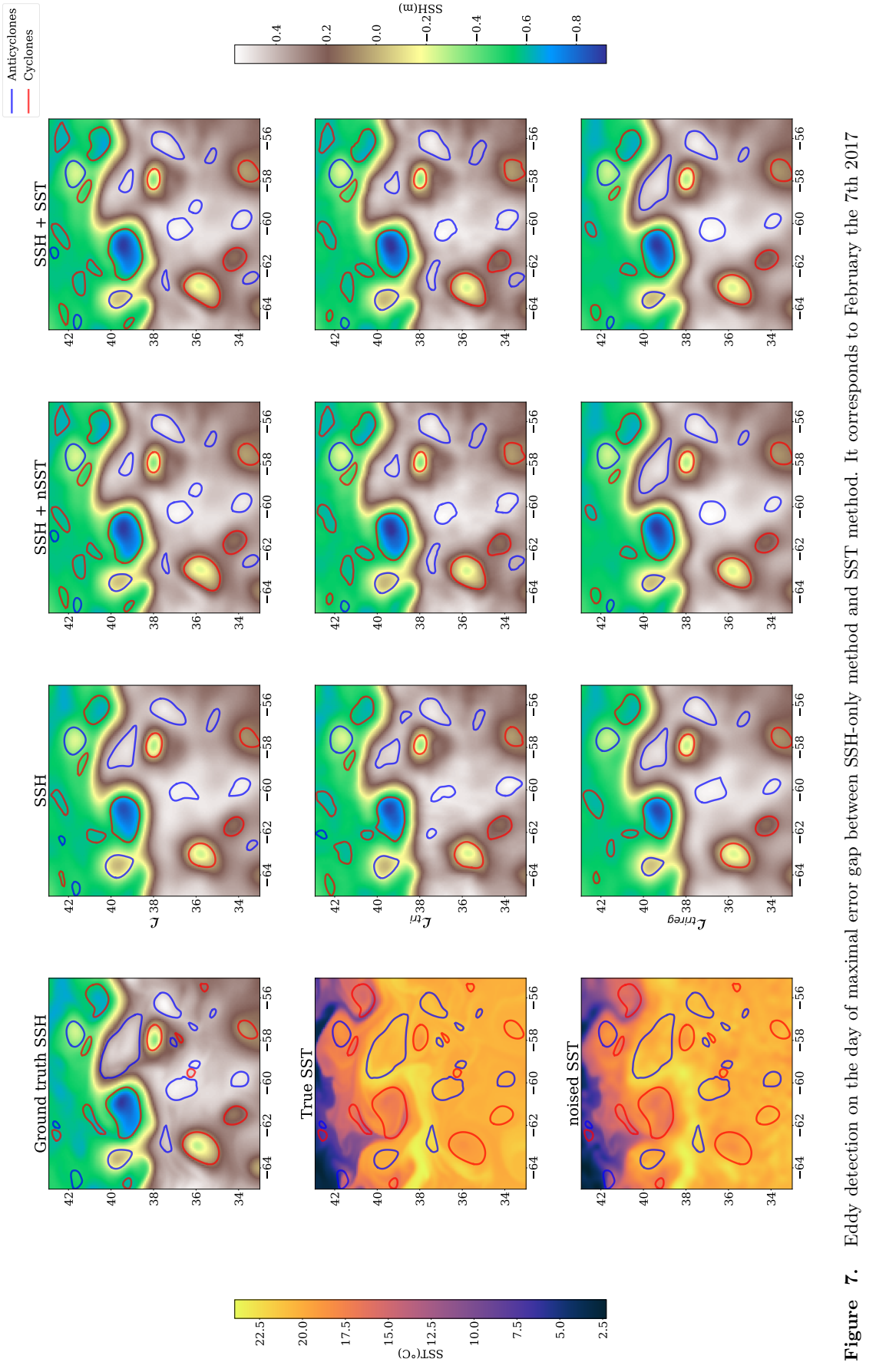

Figure 7. Eddy detection on the day of maximal error gap between SSH-only method and SST method. It corresponds to February the 7th 2017

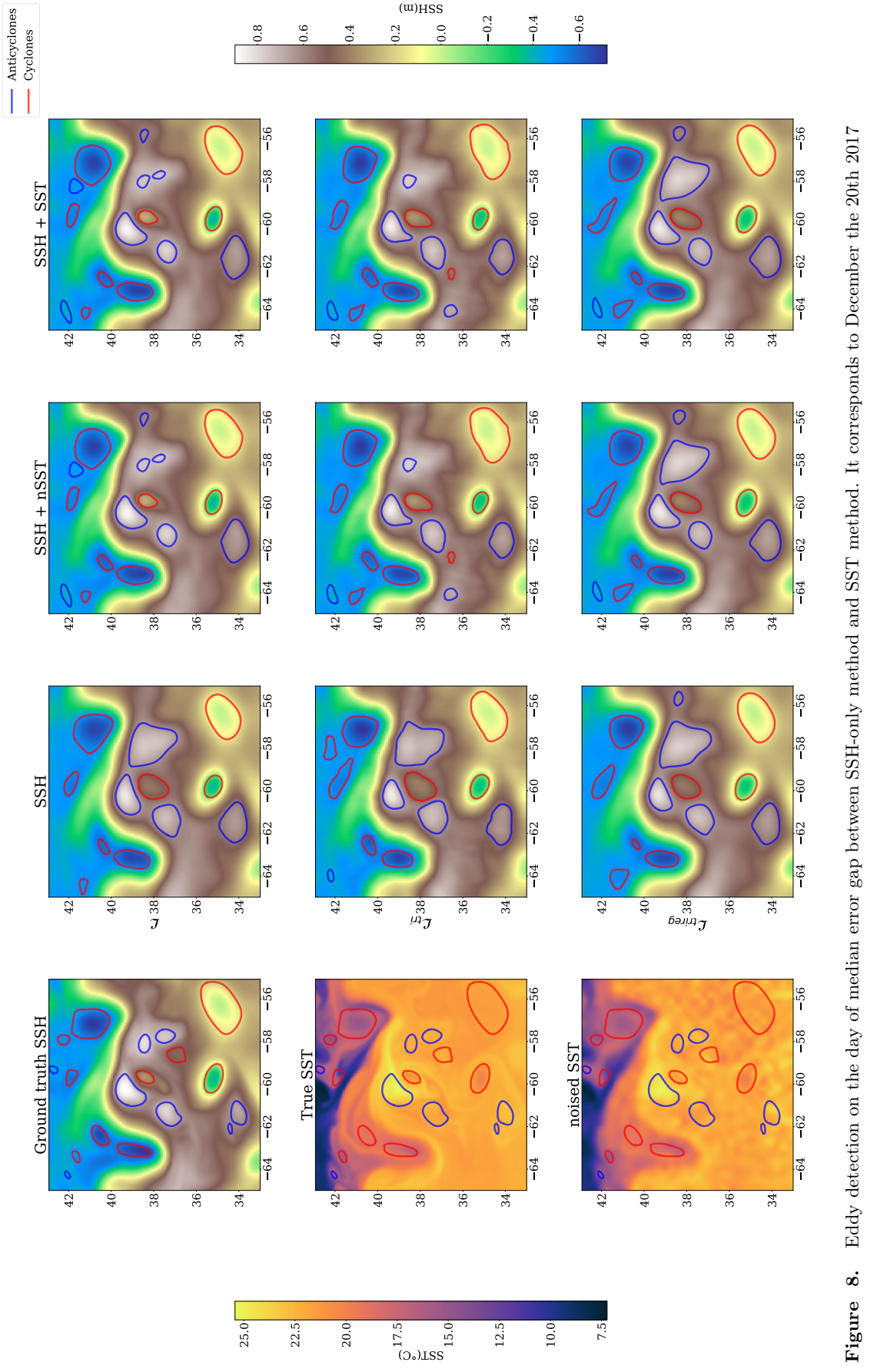

Figure 8. Eddy detection on the day of median error gap between SSH-only method and SST method. It corresponds to December the 20th 2017

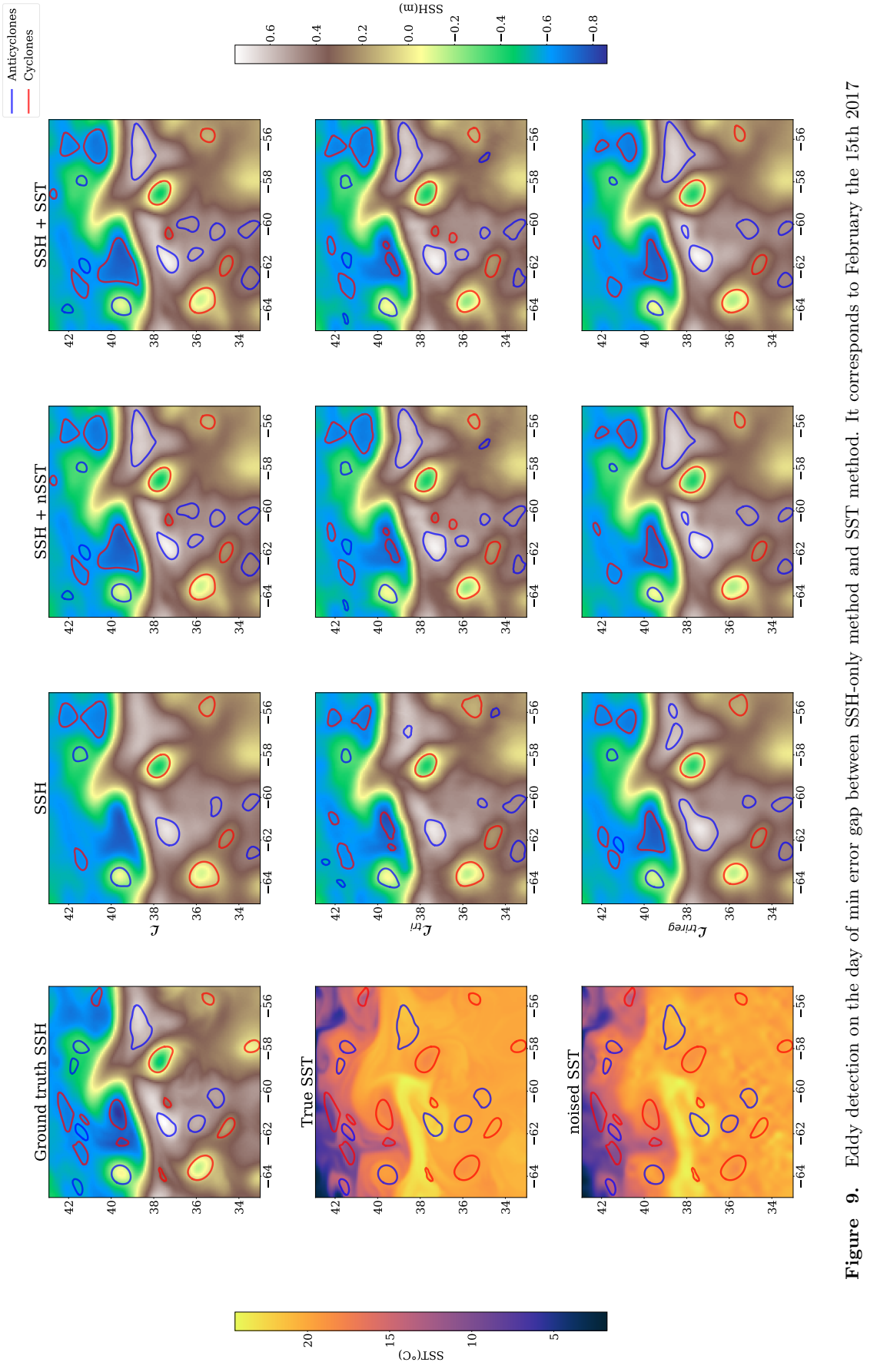

Figure 9. Eddy detection on the day of min error gap between SSH-only method and SST method. It corresponds to February the 15th 2017

References

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. Retrieved from `https://www.essoar.org/doi/abs/10.1002/essoar.10501076.1` 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, 1). 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. Retrieved from `https://hal.sorbonne-universite.fr/hal-03934647`

Ardhuin, F., Ubelmann, C., Dibarboure, G., Gaultier, L., Ponte, A., Ballarotta, M., & Faugère, Y. (2020, 11). *Reconstructing ocean surface current combining altimetry and future spaceborne doppler data.* Earth and Space Science Open Archive. Retrieved from `http://www.essoar.org/doi/10.1002/essoar.10505014.1` doi: 10.1002/ESSOAR.10505014.1

Ballarotta, M., Ubelmann, C., Rogé, M., Fournier, F., Faugère, Y., Dibarboure, G., ... Picot, N. (2020, 9). Dynamic mapping of along-track ocean altimetry: Performance from real observations. *Journal of Atmospheric and Oceanic Technology*, *37*, 1593-1601. Retrieved from `https://journals.ametsoc.org/view/journals/atot/37/9/jtechD200030.xml` 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, 10). 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, 10). Global observations of nonlinear mesoscale eddies. *Progress in Oceanography*, *91*, 167-216. doi: 10.1016/J.POCEAN.2011.01.002

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. Retrieved from `https://www.mdpi.com/2072-4292/12/10/1601/htmhttps://www.mdpi.com/2072-4292/12/10/1601` doi: 10.3390/RS12101601

CLS/MEOM. (2020). *Swot data challenge natl60 [dataset].* CNES. doi: https://doi.org/10.24400/527896/A01-2020.002

CMEMS. (2020). *Global ocean physics reanalysis [dataset].* Mercator Ocean International. doi: https://doi.org/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: https://doi.org/10.48670/MOI-00146

CMEMS. (2023). *Global oceans sea surface temperature multi-sensor l3 observations [dataset].* Mercator Ocean International. doi: https://doi.org/10.48670/MOI-00164

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. *ISPRS 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).* Retrieved from `https://hal.sorbonne-universite.fr/hal-03765735`

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

Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., ... Hu, S.-M. (2021, 11). Attention mechanisms in computer vision: A survey. *Computational Visual Media*, *8*, 331-368. Retrieved from `http://arxiv.org/abs/2111.07624http://dx.doi.org/10.1007/s41095-022-0271-y` doi: 10.1007/s41095-022-0271-y

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. Retrieved from `http://www.ecco-group.org` 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, 12). 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, 03). Spatial and temporal statistics of sea surface temperature and chlorophyll fronts in the California Current. *Journal of Plankton Research*, *34*(9), 749-760. Retrieved from `https://doi.org/10.1093/plankt/fbs010` doi: 10.1093/plankt/fbs010

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, 10). *Mapping altimetry in the forthcoming SWOT era by back-and-forth nudging a one-layer quasi-geostrophic model.* Earth and Space Science Open Archive. Retrieved from `http://www.essoar.org/doi/10.1002/essoar.10504575.1` 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.*

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), e2022MS003589. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003589` (e2022MS003589 2022MS003589) doi: https://doi.org/10.1029/

2022MS003589

McCann, M., Jin, K., & Unser, M. (2017, 11). Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, *34*, 85–95. Retrieved from http://arxiv.org/abs/1710.04011http://dx.doi.org/10.1109/MSP.2017.2739299 doi: 10.1109/MSP.2017.2739299

Mkhinini, N., Coimbra, A. L. S., Stegner, A., Arsouze, T., Taupier-Letage, I., & Béranger, K. (2014, 12). 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. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1002/2014JC010176 doi: 10.1002/2014JC010176

Nardelli, B., Cavaliere, D., Charles, E., & Ciani, D. (2022, February). Super-resolving ocean dynamics from space with computer vision algorithms. *Remote Sensing*, *14*, 1159. Retrieved from https://www.mdpi.com/2072-4292/14/5/1159/htmhttps://www.mdpi.com/2072-4292/14/5/1159 doi: 10.3390/RS14051159

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.

Qin, Z., Zeng, Q., Zong, Y., & Xu, F. (2021, 9). 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*.

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. Retrieved from https://journals.ametsoc.org/view/journals/atot/33/8/jtech-d-15-0163_1.xml doi: 10.1175/JTECH-D-15-0163.1

Vu, B. L., Stegner, A., & Arsouze, T. (2018, 4). Angular momentum eddy detection and tracking algorithm (ameda) and its application to coastal eddy formation. *Journal of Atmospheric and Oceanic Technology*, *35*, 739-762. Retrieved from https://journals.ametsoc.org/view/journals/atot/35/4/jtech-d-17-0010.1.xml 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*.