Pre-training and Fine-tuning Attention Based Encoder Decoder Improves Sea Surface Height Multi-variate Inpainting

Théo Archambault, Arthur Filoche, Anastase Charantonis, Dominique Béréziat

Pre-training and Fine-tuning Attention Based Encoder Decoder Improves Sea Surface Height Multi-variate Inpainting

Théo Archambault*[1,2][a], Arthur Filoche[1,3][b], Anastase Charantonis[4,5][c] and Dominique Béréziat[1][d]

[1]*Sorbonne Université, CNRS, LIP6, Paris, France*

[2]*Sorbonne Université, LOCEAN, Paris, France*

[3]*UWA, Perth, Australia*

[4]*ENSIIE, LaMME, Evry, France*

[5]*Inria, Paris, France*

**Corresponding author: theo.archambault@lip6.fr*

Keywords: Image inverse problems, Deep neural network, Spatiotemporal inpainting, Multi-variate observations, Transfer learning, Satellite remote sensing

Abstract: The ocean is observed through satellites measuring physical data of various natures. Among them, Sea Surface Height (SSH) and Sea Surface Temperature (SST) are physically linked data involving different remote sensing technologies and therefore different image inverse problems. In this work, we propose to use an Attention-based Encoder-Decoder to perform the inpainting of the SSH, using the SST as contextual information. We propose to pre-train this neural network on a realistic twin experiment of the observing system and to fine-tune it in an unsupervised manner on real-world observations. We show the interest of this strategy by comparing it to existing methods. Our training methodology achieves state-of-the-art performances, and we report a decrease of 25% in error compared to the most widely used interpolations product.

1 INTRODUCTION

In the past decades, satellite remote sensing produced an unprecedented amount of data, which led to a deeper understanding of the Earth system. For instance, out of the 50 essential climate variables defined by the Global Climate Observing System (GCOS) 26 are estimated through satellite (Yang et al., 2013). In the field of oceanography, satellites are used to measure various ocean surface variables, such as height, temperature, ice fraction, or chlorophyll concentration. The nature of sea-surface satellite observations requires solving various image inverse problems, such as inpainting, super-resolution, denoising, etc.

In this study, we focus on the inpainting of the Sea Surface Height (SSH), which is a very important variable of the ocean state, as it is used to retrieve surface currents through the geostrophic approximation. The altimeters embarked in satellites measure their distance to the sea surface through the return time of a radar pulse. Because of this technique, the nadir-pointing altimeters sensors are only able to take measures vertically, along their ground tracks (Martin, 2014). Therefore, producing a fully grided map of the SSH is a challenging spatiotemporal inpainting problem. It is currently tackled by the Data Unification and Altimeter Combination System, DUACS, (Taburet et al., 2019) which is a linear optimal interpolation of the along-track data from several satellites (Bretherton et al., 1976). However, prior works show that DUACS produces overly smooth maps, and is missing a lot of small structures and eddies (Amores et al., 2018; Stegner et al., 2021). To enhance the quality of this reconstruction, we are interested in exploiting contextual variables physically related to SSH, and with a similar or finer resolution. Among different possibilities, Sea Surface Temperature (SST) is linked to surface currents, as the heat is passively transported by oceanic circulation, and is acquired through direct infrared measures leading to images with higher spatiotemporal sampling.

In the last years, deep learning has emerged as one of the leading methods to solve image inverse prob-

[a] https://orcid.org/0000-0001-8051-0534

[b] https://orcid.org/0000-0001-7779-6105

[c] https://orcid.org/0000-0003-4953-2684

[d] https://orcid.org/0000-0003-1444-8212

lems (McCann et al., 2017) and specifically inpainting problems (Jam et al., 2021; Qin et al., 2021). Their flexibility allows neural networks to include contextual information such as SST, and several studies conclude that using it in a multi-variate neural network leads to a significant improvement of the SSH reconstruction (Nardelli et al., 2022; Fablet et al., 2023; Thiria et al., 2023; Martin et al., 2023; Archambault et al., 2023). However, training neural networks usually requires pairs of ground truth and observations which we lack in real-world situations. To overcome this limitation, previous works have examined two main strategies: training the network on a realist simulation of the observing system (Fablet et al., 2023) or using loss functions not requiring ground truth (Archambault. et al., 2023; Martin et al., 2023; Archambault et al., 2023). As the first method entirely relies on the realism of the twin experiment its application to real-world observations suffers from a domain gap, especially for multi-variate approaches. To this day, if the feasibility of training SSH-only networks on simulation alone has been demonstrated to be possible (Fablet et al., 2023), transferring multi-variate approaches has never been successfully performed. On the other hand, our previous work (Archambault et al., 2023) shows that the method trained using only observations suffers from a drop in performance compared to fully supervised ones.

In this work, we are interested in combining the advantages of these two methods. We propose to perform a pre-training on a multi-variate simulation of the observing system, and a fine-tuning using only real-world observations. This paper is structured as follows: first, we introduce the different data used in this study and present the inpainting methods. Then we compare the different learning strategies and present a benchmark of this application.

2 DATA

2.1 Observing System Simulation Experiment

In geosciences, one of the major difficulties is that the ground truth we aim to estimate is often inaccessible. To understand the impact of observation systems on the reconstruction process, researchers employ a method known as the Observing System Simulation Experiment (OSSE). This technique involves simulating the observation operator on a physical simulation, in our case to replicate realistic satellite measurements. The oceanographic community widely uses it as it provides ways to test reconstruction methods and errors (Amores et al., 2018; Stegner et al., 2021; Gaultier et al., 2016). In this context, we use the SSH and SST variables from a realistic simulation as the ground truth upon which we simulate satellite measures. In this study, we select a portion of the North Atlantic Ocean (from latitudes 33° to 43° and longitudes -65° to -55°) of the Global Ocean Physical Reanalysis, GLORYS (CMEMS, 2020). We retrieve 7194 daily images of data starting from Mars 20, 2000 to December 29, 2019. Hereafter, we call $\mathbf{X}$ the ground truth variable, $\mathcal{H}$ our simulated observing operator, and $\mathbf{Y} = \mathcal{H}(\mathbf{X})$ their associated simulated observations. Following our previous work (Archambault et al., 2023), we present a multivariate OSSE with enough pairs of observations and ground truth to train a neural network.

2.1.1 Sea Surface Height

The nadir-pointing SSH observations are localized on a precise spatiotemporal support (denoted Ω) which we want to reproduce in our OSSE. Using the support from the Copernicus sea level real-world observations (CMEMS, 2021), and the ground truth data $\mathbf{X}^{ssh}$ from GLORYS we simulate SSH observations $\mathbf{Y}^{ssh}$ as the trilinear interpolation of $\mathbf{X}^{ssh}$ on each point of the support. We add an instrumental error $\varepsilon \sim \mathcal{N}(0,\sigma)$ with $\sigma = 1.9\,\text{cm}$, which is the distribution used in the Ocean data challenge 2020 (CLS/MEOM, 2020). The SSH observing system is thus defined as follows:

$$\mathbf{Y}^{ssh} = \mathcal{H}^{ssh}\left(\mathbf{X}^{ssh}, \Omega\right) + \varepsilon \tag{1}$$

where $\mathcal{H}^{ssh}$ SSH observation operator. An example of these simulated along-track measurements is presented in Figure 1.

2.1.2 Sea Surface Temperature

SST remote sensing relies on direct infrared measurements, enabling broader coverage but making the data susceptible to cloud interference. To address gaps left by the clouds, the oceanographic community merges the images taken by several satellites through linear interpolation. The interpolated images present artificially smoothed structures in thick cloud regions. We simulate this process as follows:

$$\begin{aligned}\mathbf{Y}^{sst} &= \mathcal{H}^{sst}\left(\mathbf{X}^{sst}, C\right) \\ &= (1-C)\odot\left(\mathbf{X}^{sst}+\varepsilon\right) + C\odot\mathcal{G}_\sigma \star \left(\mathbf{X}^{sst}+\varepsilon\right)\end{aligned} \tag{2}$$

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, and C is the cloud cover (1 when a cloud is present and 0 elsewhere). We first add ε to the SST ground

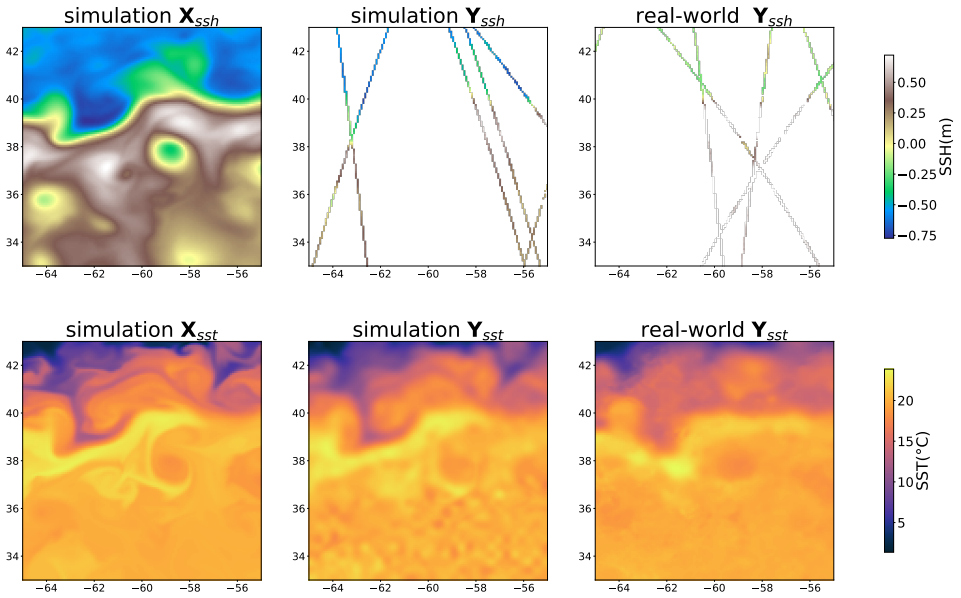

Figure 1: Daily example of SSH and SST data. The first column is the ground truth from the physical model, the second column is our simulation of the observations, and the last is the real-world satellite data.

truth to simulate the instrumental error, and then perform a Gaussian blur with a kernel $\mathcal{G}_\sigma$ ($\sigma = 16$km). This smoothing is then applied only when clouds are present, to mimic in-equal spatial resolution of the satellite SST images. $\mathcal{H}^{sst}$ adds a noise with a standard deviation of 0.5°C out of the 4.96°C of the SST standard deviation.

2.2 Real-world data

For SSH real-world data we propose to use the ungrided data used as inputs of the DUACS protocol (CMEMS, 2021). Concerning SST data, we use the Multiscale Ultrahigh Resolution (MUR) SST (NASA/JPL, 2019). These products in cloud-free, as missing values are inpainted using a linear optimal interpolation, which leads to a smoothing of high spatial frequencies when could are present.

3 PROPOSED METHOD

To exploit the temporal coherence of the sparsely observed ocean structures, we propose to perform the SSH inpainting on a time series of 21 daily images. The neural network f_θ estimates the SSH fields $\hat{\mathbf{X}}^{ssh}$ from observations $\mathbf{Y}$, which could be $\mathbf{Y}^{ssh}$ for SST-agnostic networks or $(\mathbf{Y}^{ssh}, \mathbf{Y}^{sst})$ for SST-aware networks. $\mathbf{Y}^{ssh}, \mathbf{Y}^{sst}, \hat{\mathbf{X}}^{ssh}$ have the same size (21 images of size 128 by 128).

3.1 Architecture

Following our previous work (Archambault et al., 2023) we propose to use an Attention Based Encoder Decoder (ABED) to perform the inpainting. The architecture of the network is presented in Figure 2. It starts with two encoding blocks that divide the spatial dimensions of the images by 2. Then spatiotemporal attention and decoding blocks are performed successively to get back to the original size. This mechanism allows the network to highlight essential features in the input images such as oceanic eddies, while reducing the importance of irrelevant ones such as cloudy areas for instance. Attention modules are widely used in many computer vision tasks including image inpainting (Guo et al., 2021) and can be transposed to geoscience applications (Che et al., 2022; Archambault et al., 2023). Furthermore, the nature of attention modules is well suited to fine-tuning as irrelevant pre-trained filters can easily be weighed with small values during the refitting of attention layers.

Our spatiotemporal attention block is divided into two steps: temporal and spatial attention. Our approach follows the Convolutional Block Attention Module (CBAM) principle introduced by (Woo et al., 2018), which proposed to compute consecutively channel and spatial attention. We adapt this concept to spatiotemporal data by integrating temporal information into the channel attention mechanism. The temporal attention begins by computing the spatial average for each channel and instant, resulting in a tensor of dimensions $C \times T$, where C denotes the

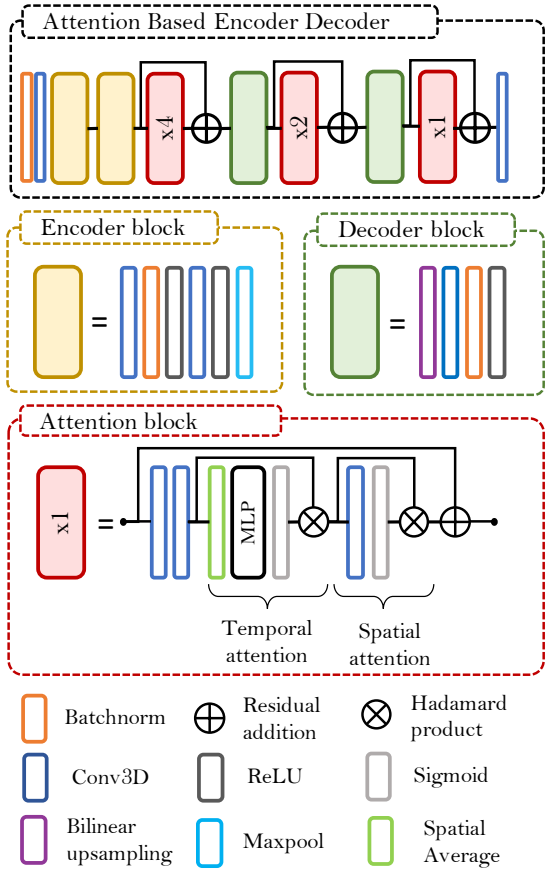

Figure 2: Overview of the Attention-Based-Encoder-Decoder. The network starts with two encoding blocks each dividing the spatial dimensions of the images by 2. Attentions modules are then performed, followed by a residual connection and a decoding block.

number of channels and T represents the length of the time series. We then apply a two-layer perceptron with shared weights across all time steps followed by a sigmoid activation. The resulting tensor has values between 0 and 1 and we multiply it to the input tensor to perform temporal attention. To proceed with spatial attention, we use a 3-dimensional convolution, where the kernel's temporal size equals the length of the time series. We also use a sigmoid activation to obtain a 2D image between 0 and 1 before multiplying with the input tensor. Subsequently, a residual skip connection is computed. This described block is iteratively applied 4 times for the first block, 2 times for the second block, and once for the final block as Figure 2 shows[1].

[1]Our implementation and training data are available here: https://gitlab.lip6.fr/archambault/visapp2024

3.2 Loss functions

We propose two loss functions to train this neural network. The first method takes the Mean Square Error (MSE) on the entire image in a supervised fashion. This is applicable exclusively in the OSSE setting where we have access to the ground truth during training. The second approach uses an unsupervised loss function, enabling training in scenarios where only observations are accessible. Figure 3 gives a visual overview of the two training methods.

We detail hereafter the unsupervised loss function. Prior studies suggest that it is possible to perform the SSH interpolation from observations only. Using a spatiotemporal Deep Image Prior strategy (Ulyanov et al., 2017), meaning overfiting observations from a white noise, (Filoche et al., 2022) showed that if chosen correctly, the architecture of the neural network acts as a regularization that outperforms DUACS. Following the same principle (Archambault. et al., 2023) estimated an SSH map from gridded SST images on one year of data. One of the major limitations of these two methods is that they need to be refitted if applied to unseen data which makes them extremely inefficient computationally speaking. To overcome this limitation, (Archambault et al., 2023; Martin et al., 2023) proposed to train the neural network in a way that doesn't require refitting and that enables the use of contextual information such as SST.

To train the neural network in a context where only observations are available, we apply the observing operator $\mathcal{H}^{ssh}$ on the estimate field $\hat{\mathbf{X}}^{ssh} = f_\theta(\mathbf{Y}^{ssh})$ before computing the MSE. This allows us to get back to the observation domain where we have data to constrain the method. The unsupervised loss function is defined as follows:

$$\begin{aligned}\mathcal{L}_{unsup}(\mathbf{Y}^{ssh}, \hat{\mathbf{X}}^{ssh}) &= \frac{1}{N}\sum_k \left(\mathbf{Y}_k^{ssh} - \mathcal{H}^{ssh}(\hat{\mathbf{X}}^{ssh})_k\right)^2 \\ &= \frac{1}{N}\sum_k \left(\mathbf{Y}_k^{ssh} - \hat{\mathbf{Y}}_k^{ssh}\right)^2 \end{aligned} \quad (3)$$

N is the number of SSH samples in the observation vector $\mathbf{Y}^{ssh}$ and $\hat{\mathbf{Y}}^{ssh}$ is the estimation of the observations. To ensure that the network produces a valid estimation outside of the SSH measures given as inputs, we leave aside the data from one satellite (out of three to six satellites depending on the period) from the neural network's inputs. In Figure 3 we call this input vector $\mathbf{Y}^{ssh_{in}}$. The network is then controlled on all the observations, the ones given in input and the left-aside. Therefore, to accurately estimate the withheld observations, the network is forced to generalize well on the whole image.

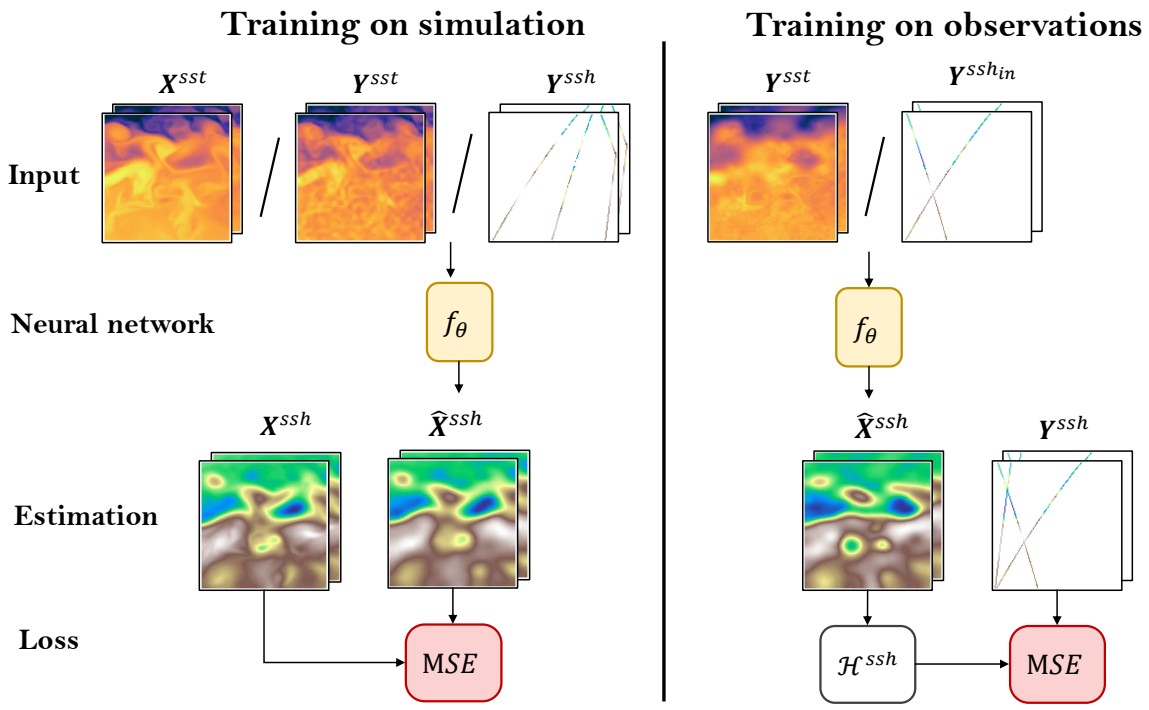

Figure 3: Computational graph of the two different training methods. In the supervised simulation situation (left) the neural network f_θ takes as inputs either $\mathbf{Y}^{ssh}$ alone, $\mathbf{Y}^{ssh}$ and $\mathbf{Y}^{sst}$, or $\mathbf{Y}^{ssh}$ and $\mathbf{X}^{sst}$. The network is then controlled in a supervised manner. In the real-world observations framework (right), only a portion of satellite measures is passed to the network inputs ($\mathbf{Y}_i^{ssh}$) with the SST optionally. The observing operator $\mathcal{H}^{ssh}$ is then applied to the estimation $\hat{\mathbf{X}}^{ssh}$ so that the network can be controlled at the location where we have access to observations only.

3.3 Training procedures

Using the two losses given in Section 3.2 and the two datasets described in Section 2 several training and fine-tuning strategies are possible.

First, we can perform a supervised training on the OSSE data, and directly infer real-world data. This approach was tested by (Fablet et al., 2023) and has the advantage of being straightforward, but could suffer from the domain gap between the simulation and the real world. Specifically, (Fablet et al., 2021) achieved the training from SSH-only observations, but, to this day, no SST-using method has successfully been transferred to real-world data.

Another possibility is to train on real-world observations only using the loss described in Equation 3. This approach was tested by (Archambault. et al., 2023; Martin et al., 2023) which successfully included SST information in SSH inpainting. However, comparing supervised and unsupervised methods on this OSSE, our previous work shows a significant drop in reconstruction performances for the unsupervised interpolations (Archambault et al., 2023).

To benefit from the supervised training on simulated data without suffering from the domain gap, we propose to pre-train ABED on the OSSE and to fine-tune it on real-world data. In the pre-training step, we test three input settings: $\mathbf{Y}^{ssh}$ for SST-agnostics networks, $(\mathbf{Y}^{ssh}, \mathbf{Y}^{sst})$ for networks pre-trained with SST observations, and $(\mathbf{Y}^{ssh}, \mathbf{X}^{sst})$ for networks pre-trained using SST ground truth. Specifically, comparing methods pre-trained with $\mathbf{X}^{sst}$ or $\mathbf{Y}^{sst}$ will help us to understand the impact of our OSSE. However, while training directly from observations or fine-tuning, only $\mathbf{Y}^{sst}$ is available, therefore we refit every network using the same SST observations.

3.4 Training details

Train, validation, test split. The dataset is partitioned as follows: we use the year 2017 for testing, and we validate our methods on three periods: (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 is used for training, except for 15-day periods set aside to prevent data leakage on the validation or the test set. The partition is the same for the OSSE data and

Learning method \ Input data	$\mathbf{Y}_{ssh}$			$\mathbf{Y}_{ssh}+\mathbf{Y}_{sst}$			$\mathbf{Y}_{ssh}+\mathbf{X}_{sst}$		
	μ	σ_t	λ_x	μ	σ_t	λ_x	μ	σ_t	λ_x
Observation	6.52	1.95	111	6.13	1.84	104	—	—	—
Simulation	6.35	1.9	112	6.2	1.87	108	6.85	2.22	111
Both	6.27	1.85	110	**5.77**	1.64	**102**	**5.77**	**1.6**	103

Table 1: Scores of the 10-member ensemble of the ABED inpainting. We test the following learning methods: "Observations" (trained only with real-world data), "Simulation" (trained only with simulated data), and "Both" (pre-trained on simulation and fine-tuned on real-world data). When the network is pre-trained using $\mathbf{X}^{sst}$ it is still fine-tuned with $\mathbf{Y}^{sst}$.

the real-world data.

Normalization and preprocessing. We center and reduce the data of the neural network using the mean and standard deviation of the training data. We grid SSH along-track data to a series of images of size $21 \times 128 \times 128$, and we set every pixel without information to 0. We subtract to the daily SST images the seasonal mean of SST, i.e. the mean of SST maps across all years in our training dataset, taken for this day of the year.

Optimization. We use the ADAM optimizer (Kingma and Ba, 2017), with a starting learning rate of 5.10^{-5} and a multiplicative decay of 0.99. While fine-tuning, the initial learning rate is set to 10^{-5} and the decay to 0.9.

Ensemble. To address the sensitivity of neural network optimization to weight initialization, we adopt an ensemble strategy by training ten networks for each configuration. Referred to as the "Ensemble estimation", this approach involves averaging the SSH maps generated by the networks. The Ensemble estimation usually produces better estimations than each member separately (Hinton and Dean, 2015), and specifically for SSH estimation (Archambault et al., 2023; Archambault. et al., 2023; Filoche et al., 2022).

4 RESULTS

4.1 Comparison of the different methods

In the following analysis, we compare ABED interpolations based on two distinct criteria: the learning approach employed and the input data. The comparison involves three learning methodologies: unsupervised learning on observations only, training on simulated data with a direct inference on real-world data, and a hybrid approach involving pre-training on simulation and fine-tuning on observations. Simultaneously, we evaluate three distinct sets of inputs: SSH-only, SSH and noised SST, and SSH and SST ground truth (a configuration only possible while training on simulation). We evaluate all methods on the along-track data from a satellite left aside from the inputs. To be coherent with some of the interpolations of the ocean data challenge (CLS/MEOM, 2021), the evaluation is done on a smaller area than the one used for training (from 34° to 42° North and -65° to -55° West). We want to stress that the data used for the evaluation are not used as inputs by any of these methods and present an instrumental white noise with a standard deviation from 2 to 3 cm. We still evaluate with this data keeping in mind that the noise is leading to overestimating the errors of the methods.

We consider Root Mean Squared Error (RMSE) on the independent satellite data and compute μ, its temporal mean, and σ_t, its temporal standard deviation. We also compute the spatial power density spectrum (PSD) of the error and of the independent data and retrieve λ_x (in km), the wavelength where the PSD of the error equals the PSD of the reference. It can be seen as the smallest wavelength that is at least half resolved by the interpolation method. For further details about the implementation of λ_x, we refer the reader to (Le Guillou et al., 2020). Table 1 presents the scores of the different settings.

Is our OSSE realistic? Through this experiment, we are able to assess the realism of our OSSE on SSH and SST simulated observations. When examining the SSH-only methods, we find a substantial improvement in the methods trained on simulation compared to the ones trained on the observations alone. Fine-tuning also leads to a reconstruction improvement although smaller than the one brought by the pre-training. We conclude that the SSH observations are correctly simulated. However, the SST-aware methods trained on real-world data perform better than the one trained on simulation and even more so for the one trained using the SST ground truth. This underlines the fact that the SST noise is not perfectly simulated, even if it is still more realistic than the ground truth SST. We also see that the two SST methods achieve very similar performances after being fine-tuned, this shows that given an efficient transfer learning strategy, we do not necessarily need to pre-train

Method	SST	NN	Learning	$\mu(cm)$	$\sigma_t(cm)$	$\lambda_x(km)$
DUACS	✗	✗	✗	7.66	2.66	138
DYMOST	✗	✗	✗	6.75	2.00	121
MIOST	✗	✗	✗	6.75	2.00	121
BFN	✗	✗	✗	7.46	2.59	114
4DVarNet	✗	✓	simulation	6.56	1.84	104
MUSTI	✓	✓	observation	6.26	1.96	107
ConvLTSM-SSH	✗	✓	observation	6.82	1.86	108
ConvLTSM-SSH-SST	✓	✓	observation	6.29	**1.60**	**102**
ABED-SSH	✗	✓	both	6.27	1.85	110
ABED-SSH-SST	✓	✓	both	**5.74**	1.61	**102**

Table 2: Benchmark of the interpolations provided by (CLS/MEOM, 2021), including methods using SST or not, using neural networks or not (NN), with different learning strategies. ABED interpolations are given for the pre-trained and fine-tuned version using SSH or SST.

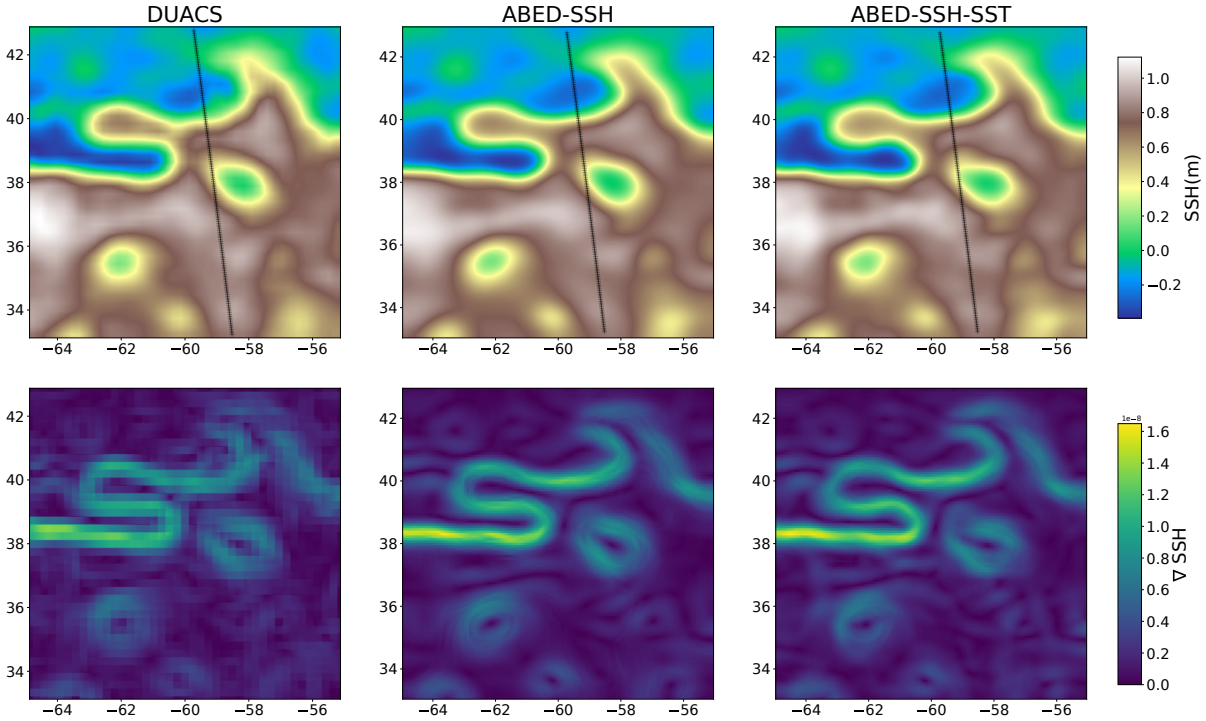

Figure 4: Estimated SSH maps from DUACS, ABED-SSH, ABED-SSH-SST and the norm of their spatial gradient. We plot the trajectory of the satellite used for evaluation in Figure 5

the network in a realistic setting. This point will be further discussed in Section 5.2.

4.2 Comparison with state-of-the-art data.

We access the performance of our method by comparing the ABED pre-trained and fine-tuned inpaintings to the state-of-the-art interpolations methods, on the left-aside satellite data. The benchmarked methods include DUACS, the most widely used product in operational applications (Taburet et al., 2019). We include three physic-based data assimilation schemes: DYMOST (Ubelmann et al., 2016; Ballarotta et al., 2020), MIOST (Ardhuin et al., 2020) and BFN (Le Guillou et al., 2020). We also compare with the supervised neural network 4DVarNet (Fablet et al., 2021), and with neural networks trained using observations only such as MUSTI (Archambault. et al., 2023) and the ConvLTSM introduced by (Martin et al., 2023).

The results summarized in Table 2 show that our method achieves state-of-the-art performances in terms of RMSE among methods using only SSH and methods using SST. More specifically, we see a clear predominance of neural network-based methods as

well as SST-aware methods. ABED-SSH-SST thanks to its pre-training and refitting improves the reconstruction of DUACS by 1.92 cm which acounts for 25% of its RMSE.

An example of improvement brought by the SST. Because of the absence of fully gridded ground truth data in the real-world setup, the interpretation of the results is difficult. Nonetheless in Figure 4 we present the estimated maps of the DUACS operational product, as well as the one of ABED-SSH and ABED-SSH-SST. We also compute the norm of the spatial gradient of the SSH to highlight the areas of strong variations. Visually, we see smaller and more precise eddies in the ABED inpainting, especially in the SST-aware version. However, as it is still hard to show the impact of SST on the reconstruction, we plot in Figure 5 the interpolation of the three different maps with the targeted independent data. We select an area where an improvement is brought by the SST, as the SST-agnostic methods clearly overestimate the SSH. When we plot the trajectory of the satellite on the SST image we see that the selected area corresponds to a small drop of temperature. This is a typical example of the interest in using temperature to constrain the inpainting as this high-resolution information lacking in input SSH observations.

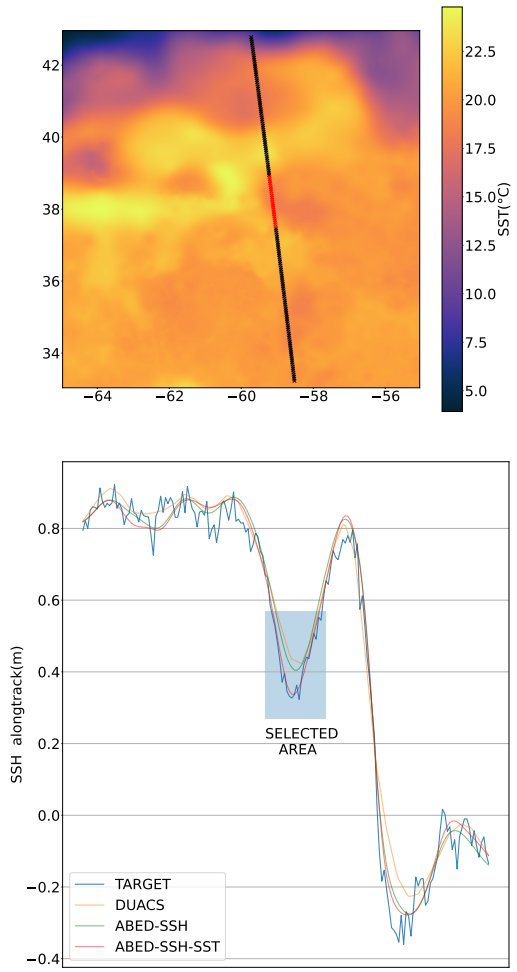

Figure 5: An example of reconstruction improvement brought by the SST. The SST image is represented along with the trajectory of the evaluation satellite, as well as the along-track interpolations of the three estimations presented in Figure 4.

5 CONCLUSIONS AND PERSPECTIVES

5.1 Summary

Throughout this study, we successfully applied a transfer learning strategy to perform the interpolation of SSH using SST with an Attention-Based Encoder Decoder. As in an operational scenario no fully gridded ground truth is accessible to train the neural network, we developed an Observing System Simulation Experiment, a twin experiment that simulates the observation system of the satellites. Doing so, we were able to compute pairs of realistic input/output based on a physical simulation of the ocean and pre-train our neural network. Then using an unsupervised loss function we fine-tuned the neural network on real-world data. We show that the pre-training enhances the reconstruction as our method achieves better results than the same network trained directly from observations. This is also the case for the fine-tuned version which outperforms the model solely trained on simulated data proving the efficiency of the refitting. Benchmarking ABED with standard interpolation methods, either based on physical prior models, neural networks trained on simulations, or directly from observations, we show that our training strategy achieves state-of-the-art performances among SST-aware or SSH-only methods. Compared to DUACS, the most widely used oceanography product, we report an RMSE improvement of 1.92 cm out of the 7.66 cm of error (25% of the RMSE of DUACS) on noised independent along-track data. We conclude from this experiment that pre-train and fine-tune neural networks help the reconstruction of variables, in settings where no ground truth is available to constrain the inversion.

5.2 Discussions and perspectives

Extension of the methodology to other variables. The proposed method could be applied to other input

or target variables if the following conditions are fulfilled. First, the variables must be correlated to each other, and we must have access to a realistic physical model that we can use to build a multi-variate OSSE. The data generated through this mean can then be used in the pre-training, enabling the network to accurately learn the physical underlying link. Then the fine-tuning will adapt this learning to the noise of the real-world data. One of the most obvious candidates to serve as well as a target or input variable is the sea's Chlorophyll, which is a passive tracer of the oceanic currents (Chelton et al., 2011).

Realism of the OSSE. We show that the multivariate OSSE performed in this study was realistic, as well for the SSH noise than for the SST noise. However, given an appropriate transfer strategy, the networks trained on the noised version of the SST and networks trained on the ground truth SST achieve similar results once retrained. This leads us to reconsider the necessity of computing a very realistic noise on contextual information, as the fine-tuning process will get rid of the learned features that do not appear in real-world data.

Toward a global gridded image. The experiment that we performed in this work was focusing on a single geographic area. Training a method able to estimate SSH on a global scale would require further work. For instance, as the physical relationship between SSH and SST depends on latitude, we are curious to know if a global model would be competitive compared to several local models.

REFERENCES

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

Archambault., T., Filoche., A., Charantonis., A., and Béréziat., D. (2023). Multimodal unsupervised spatio-temporal interpolation of satellite ocean altimetry maps. In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 4: VISAPP*, pages 159–167. INSTICC, SciTePress.

Archambault, T., Filoche, A., Charantonis, A., Béréziat, D., and Thiria, S. (2023). Unsupervised learning of sea surface height interpolation from multi-variate simulated satellite observations. *Submitted to Journal of Advances of Modeling Earth Systems (JAMES)*.

Ardhuin, F., Ubelmann, C., Dibarboure, G., Gaultier, L., Ponte, A., Ballarotta, M., and Faugère, Y. (2020). Reconstructing ocean surface current combining altimetry and future spaceborne doppler data. Earth and Space Science Open Archive.

Ballarotta, M., Ubelmann, C., Rogé, M., Fournier, F., Faugère, Y., Dibarboure, G., Morrow, R., and Picot, N. (2020). Dynamic mapping of along-track ocean altimetry: Performance from real observations. *Journal of Atmospheric and Oceanic Technology*, 37:1593–1601.

Bretherton, F., Davis, R., and Fandry, C. (1976). A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Research and Oceanographic Abstracts*, 23:559–582.

Che, H., Niu, D., Zang, Z., Cao, Y., and Chen, X. (2022). Ed-drap: Encoder–decoder deep residual attention prediction network for radar echoes. *IEEE Geoscience and Remote Sensing Letters*, 19.

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

CLS/MEOM (2020). Swot data challenge natl60 [dataset].

CLS/MEOM (2021). Data challenge ose - 2021a_ssh_mapping_ose [dataset].

CMEMS (2020). Global ocean physics reanalysis [dataset].

CMEMS (2021). Global ocean along-track l3 sea surface heights reprocessed (1993-ongoing) tailored for data assimilation [dataset].

Fablet, R., Amar, M., Febvre, Q., Beauchamp, M., and Chapron, B. (2021). 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.

Fablet, R., Febvre, Q., and Chapron, B. (2023). Multimodal 4dvarnets for the reconstruction of sea surface dynamics from sst-ssh synergies. *IEEE Transactions on Geoscience and Remote Sensing*, 61.

Filoche, A., Archambault, T., Charantonis, A., and 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., and Fu, L. (2016). The challenge of using future SWOT data for oceanic field reconstruction. *Journal of Atmospheric and Oceanic Technology*, 33:119–126.

Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., Zhang, S.-H., Martin, R. R., Cheng, M.-M., and Hu, S.-M. (2021). Attention mechanisms in computer vision: A survey. *Computational Visual Media*, 8:331–368.

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

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

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

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

Martin, S. (2014). *An Introduction to Ocean Remote Sensing*. Cambridge University Press, 2 edition.

Martin, S. A., Manucharyan, G. E., and Klein, P. (2023). Synthesizing sea surface temperature and satellite altimetry observations using deep learning improves the accuracy and resolution of gridded sea surface height anomalies. *Journal of Advances in Modeling Earth Systems*, 15(5):e2022MS003589. e2022MS003589 2022MS003589.

McCann, M., Jin, K., and Unser, M. (2017). Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34:85–95.

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

NASA/JPL (2019). Ghrsst level 4 mur 0.25deg global foundation sea surface temperature analysis (v4.2) [dataset].

Qin, Z., Zeng, Q., Zong, Y., and Xu, F. (2021). Image inpainting based on deep learning: A review. *Displays*, 69:102028.

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

Taburet, G., Sanchez-Roman, A., Ballarotta, M., Pujol, M.-I., Legeais, J.-F., Fournier, F., Faugere, Y., and Dibarboure, G. (2019). DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Sci*, 15:1207–1224.

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

Ubelmann, C., Cornuelle, B., and Fu, L. (2016). Dynamic mapping of along-track ocean altimetry: Method and performance from observing system simulation experiments. *Journal of Atmospheric and Oceanic Technology*, 33:1691–1699.

Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2017). Deep image prior. *International Journal of Computer Vision*, 128:1867–1888.

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

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., and Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature Climate Change 2013 3:10*, 3:875–883.