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

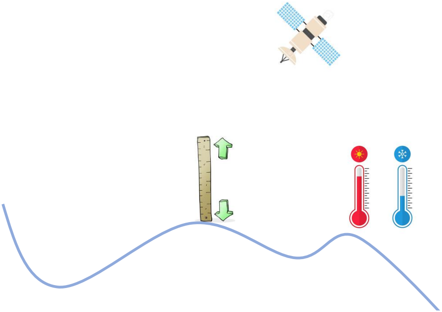
Théo Archambault *Ph.D. candidate at Sorbonne Université, LIP6 and LOCEAN*
Arthur Filoche, Anastase Charantonis, Dominique Béréziat

24 avril 2024

Summary

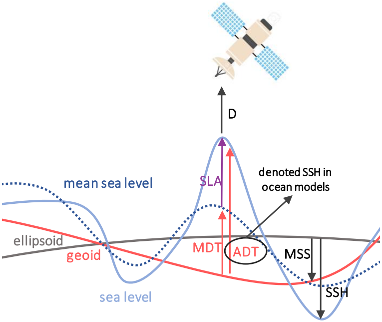
1. Multi-Variate satellite surface observations of the ocean
2. Problem Statement
3. Proposed Method
4. Results
5. Conclusion and perspectives

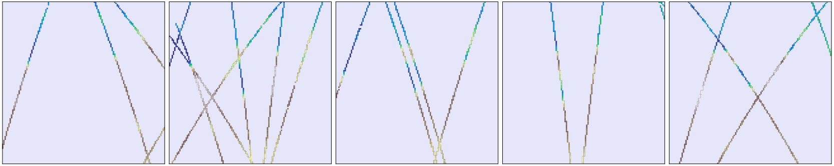
Multi-Variate satellite surface observations of the ocean



Sea Surface Height

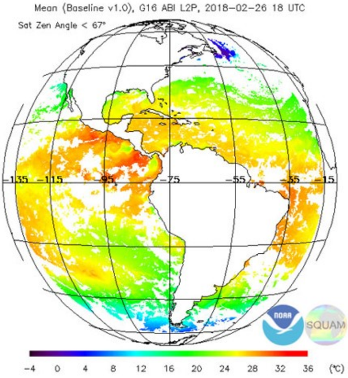
- Used to derive surface currents to geostrophic approximation
- Measurement principle : return time of a radar pulse
- Nadir-pointing altimeters : only measure data along their ground tracks

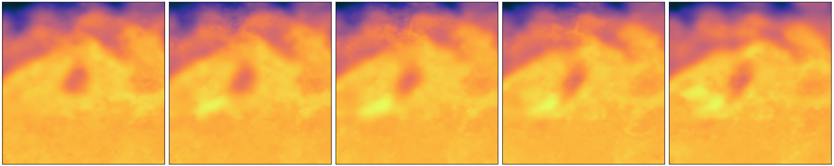




Sea Surface Temperature

- Measurement principle : Direct infra-red image with high resolution (1/25°)
- Cloud introduce gaps in data
- Fully gridded images : obtained through linear Optimal Interpolation combining several satellites and in-situ measures.
- Present noise of different frequencies : high-frequency instrumental errors and low-frequency blurring.





Physical relationship

The two variables are linked to ocean surface currents.

- SSH through geostrophic approximation (equilibrium Coriolis and pressure force)

$$\mathbf{w}_g = \begin{pmatrix} u_{\mathrm{g}} \\ v_{\mathrm{g}} \end{pmatrix} = \begin{pmatrix} -\frac{g}{f_0}\frac{\partial h}{\partial y} \\ \frac{g}{f_0}\frac{\partial h}{\partial x} \end{pmatrix} \tag{1}$$

- SST through advection dynamics

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

Problem Statement

Inverse Problem Formulation

Let $\mathbf{X}_{ssh}$, $\mathbf{X}_{sst}$ be the ocean state of the target and contextual variables, respectively. Let $\mathbf{Y}_{ssh}$, $\mathbf{Y}_{sst}$ be their associated observations obtained through $\mathcal{H}_{ssh}$, $\mathcal{H}_{sst}$ as follows :

$$\begin{cases} \mathbf{Y}_{ssh} & = & \mathcal{H}_{ssh}(\mathbf{X}_{ssh}) + \varepsilon_{ssh} \\ \mathbf{Y}_{sst} & = & \mathcal{H}_{sst}(\mathbf{X}_{sst}) + \varepsilon_{sst} \end{cases}$$

Inverse Problem Formulation

Let $\mathbf{X}_{ssh}$, $\mathbf{X}_{sst}$ *be the ocean state of the target and contextual variables, respectively. Let* $\mathbf{Y}_{ssh}$, $\mathbf{Y}_{sst}$ *be their associated observations obtained through* $\mathcal{H}_{ssh}$, $\mathcal{H}_{sst}$ *as follows :*

$$\begin{cases} \mathbf{Y}_{ssh} & = & \mathcal{H}_{ssh}\left(\mathbf{X}_{ssh}\right) + \varepsilon_{ssh} \\ \mathbf{Y}_{sst} & = & \mathcal{H}_{sst}\left(\mathbf{X}_{sst}\right) + \varepsilon_{sst} \end{cases}$$

$\mathcal{H}_{ssh}$: the observation operator of the SSH that we aim to inverse using SSH observations and SST contextual

Observing System Simulation Experiment

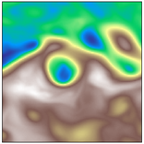
In geosciences, the ground truth is often not accessible : we must use an Observing System Simulation Experiment (OSSE).

- The ground truth is a realistic physical simulation
- We simulate satellite observations to retrieve $(\mathbf{Y}, \mathbf{X})$ pairs
- We use these data to train and evaluate reconstruction methods

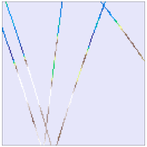
Observing System Simulation Experiment

We use the Global Physical Reanalysis (GLORYS) as simulated ground truth. We take around 20 years of data in the Gulf Stream area (Latitude 33 to 43, longitude -65 to -55)

- We simulate $\mathcal{H}_{ssh}$ as the trilinear interpolation of this ground truth on the path of the satellite. We use the support from real-world observations.



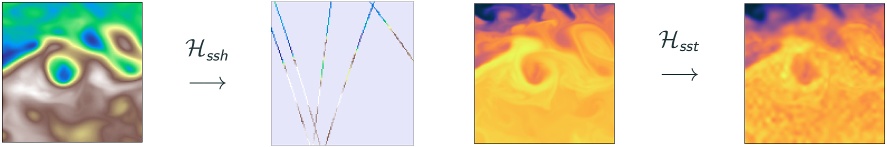
$\xrightarrow{\mathcal{H}_{ssh}}$



Observing System Simulation Experiment

We use the Global Physical Reanalysis (GLORYS) as simulated ground truth. We take around 20 years of data in the Gulf Stream area (Latitude 33 to 43, longitude -65 to -55)

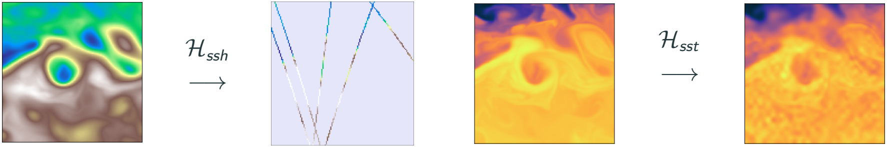
- We simulate $\mathcal{H}_{ssh}$ as the trilinear interpolation of this ground truth on the path of the satellite. We use the support from real-world observations.
- To simulate $\mathcal{H}_{sst}$ we first add a gaussian noise, and we use a real world cloud cover to select areas where we smooth the image using a gaussian blur.



Observing System Simulation Experiment

We use the Global Physical Reanalysis (GLORYS) as simulated ground truth. We take around 20 years of data in the Gulf Stream area (Latitude 33 to 43, longitude -65 to -55)

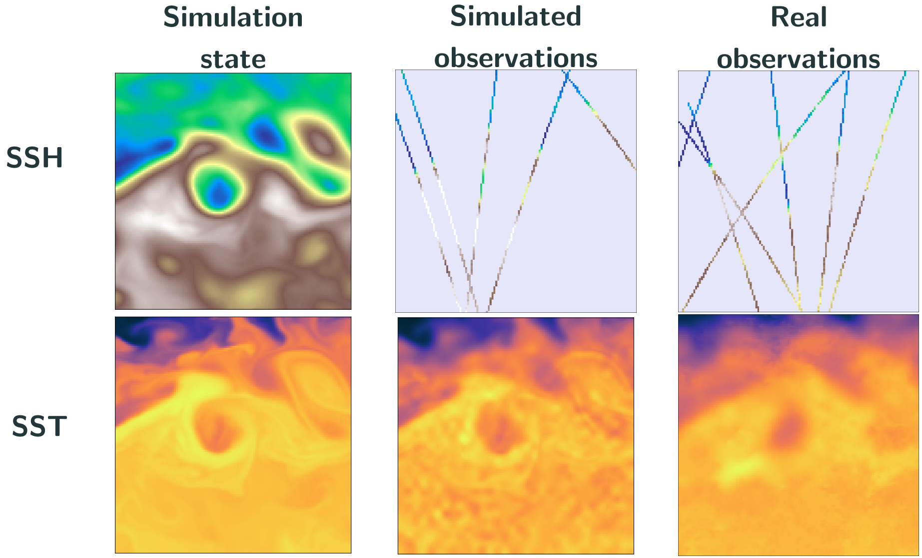
- We simulate $\mathcal{H}_{ssh}$ as the trilinear interpolation of this ground truth on the path of the satellite. We use the support from real-world observations.
- To simulate $\mathcal{H}_{sst}$ we first add a gaussian noise, and we use a real world cloud cover to select areas where we smooth the image using a gaussian blur.



Doing so we have access to :

- Gappy and noisy SSH measures which emulates SSH along tracks measures
- Noisy and blurred SST maps which emulates the interpolation process, and unequal noise resolutions

Domain gap ?


Simulation state
Simulated observations
Real observations
SSH
SST

Proposed Method

- Neural network inpainting of SSH
- Avoid domain gap

Neural network inpainting

We perform the interpolation using a neural network f_θ estimating the SSH field from 21 days on data.

$$\hat{\mathbf{X}}_{ssh} = f_\theta(\mathbf{Y}_{ssh}, \mathbf{Y}_{sst}) \quad \text{with } \hat{\mathbf{X}}_{ssh}, \mathbf{Y}_{ssh}, \mathbf{Y}_{sst} \in \mathbb{R}^{T \times H \times W}$$

Neural network inpainting

We perform the interpolation using a neural network f_θ estimating the SSH field from 21 days on data.

$$\hat{\mathbf{X}}_{ssh} = f_\theta(\mathbf{Y}_{ssh}, \mathbf{Y}_{sst}) \quad \text{with } \hat{\mathbf{X}}_{ssh}, \mathbf{Y}_{ssh}, \mathbf{Y}_{sst} \in \mathbb{R}^{T \times H \times W}$$

Supervised Learning : On simulation we have access to $\mathbf{X}_{ssh}$

- $\mathcal{L}(\hat{\mathbf{X}}_{ssh}, \mathbf{X}_{ssh}) = \mathrm{MSE}(\hat{\mathbf{X}}_{ssh}, \mathbf{X}_{ssh})$
- Domain gap issues

Neural network inpainting

We perform the interpolation using a neural network f_θ estimating the SSH field from 21 days on data.

$$\hat{\mathbf{X}}_{ssh} = f_\theta(\mathbf{Y}_{ssh}, \mathbf{Y}_{sst}) \quad \text{with } \hat{\mathbf{X}}_{ssh}, \mathbf{Y}_{ssh}, \mathbf{Y}_{sst} \in \mathbb{R}^{T \times H \times W}$$

Supervised Learning : On simulation we have access to $\mathbf{X}_{ssh}$

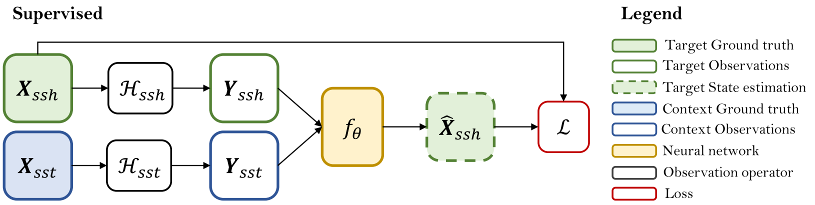
- $\mathcal{L}(\hat{\mathbf{X}}_{ssh}, \mathbf{X}_{ssh}) = \mathrm{MSE}(\hat{\mathbf{X}}_{ssh}, \mathbf{X}_{ssh})$
- Domain gap issues

Unsupervised learning : training on real observations

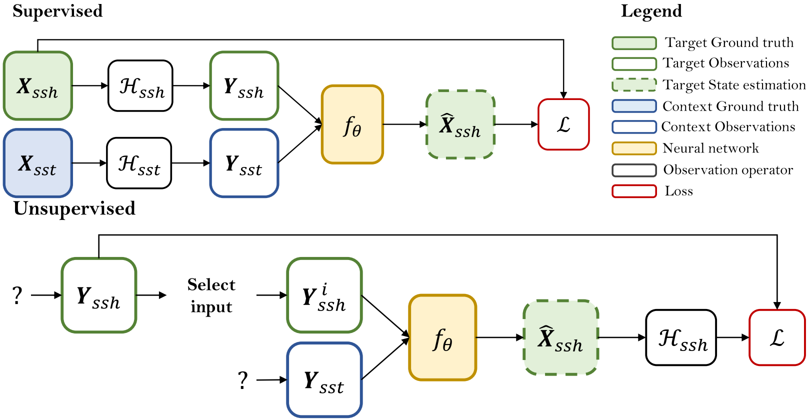
- $\mathcal{L}(\hat{\mathbf{X}}_{ssh}, \mathbf{Y}_{ssh}) = \mathrm{MSE}(\mathcal{H}_{ssh}(\hat{\mathbf{X}}_{ssh}), \mathbf{Y}_{ssh})$
- With observations only, learning the relationship between data is harder

Learning method

Supervised



Learning method

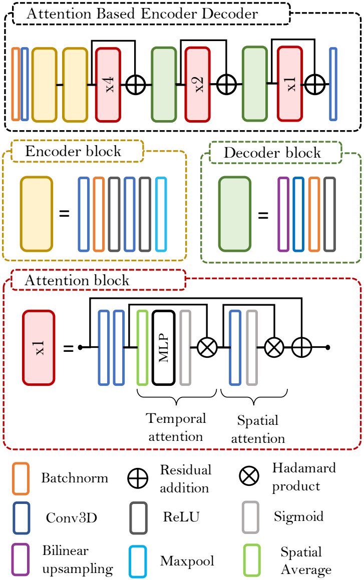


- Subset input data $\mathbf{Y}^i_{ssh}$, and estimates $\hat{\mathbf{X}}_{ssh}$ from $(\mathbf{Y}^i_{ssh}, \mathbf{Y}_{sst})$.
- Apply $\mathcal{H}_{ssh}$ to $\hat{\mathbf{X}}_{ssh}$ before computing the loss.
- It forces the network to accurately estimate the entire map.

Architecture

Attention-Based Encoder-Decoder (ABED).

- Two encoding blocks reducing spatial dimensions
- Spatio-Temporal Attention modules *(inspired by CBAM)*
  - Temporal attention : performs channel and temporal attention together
  - Spatial attention
- Decoding blocks to increase spatial dimensions



Results

Experiments

Evaluation on a real observation dataset : the *Ocean Data Challenge 2021*.

- 1 year of data on the Gulf Stream area
- Provides state-of-the-art reconstruction methods
- Evaluation on independent data
- Metrics : μ the RMSE score (in cm), σ_t its temporal std (in cm), λ_x the half-resolved spatial wavelength (in km)

Experiments

Evaluation on a real observation dataset : the *Ocean Data Challenge 2021*.

- 1 year of data on the Gulf Stream area
- Provides state-of-the-art reconstruction methods
- Evaluation on independent data
- Metrics : μ the RMSE score (in cm), σ_t its temporal std (in cm), λ_x the half-resolved spatial wavelength (in km)

We want to test :

- The impact of using SST : training using SSH, $SSH + nSST$, $SSH + SST$
- The learning strategy

Comparing training strategies

Given the supervised and the unsupervised learning, we derive 3 strategies ;

Comparing training strategies

Given the supervised and the unsupervised learning, we derive 3 strategies ;

- **Observations only** : unsupervised training on real-world observations.

Input data	SSH			SSH+nSST			SSH+SST		
Learning method	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	—	—	—

Comparing training strategies

Given the supervised and the unsupervised learning, we derive 3 strategies ;

- **Observations only** : unsupervised training on real-world observations.
- **Simulation only** : supervised training on simulation and direct inference on real-world data.

Input data	SSH			SSH+nSST			SSH+SST		
Learning method	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

Comparing training strategies

Given the supervised and the unsupervised learning, we derive 3 strategies ;

- **Observations only** : unsupervised training on real-world observations.
- **Simulation only** : supervised training on simulation and direct inference on real-world data.
- **Both** : Supervised pre-training on simulation and unsupervised fine-tuning on real-world observations.

Input data	SSH			SSH+nSST			SSH+SST		
Learning method	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

Comparing training strategies

Input data	SSH			SSH+nSST			SSH+SST		
Learning method	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

Comparing training strategies

Input data	SSH			SSH+nSST			SSH+SST		
Learning method	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

SST enhances the reconstruction :

- On observation only learning
- On simulation only (if trained with the noisy SST version)
- Once fine-tuned it leads to a very high improvement

Comparing training strategies

Input data	SSH			SSH+nSST			SSH+SST		
Learning method	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

SST enhances the reconstruction :

- On observation only learning
- On simulation only (if trained with the noisy SST version)
- Once fine-tuned it leads to a very high improvement

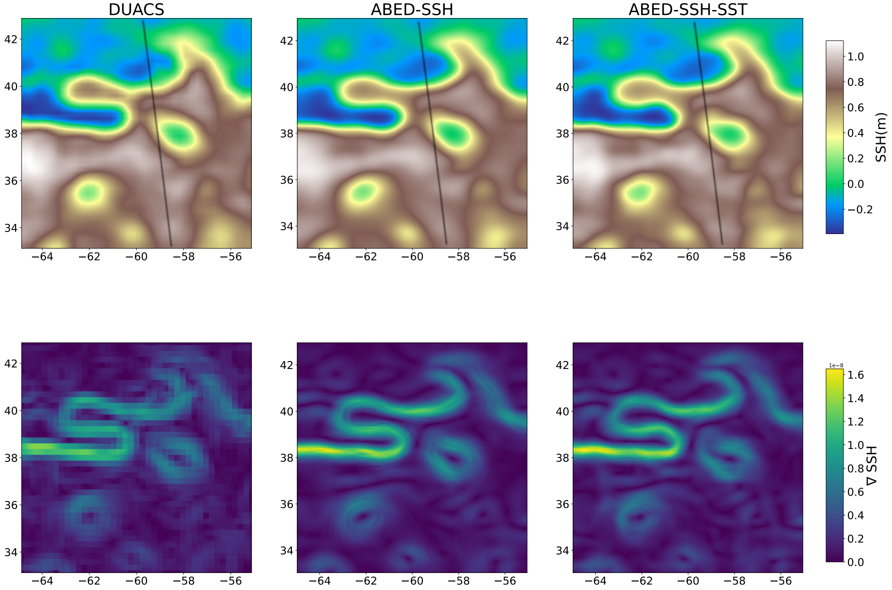
Interest of the pre-training and the fine-tuning :

- Systematically improves the reconstruction
- Once fine-tuned, pre-training on noisy SST or not makes no difference

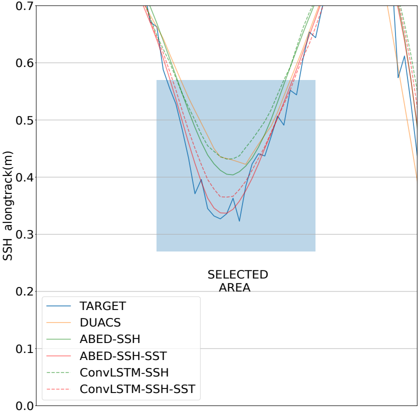
State-of-the-art comparison

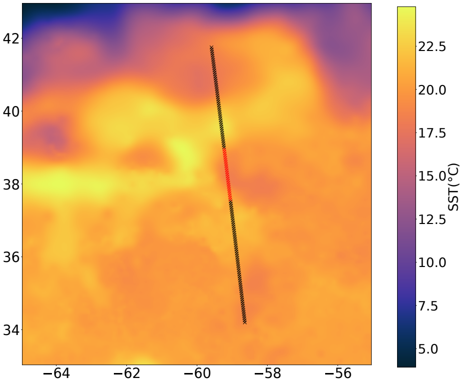
Method	SST	NN	Learning	$\mu(cm)$	$\sigma_t(cm)$	$\lambda_x(km)$
DUACS	✗	✗	✗	7.66	2.66	138
DYMOST	✗	✗	✗	6.75	2.00	121
MIOST	✗	✗	✗	6.75	2.00	121
BFN	✗	✗	✗	7.46	2.59	114
4DVarNet	✗	✓	simulation	6.56	1.84	104
MUSTI	✓	✓	observation	6.26	1.96	107
ConvLstm	✗	✓	observation	6.82	1.86	108
ConvLstm	✓	✓	observation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

Improvements brought by SST(1)



Improvements brought by SST(2)





Conclusion and perspectives

Conclusions and perspectives

- SST improves reconstruction
- Pre-training and fine-tuning neural networks leads to better reconstructions.

Currently working on

- Forecast
- Toward a global product : how to adapt this method in the context of wider geographical areas ?