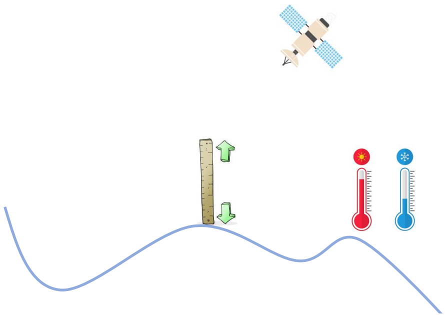
Pre-training and fine-tuning neural networks to interpolate Sea Surface Height field from multi-variate satellite observations

Théo Archambault *Ph.D. candidate at Sorbonne Université, LIP6 and LOCEAN*
Arthur Filoche, Anastase Charantonis, Dominique Béréziat
24 avril 2024

Summary

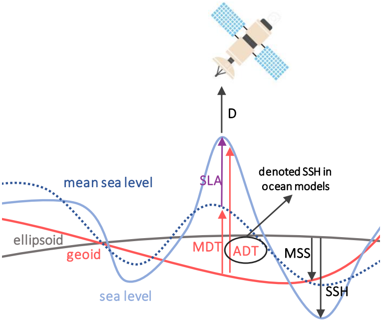
1. Multi-Variate satellite surface observations of the ocean
2. Problem Statement
3. Proposed Method
4. Results on synthetic data
5. Results on observations
6. Conclusion and perspectives

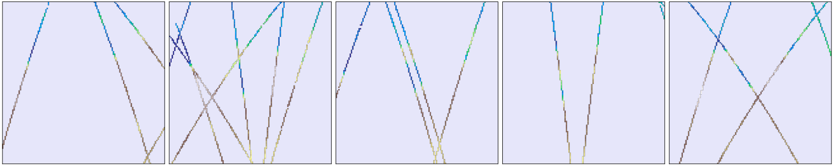
Multi-Variate satellite surface observations of the ocean



Sea Surface Height

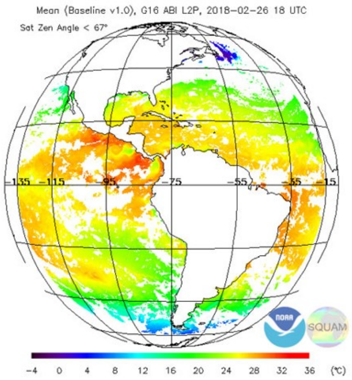
- Used to derive surface currents to geostrophic approximation
- Measurement principle : return time of a radar pulse
- Nadir-pointing altimeters : only measure data along their ground tracks

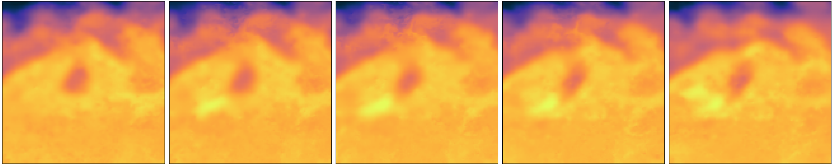




Sea Surface Temperature

- Measurement principle : Direct infra-red image with high resolution (1/25°)
- Sensible to cloud cover
- Combining several satellites through interpolation
- Physically linked to SSH as currents advect heat





Problem Statement

Inverse Problem Formulation

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

$$\begin{cases} \mathbf{Y}_a & = & \mathcal{H}_a(\mathbf{X}_a) + \varepsilon_a \\ \mathbf{Y}_b & = & \mathcal{H}_b(\mathbf{X}_b) + \varepsilon_b \end{cases}$$

$\mathcal{H}_a$ can be seen as a forward operator that we aim to inverse. Therefore, in our application, the SSH is the target variable ($\mathbf{X}_a$), and SST is the contextual information ($\mathbf{X}_b$).

The operational inversion method : DUACS

Data Unification and Altimeter Combination System (DUACS) :

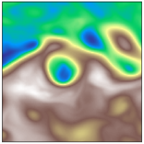
- Optimal linear interpolation of the satellite along-tracks measures from multiple satellites.
- Uses covariances matrices tuned on 25 years of data and not publicly available
- Low effective resolution
- Misses mesoscale structures and eddies

Our goal : enhance DUACS's interpolation using a neural network and SST information.

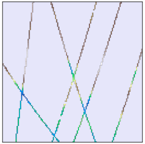
Observing System Simulation Experiment

We use the Global Physical Reanalysis (GLORYS) as simulated ground truth. We take 7194 days of data in the north stream area (Latitude 33 to 43, longitude -65 to -55)

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



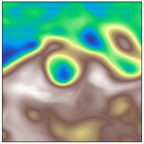
$$\xrightarrow{\mathcal{H}_a}$$



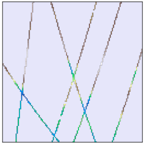
Observing System Simulation Experiment

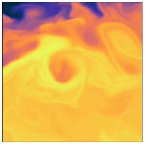
We use the Global Physical Reanalysis (GLORYS) as simulated ground truth. We take 7194 days of data in the north stream area (Latitude 33 to 43, longitude -65 to -55)

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

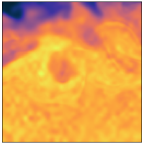


$\xrightarrow{\mathcal{H}_a}$





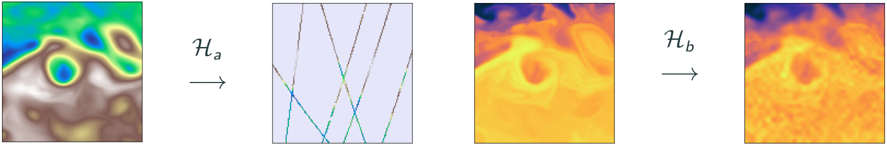
$\xrightarrow{\mathcal{H}_b}$



Observing System Simulation Experiment

We use the Global Physical Reanalysis (GLORYS) as simulated ground truth. We take 7194 days of data in the north stream area (Latitude 33 to 43, longitude -65 to -55)

- We simulate $\mathcal{H}_a$ as the trilinear interpolation of this ground truth on the path of the satellite. We use the support from real-world observations.
- To simulate $\mathcal{H}_b$ we first add a white additive 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

Proposed Method

Neural inversion

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

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

The network is trained using a loss function $\mathcal{L}$ in our case the Mean Squared Error.

Neural inversion

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

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

The network is trained using a loss function $\mathcal{L}$ in our case the Mean Squared Error.

- **Supervised Learning** : If we have access to $\mathbf{X}_a$
  - Requires a realistic twin experiment on a simulation
  - Domain gap issues

Neural inversion

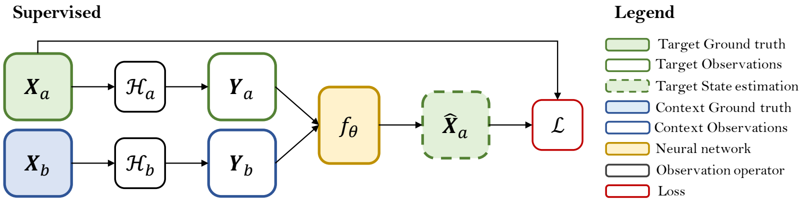
We perform the interpolation using a neural network f_θ estimating the SSH field form 21 days on data.

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

The network is trained using a loss function $\mathcal{L}$ in our case the Mean Squared Error.

- **Supervised Learning** : If we have access to $\mathbf{X}_a$
  - Requires a realistic twin experiment on a simulation
  - Domain gap issues
- **Unsupervised learning** : If we only have access to observations
  - With observations only, learning the relationship between data is harder

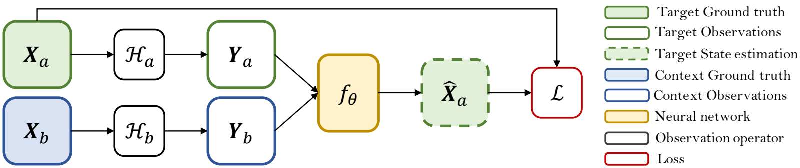
Learning method



- Apply observing operators to $\mathbf{X}_a, \mathbf{X}_b$ and estimates the state from observations.

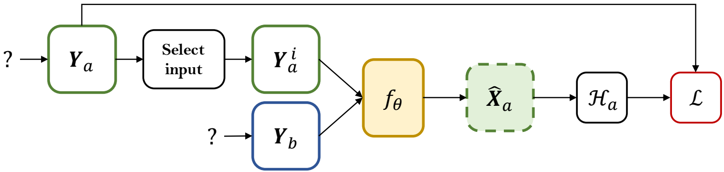
Learning method

Supervised



- Apply observing operators to $\mathbf{X}_a, \mathbf{X}_b$ and estimates the state from observations.

Unsupervised

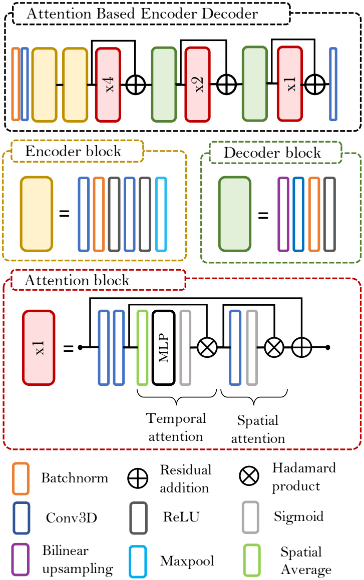


- Subset input data $\mathbf{Y}_a^i$, and estimates $\hat{\mathbf{X}}_a$ from $(\mathbf{Y}_a^i, \mathbf{Y}_b)$
- Apply $\mathcal{H}_a$ to $\hat{\mathbf{X}}_a$ before computing the loss.

Architecture

We propose an attention-based encoder-decoder (ABED) using a time series of along-track SSH and SST to estimate the gridded SSH time-series.

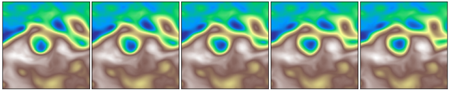
- Two encoding blocks reducing spatial dimensions
- Spatio-Temporal Attention modules
  - Temporal attention : performs channel and temporal attention together
  - Spatial attention
- Decoding blocks to increase spatial dimensions



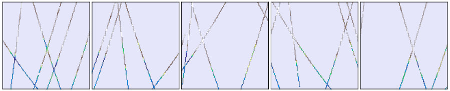
Results on synthetic data

Sea Surface Height

$\mathbf{X}_a$

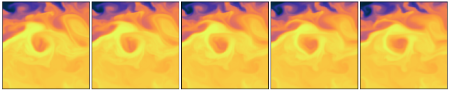


$\mathbf{Y}_a$

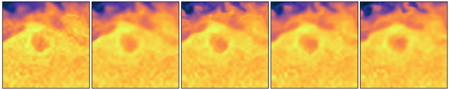


Sea Surface Temperature

$\mathbf{X}_b$



$\mathbf{Y}_b$



Experiment

On the OSSE framework, we want to assess :

Learning strategies :

- Supervised learning
- Unsupervised learning
- Unsupervised learning with regularization

Experiment

On the OSSE framework, we want to assess :

Learning strategies :

- Supervised learning
- Unsupervised learning
- Unsupervised learning with regularization

Impact of SST data :

- SSH-only interpolation
- SSH and noised SST
- SSH and model SST (to give an upper bound performance)

Experiment

On the OSSE framework, we want to assess :

Learning strategies :

- Supervised learning
- Unsupervised learning
- Unsupervised learning with regularization

Impact of SST data :

- SSH-only interpolation
- SSH and noised SST
- SSH and model SST (to give an upper bound performance)

We validate the results using the following metrics :

- SSH reconstruction : RMSE
- Structures analysis : we perform an automatic eddy detection on ground truth and estimation, compare the physical properties of the eddies
- State-of-the-art interpolations methods on a similar OSSE

SSH reconstruction

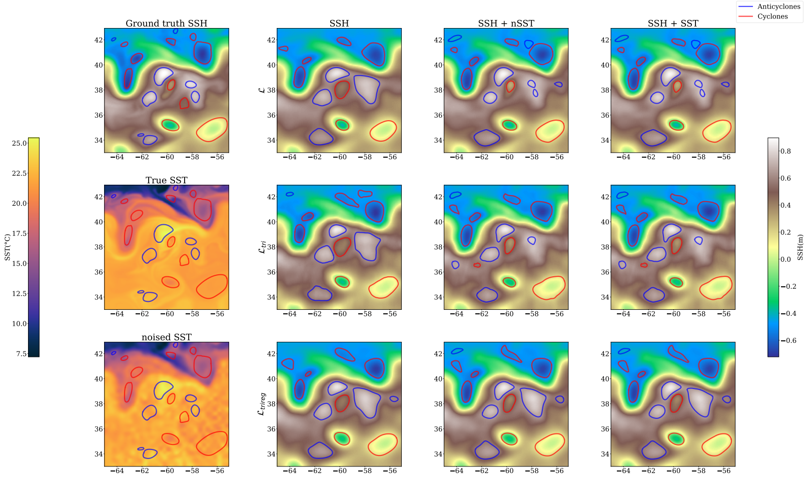
Loss	SSH	SSH+nSST	SSH+SST
supervised	4.18 — 3.85	3.23 — 2.93	**2.92 — 2.59**
unsupervised	4.52 — 4.16	3.86 — 3.51	3.62 — 3.24
unsupervised+regul	4.38 — 4.13	3.73 — 3.48	3.48 — 3.20

Table 1: SSH reconstruction RMSE in centimeters. For each method, the first score is the one of the mean RMSE of 3 neural networks, while the second one is the one of its ensemble estimation.

- SST improves reconstruction
- It is possible to train a neural network in an unsupervised way, although with a performance drop.
- Ensemble estimation improves reconstruction

Structures analysis

We perform an automatic eddy detection (AMEDA) on the generated maps and compare the eddies of the interpolations to the ones of the ground truth.



Structures analysis

Loss	SSH			SSH+nSST			SSH+SST		
	F_1	recall	precision	F_1	recall	precision	F_1	recall	precision
supervised	0.719	0.617	0.86	0.765	0.685	0.866	**0.785**	**0.728**	0.852
unsupervised	0.704	0.647	0.771	0.727	0.672	0.79	0.739	0.692	0.793
unsupervised+regul	0.714	0.609	0.863	0.725	0.623	0.865	0.742	0.644	**0.877**

Table 2: Detection scores

- Better eddy detection of SST using methods (similar precision but increased recall)
- Performance drop of the unsupervised methods

State-of-the-art comparison

We use the interpolations benchmark of the Ocean Data Challenge 2020 to compare our methods. The considered metrics are the following :

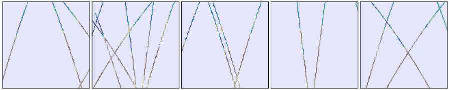
- μ and σ_t (in cm) are respectively RMSE and its temporal standard deviation
- $\lambda_x(^\circ)$ and λ_t(days) are respectively the smallest half resolved spatial and the temporal wavelength
- μ_u and μ_v are the RMSE on northward and eastward surface currents

Method	SST	SUP	μ	σ_t	λ_x	λ_t	μ_u	μ_v
DUACS	✗	✗	4.89	3.02	1.42	12.08	16.8	16.2
DYMOST	✗	✗	5.18	3.05	1.35	11.87	16.8	16.8
MIOST	✗	✗	4.21	2.5	1.34	10.34	14.9	14.5
BFN	✗	✗	4.7	2.73	1.23	10.64	15.1	15.3
4DVarNet	✗	✓	3.26	1.73	**0.84**	7.95	13.1	12.8
MUSTI	✓	✗	3.12	1.32	1.23	**4.14**	12.2	14.2
ABED-SSH	✗	✓	3.75	2.0	1.21	8.74	13.3	13.5
ABED-SSH	✗	✗	4.06	2.19	1.32	9.29	13.7	15.1
ABED-SSH-SST	✓	✓	**2.88**	**1.24**	0.95	4.51	**11.4**	**11.4**
ABED-SSH-SST	✓	✗	3.08	1.41	1.18	5.18	11.8	12.8

Results on observations

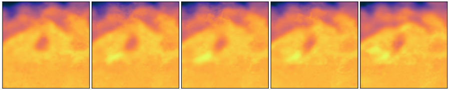
Sea Surface Height

$\mathbf{Y}_a$



Sea Surface Temperature

$\mathbf{Y}_b$



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

Table 3: μ and σ_t :RMSE and RMSE temporal standard deviation (in cm), λ_x : first half-resolved spatial wavelength (in km).

Comparing training strategies

We compare the methods on real-world observations : on the noisy measurements of a left aside satellite.

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

We compare the methods on real-world observations : on the noisy measurements of a left aside satellite.

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

We compare the methods on real-world observations : on the noisy measurements of a left aside satellite.

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 SST OSSE :

- Leads to better reconstruction on the training from simulation only
- 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**

Conclusion and perspectives

Conclusions

- SST is a key contextual information
- It is possible to learn a neural inversion from observations only
    - But it leads to a drop of performances
- Pre-training and fine-tuning neural networks leads to better reconstructions.
    - Especially for SST using networks.

Perspectives

Currently working on

- Forecast
- Simultaneous forecast and gridding

Other ideas

- Toward a global product : how to adapt this methods in the context of wider geographical areas ?
- Application of the method to other variables as well as contextual information or target data. (CHL)