Training end-to-end neural mapping schemes from simulation data for the reconstruction of global-scale sea surface fields

Daniel Zhu, Paul de Nailly, Daria Botvynko, Julien le Sommer, François Rousseau and Ronan Fablet

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Lab-STICC

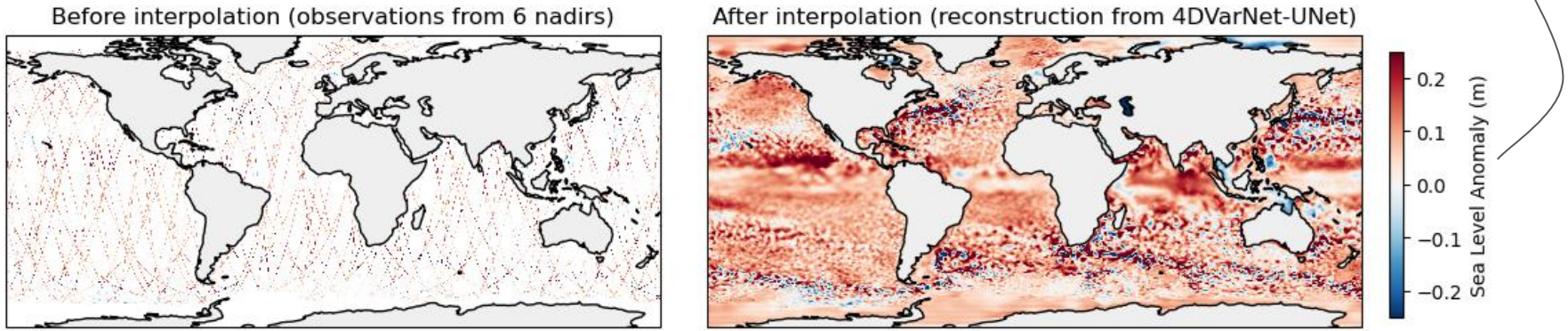
Introduction

Introduction | 1. Sea Level Anomaly from Satellite Altimetry

- Estimating ocean circulation is key for operations at sea and climate monitoring
- Key source of information : satellite altimetry measuring sea surface height (SSH)
- But altimeters provide a sparse and irregularly sampled representation of the SSH

  → so recovering the entire domain requires interpolation methods

Mean+Anomaly



Introduction | 2. Background

Name	Strategy	Technological readiness	Reference
DUACS	Based on optimal interpolation (OI), minimises a variational cost, Gaussian covariance model	Operational product	Taburet et al., 2019
MIOST	OI-based, wavelet-based covariance model, multiscale space-time feature		Ubelmann et al., 2021
NeurOST	Neural network, training on real data	State-of-the-art performance at global scale	Martin et al., 2024
4DVarNet	Bi-level optimisation, minimises a variational cost with a learnt prior,	State-of-the-art performance at regional scale	Febvre et al., 2024

Introduction | 3. Contribution

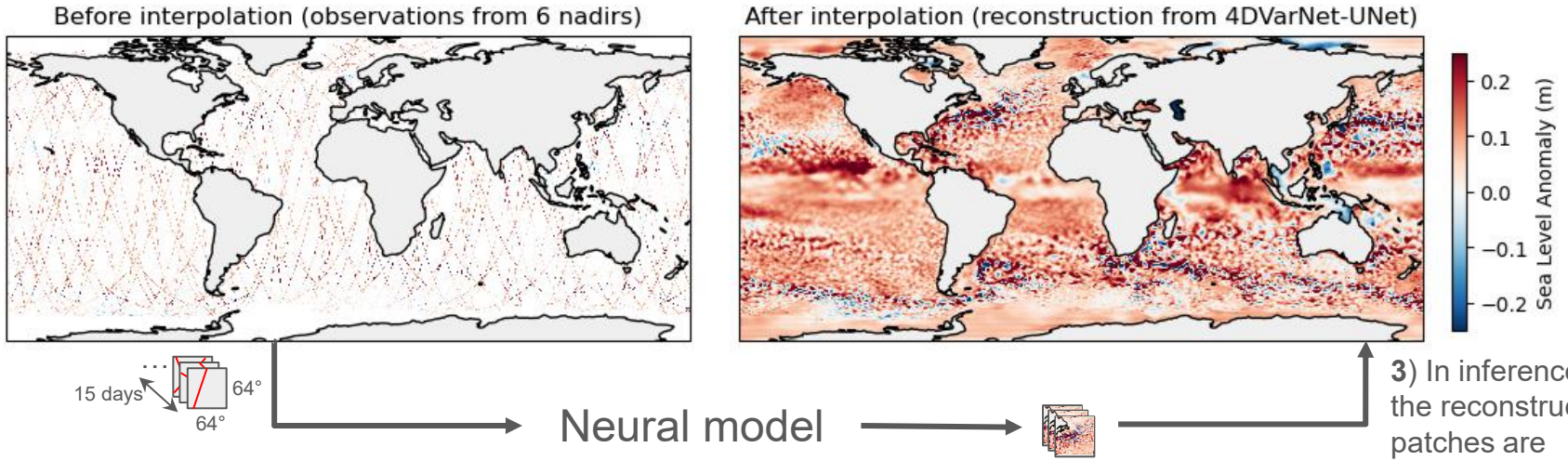
Question: **Is it possible to reach state-of-the-art performance in global mapping with neural models trained on simulated datasets?**

- We demonstrate that models trained on simulated observations generalise well on real data and can reach state-of-the-art performance
- We evaluate how to optimally draw pseudo-observations from ocean models for training neural mapping schemes
- We show that ensemble inferences from stochastic pseudo-observations improve significantly mapping performance.

Methodology

Methodology | 1. Framework & Data structure

Considering a global domain with a daily temporal resolution (daily average) and 1/4° spatial resolution,



1) For the training, we split the domain into patches of shape **15 × 64° × 64°** (days × latitude × longitude) and provide them to the model.

For inference, the patches' shape is **15 × 180° × 360°**.

2) The model predicts an estimation of the hidden state from the provided observations

3) In inference, the reconstructed patches are stitched together

In training, we just backpropagate the error of the patches

Methodology | 2. Evaluation framework

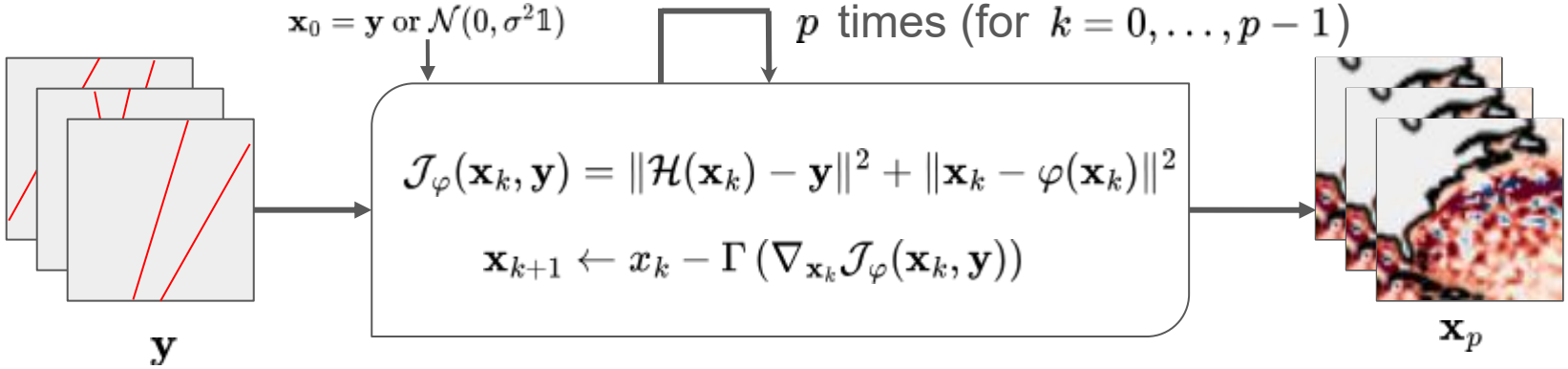
- We use the Ocean Data Challenge Global OSE Mapping (by Datlas & CLS)
- Sea Level Anomaly (SLA) from a constellation of 6 nadirs (Jason 3, Sentinel 3A, Sentinel 3B, Haiyang 2A, Haiyang 2B and Cryosat 2) flying in 2019
- Evaluation metrics based on Saral/Altika ($14 \cdot 10^6$ measurements): **RMSE** (in cm), **normalised RMSE** (μ, in %) & **effective spatial resolution** (λ, in km)

↑ QR code to the data challenge

Methodology | 3. Neural mapping schemes

Three models:

- **UNet**: 12 million parameters, 3 floors, 2D convolutions (3×3 kernel)
- **4DVarNets (**4DVarNet-ConvLSTM & 4DVarNet-Unet)



where φ : prior operator:

Γ : neural solver:

	4DVarNet-ConvLSTM (Febvre et al., 2024)	4DVarNet-UNet (Dwariwal & Nichol, 2021)
	Light weight CNN	UNet (guided diffusion)
	ConvLSTM	UNet (guided diffusion)
	570 000 params	56 million params

Methodology | 4. Training strategy

- Models are trained on the SLA, at 1/4° resolution from the **reanalysis Glorys12**, from 2010 to 2017 (validation on 2018)
- Spatial grid of shape : latitude × longitude = 680 × 1440
- The training loss involves two Mean Squared Error (MSE) terms:

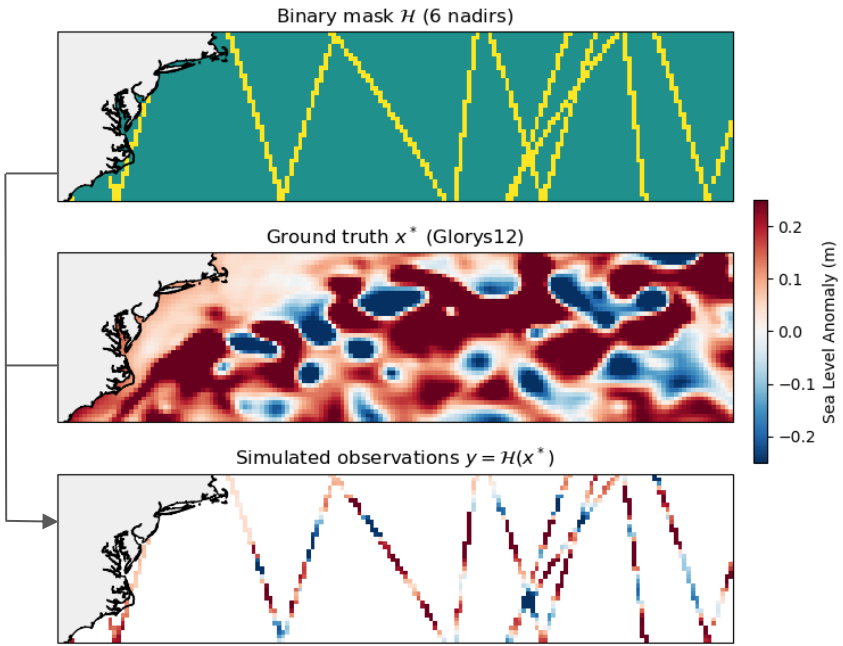
$$\mathcal{L}(\mathbf{x}^*, \hat{\mathbf{x}}) = \alpha \cdot \mathrm{MSE}(\mathbf{x}^*, \hat{\mathbf{x}}) + \beta \cdot \mathrm{MSE}(\nabla \mathbf{x}^*, \nabla \hat{\mathbf{x}})$$

where $\mathbf{x}^*$ and $\hat{\mathbf{x}}$ are respectively the ground truth and the predicted states and α and β are fixed scalars, balancing the two MSE.

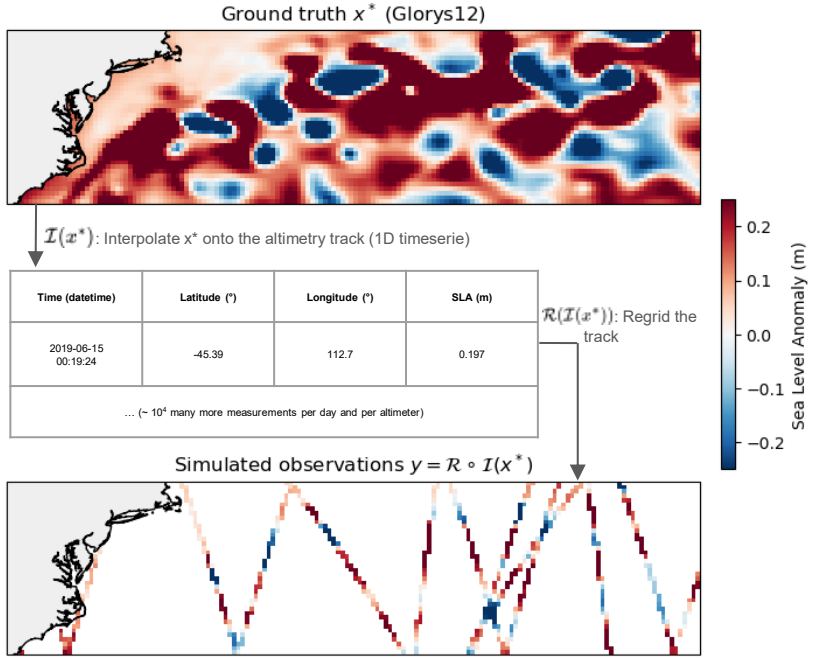
Methodology | 4. Training strategy

Two datasets to simulate the altimetry tracks (observations) are considered:

Dataset A

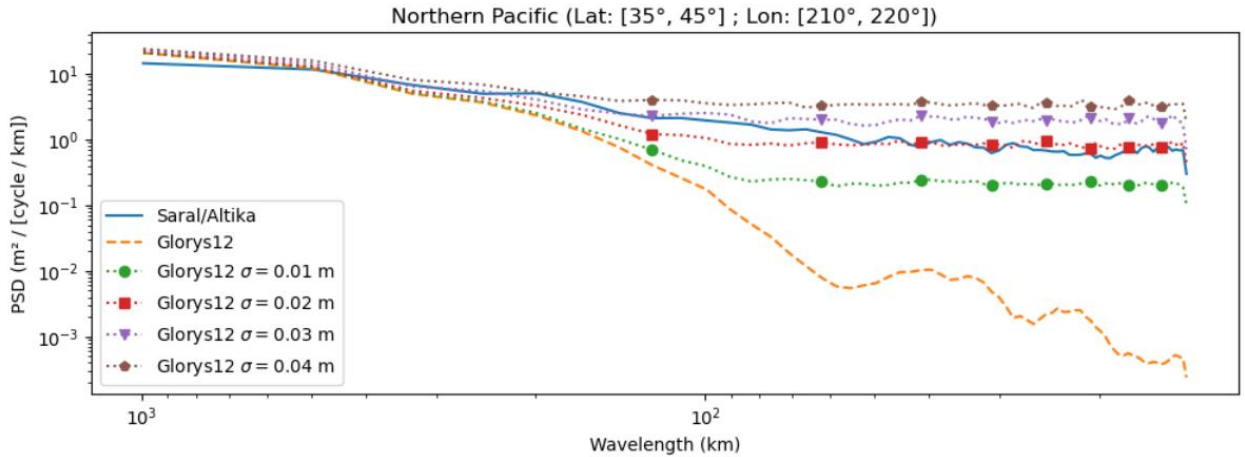


Dataset B



Methodology | 4. Training strategy

- Dataset B introduces noise to observation (standard deviation = 8.4 mm, which corresponds to 9.8 % of the SLA variability on global scale)
- We also consider altered versions of datasets A and B with a Gaussian additive noise scaled between 0 and 0.03 m



Methodology | 5. Ensemble inference

The inference output is averaged among n sub-inferences from the same input, with a Gaussian additive noise:

$$\hat{\mathbf{x}} = \frac{1}{n}\sum_{k=1}^{n} \Psi_\theta(\mathbf{y} + \eta_k) \qquad \text{with } \eta_k \sim \mathcal{N}(0, \sigma^2 \mathbb{1})$$

As opposed to the training (which is deterministic with a fixed seed), this approach is stochastic and reduces variance.

All inferences in our results sections are from this ensemble inference approach.

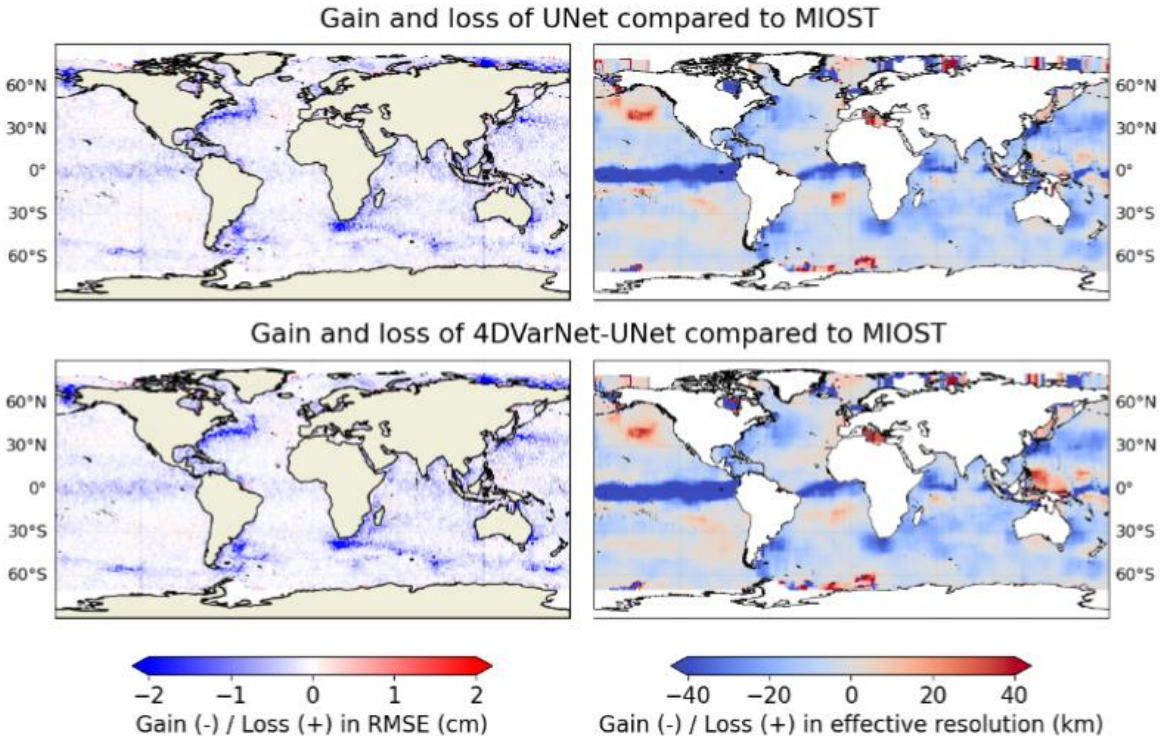
Results

Results | 1. Benchmarking against the state-of-the-art

	Global metrics			μ scores by regimes (%) ↑			
Methods	μ (%) ↑	RMSE (cm) ↓	λ (km) ↓	Coastal	High var.	Low var.	Equatorial
DUACS	68.78	4.06	216.83	57.95	77.74	69.23	68.76
MIOST	69.35	3.99	214.57	58.63	78.54	69.73	68.31
NeurOST SSH	70.40	3.85	207.42	60.55	79.53	70.45	69.66
UNet	70.82	3.80	202.74	60.54	81.2	70.77	69.68
4DVarNet-ConvLSTM	69.94	3.91	221.87	59.64	79.75	70.01	69.06
4DVarNet-UNet	**71.02**	**3.77**	**200.56**	**60.73**	**82.06**	**71.00**	**69.94**
NeurOST SSH-SST	*70.91*	*3.78*	*200.57*	*60.95*	80.71	*70.89*	*69.80*

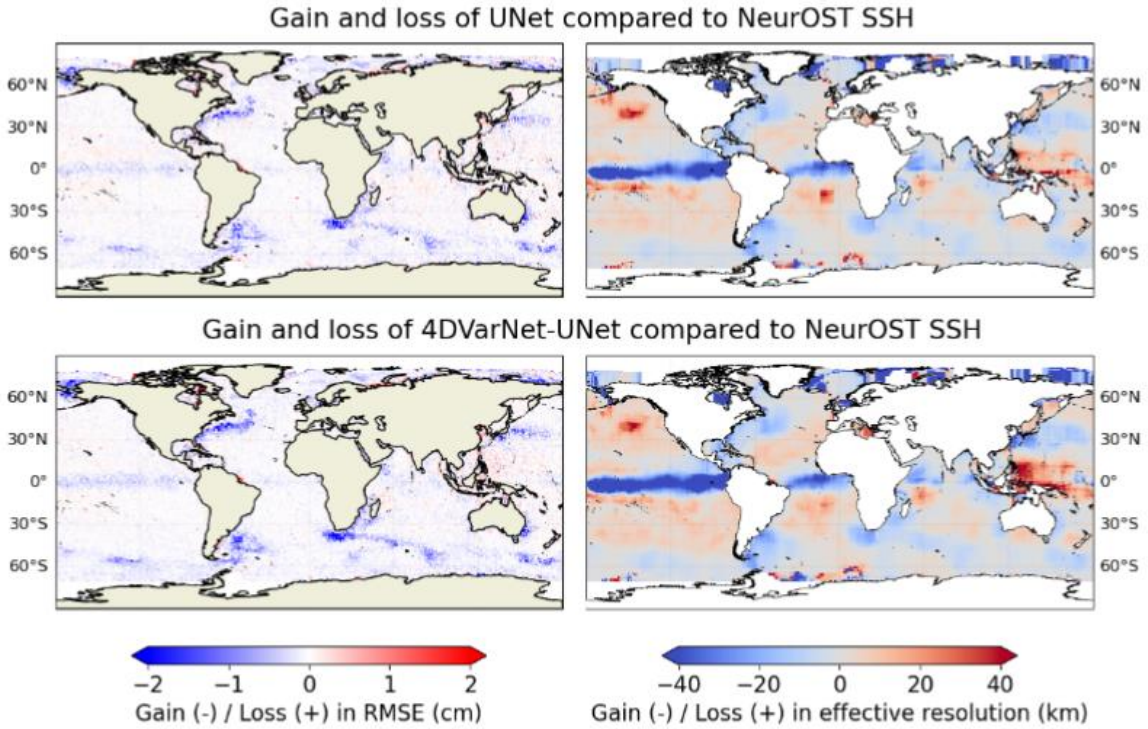
Results | 1. Benchmarking against the state-of-the-art

- We outperform the baseline MIOST in average
- Improvements are however not homogenous across the globe (worse in Pacific)



Results | 1. Benchmarking against the state-of-the-art

- Significant gains in equatorial band;
- Gains in high variability regions (Gulfstream, Kuroshio, etc): gradient terms from the training loss;
- Worse performance in low variability and coastal regions



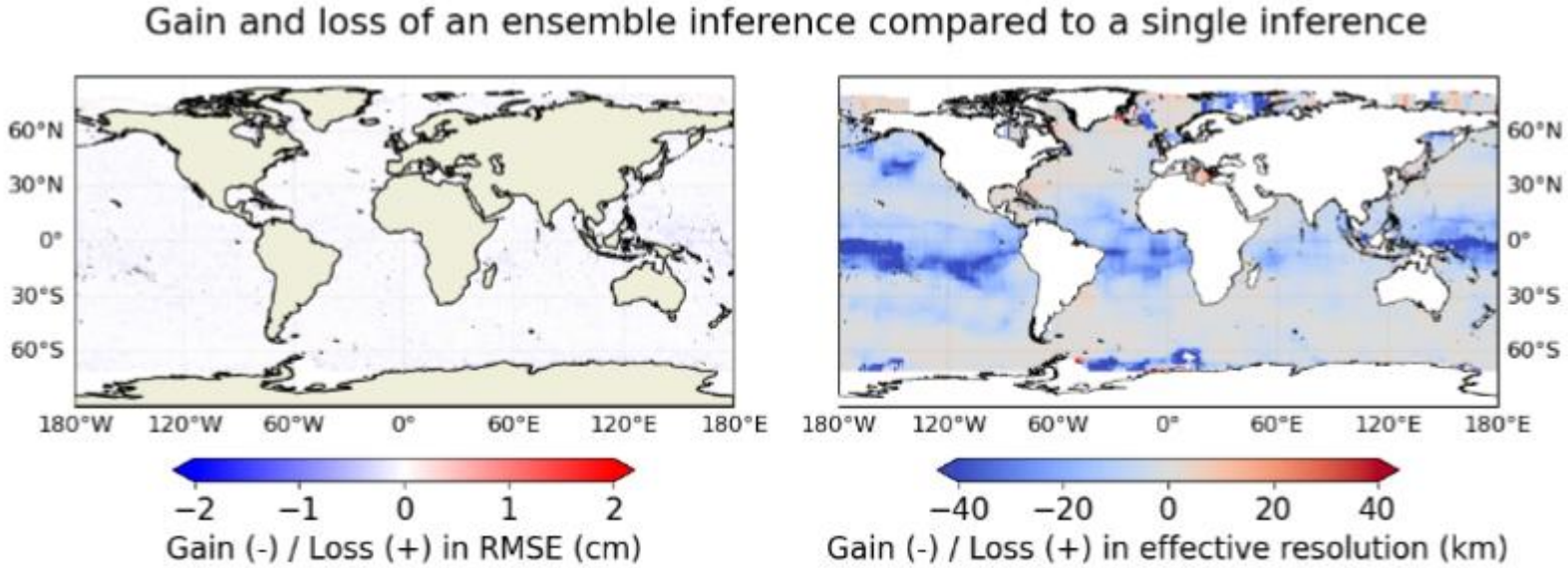
Results | 2. Impact of pseudo-observations used for training

	Global metrics			μ scores by regimes (%) ↑			
Training datasets	μ (%) ↑	RMSE (cm) ↓	λ (km) ↓	Coastal	High var.	Low var.	Equatorial
A	69.72	3.94	222.05	59.60	79.88	69.58	68.34
A + Gaussian noise	70.64	3.82	203.46	60.53	80.59	70.62	**69.71**
B	69.83	3.92	226.13	59.27	81.18	69.62	67.69
B + Gaussian noise	**70.82**	**3.80**	**202.74**	**60.54**	**81.28**	**70.77**	69.68

(**UNet** model) Training on different datasets.

- Dataset A. Binary mask applied on Glorys12
- Dataset B. Glorys12 interpolated onto altimetry tracks, then regridded

Results | 3. Importance of ensemble inference



(**UNet** model) Ensemble inference provides significant gains in effective resolution, on the equatorial band and low variability regions (but slight degradations in high variability regions)

Conclusion

Conclusion

- Training on simulated datasets leads to state-of-the-art mapping performance for real altimetry
- Significant impact of simulated trained dataset onto the mapping performance
- Weaker impact of the selected neural architecture (Unet vs. 4DVarNet)

Thank you for your attention!

Contact: daniel.zhu@imt-atlantique.fr

Appendix

Evaluation metrics

- Root Mean Square Error (**RMSE**), in cm, between the ground truth $\mathbf{x}^*$ and the reconstructed state $\hat{\mathbf{x}}$;
- **Normalised RMSE** μ , in %:

$$\mu = 100 \times \left(1 - \frac{\mathrm{RMSE}(\mathbf{x}^*, \hat{\mathbf{x}})}{\sigma_{\mathbf{x}^*}}\right) \tag{6}$$

  provides reconstruction performance regardless of the variability of the local dynamics;
- **Effective (spatial) resolution** λ , in km. It is the wavelength for which the Power Spectral Density (PSD) verifies:

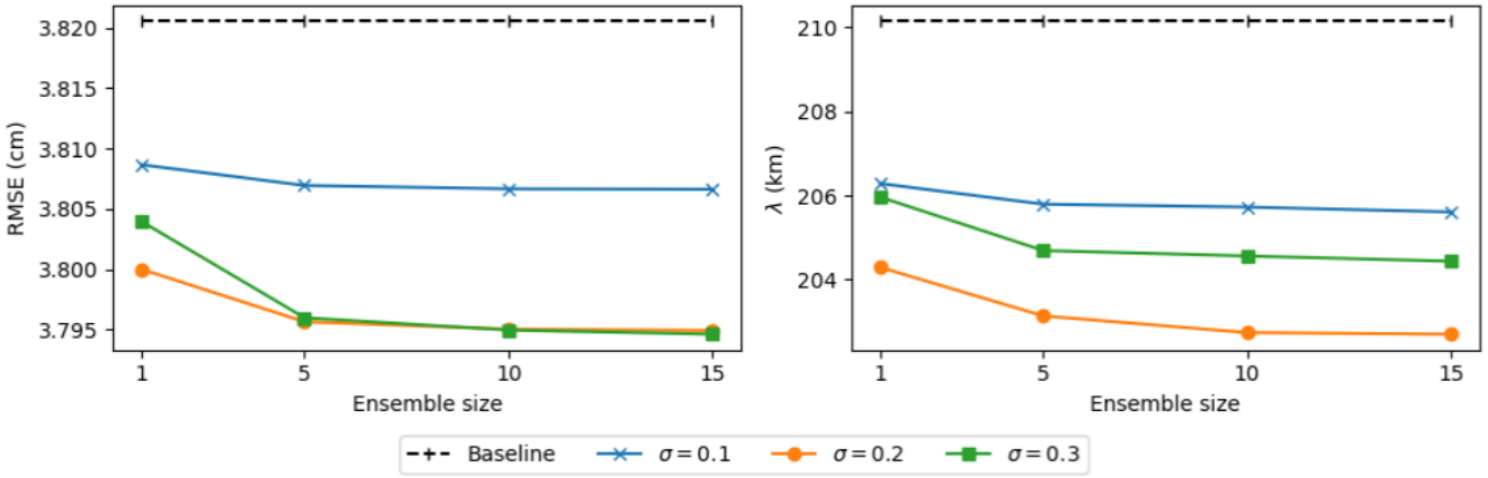
$$\frac{\mathrm{PSD}_\lambda(\mathbf{x}^* - \hat{\mathbf{x}})}{\mathrm{PSD}_\lambda(\mathbf{x}^*)} = 0.5$$

Sensibility analysis in inference

	Global metrics			μ scores by regimes (%) ↑			
Methods	μ (%) ↑	RMSE (cm) ↓	λ (km) ↓	Coastal	High var.	Low var.	Equatorial
Baseline	**70.82**	**3.80**	**202.74**	**60.54**	**81.28**	**70.77**	69.68
No LWE correction	69.82	3.92	207.87	58.30	80.79	70.16	68.60
Unfiltered SLA	70.78	**3.80**	204.34	**60.54**	81.24	70.71	**69.71**

(**UNet** model) Baseline = *with* LWE correction, *filtered* SLA.

Sensibility analysis in inference



(**UNet** model)