OceanBench-STOF

Short-term end-to-end neural forecasts of sea surface dynamics

Daria Botvynko (IMT Atlantique), Pierre Haslée (IMT Atlantique),
Bertrand Chapron (Ifremer), Clément de Boyer Montégut (Ifremer), Lucile Gaultier (ODL),
Julien Le Sommer (IGE), Ronan Fablet (IMT Atlantique)

Context

GLONET: MERCATOR'S END-TO-END NEURAL FORECASTING SYSTEM

A PREPRINT

Anass El Aouni * Quentin Gaudel Charly Regnier Simon Van Gennip

Marie Drevillon Yann Drillet Jean-Michel Lellouche

Mercator Ocean International

OCEANBENCH: The Sea Surface Height Edition

J. Emmanuel Johnson* CNRS UMR IGE johnsonj@univ-grenoble-alpes.fr

Quentin Febvre* IMT Atlantique quentin.febvre@imt-atlantique.fr

Anastasia Gorbunova CNRS UMR IGE

Sammy Metref DATLAS

Maxime Ballarotta CLS

Julien Le Sommer CNRS UMR IGE

Ronan Fablet IMT Atlantique

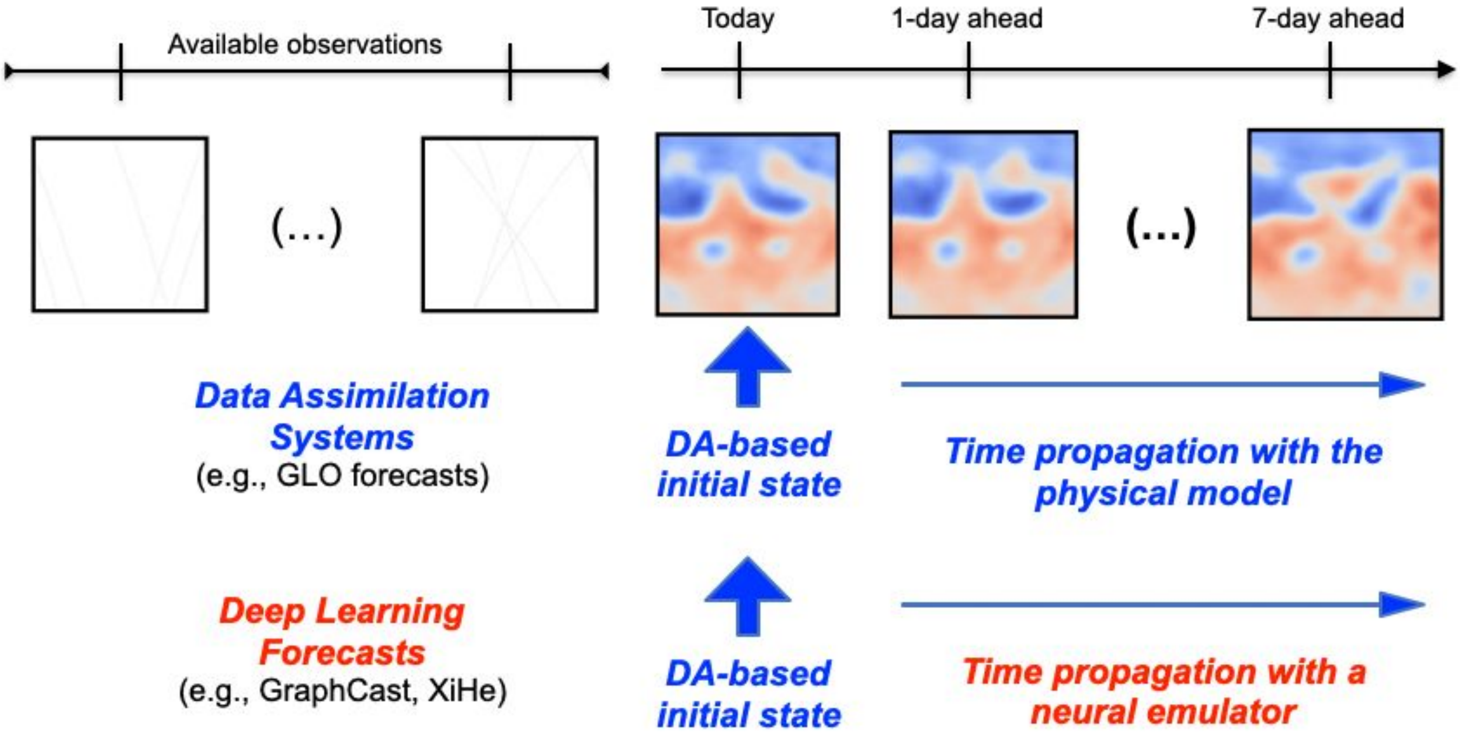
XiHe: A Data-Driven Model for Global Ocean Eddy-Resolving Forecasting

iang Wang, Renzhi Wang, Ningzi Hu, Pinqiang Wang, Peng Huo, Guihua Wang, Huizan Wang, ang Wang, Junxing Zhu, Jianbo Xu, Jun Yin, Senliang Bao, Ciqiang Luo, Ziqing Zu, Yi Han, Weimin Zhang, Kaijun Ren, Kefeng Deng, Junqiang Song

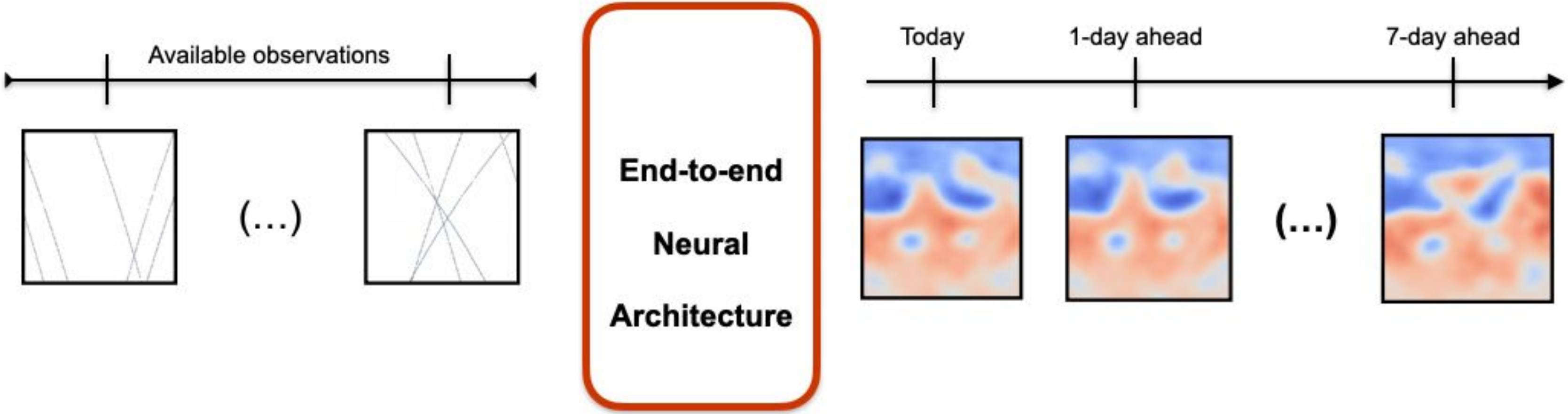
ORCAst: Operational High-Resolution Current Forecasts

Pierre Garcia, Inès Larroche, Amélie Pesnec, Hannah Bull, Théo Archambault, Evangelos Moschos, Alexandre Stegner, Anastase Charantonis, and Dominique Béréziat

Ocean Forecasting Systems



End-to-end neural forecasts

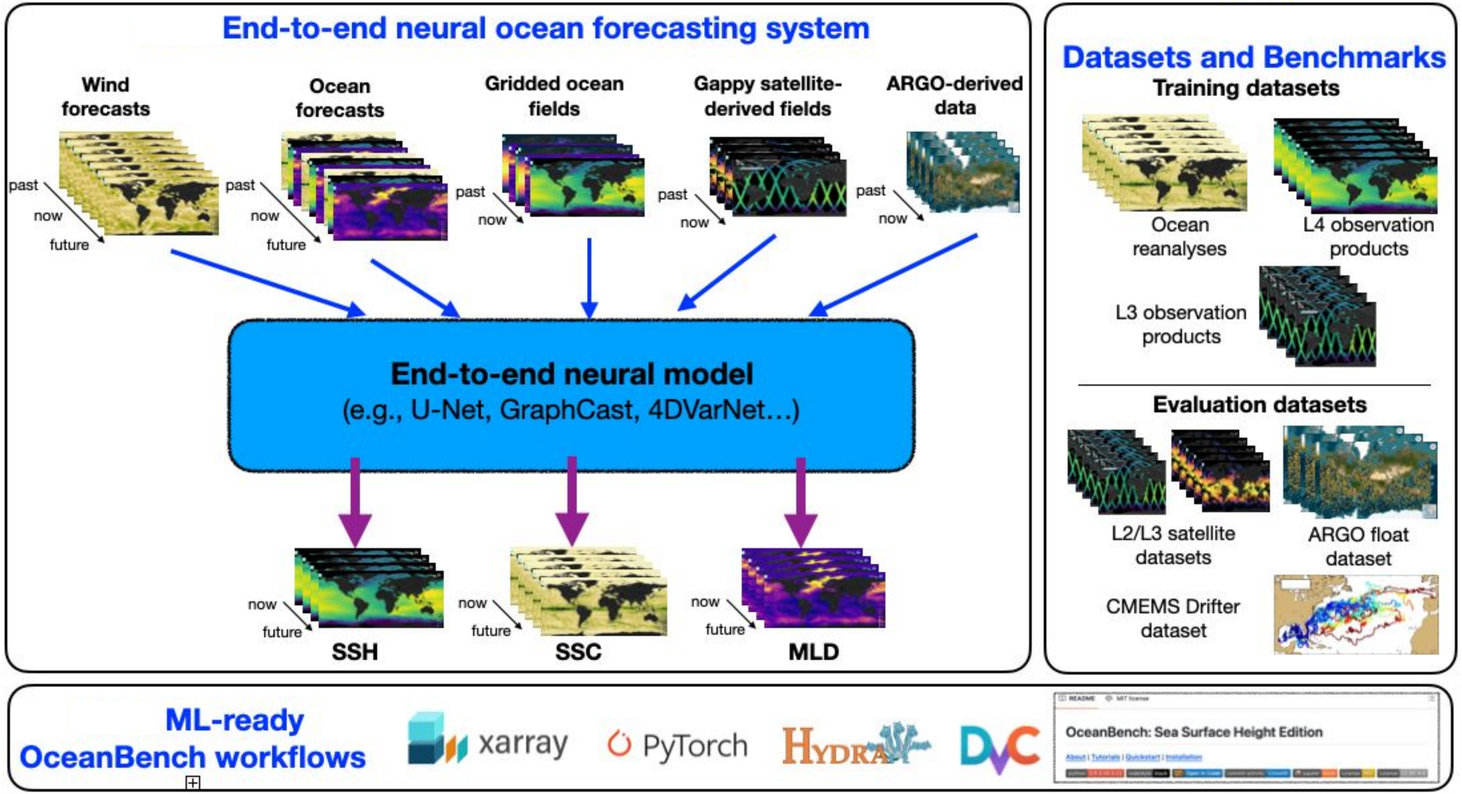


Objectives

Design, evaluation and benchmarking
of *end-to-end neural architectures for short-term ocean forecasting*.

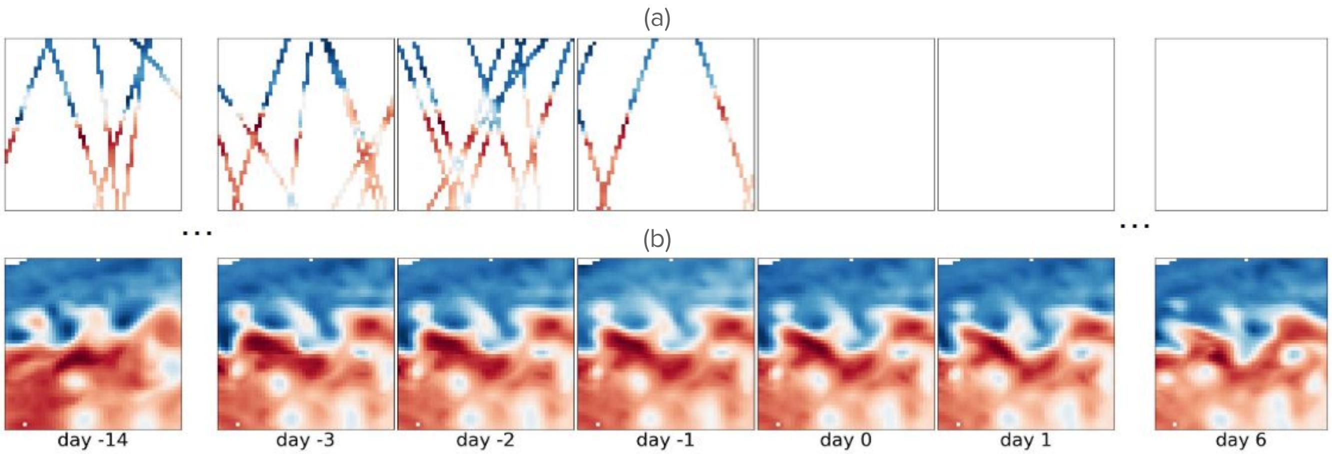
Can end-to-end approaches outperform state-of-the-art forecasting systems?

Overall approach



Proposed training workflow

- Target: from reference SLA dataset [t, ..., t + 6]
- Input: SLA L3 [t-14, ... , t-1]
- Globally, 1/4°, daily



Simulated / real SLA observations

Reference SLA (from a model or an observations-based product)

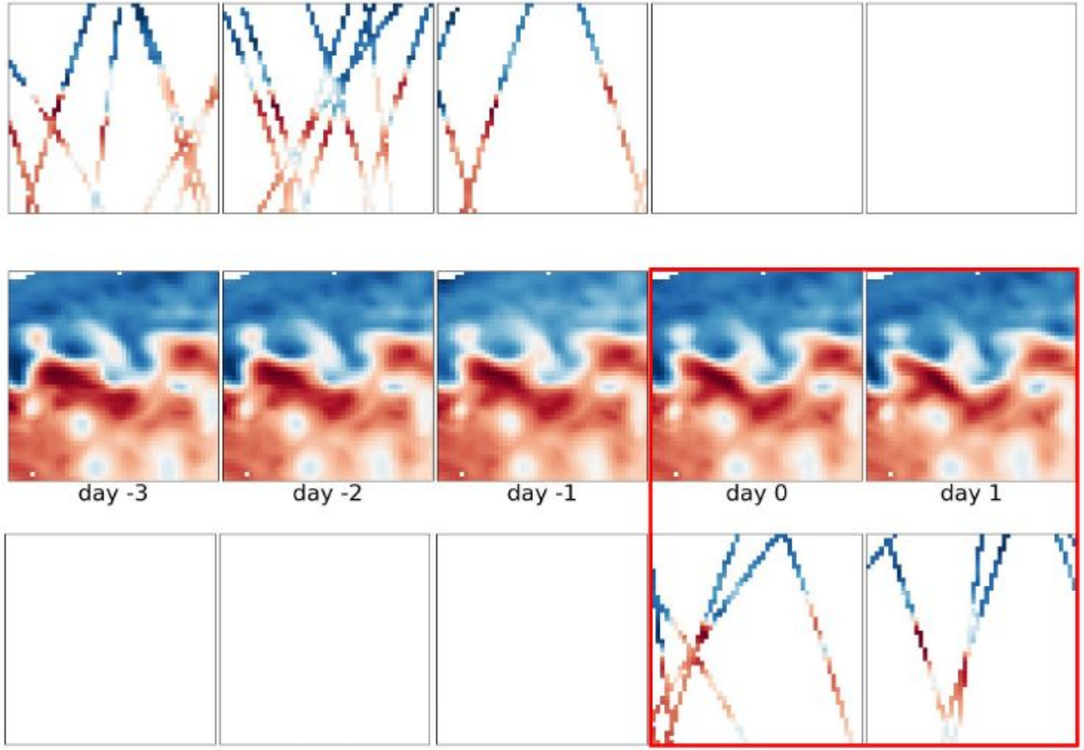
Evaluation procedure

Input: Nadirs / SWOT KaRIn

Trained neural forecast model

Output: gap-free future SLA

Reference: SARAL / AltiKa



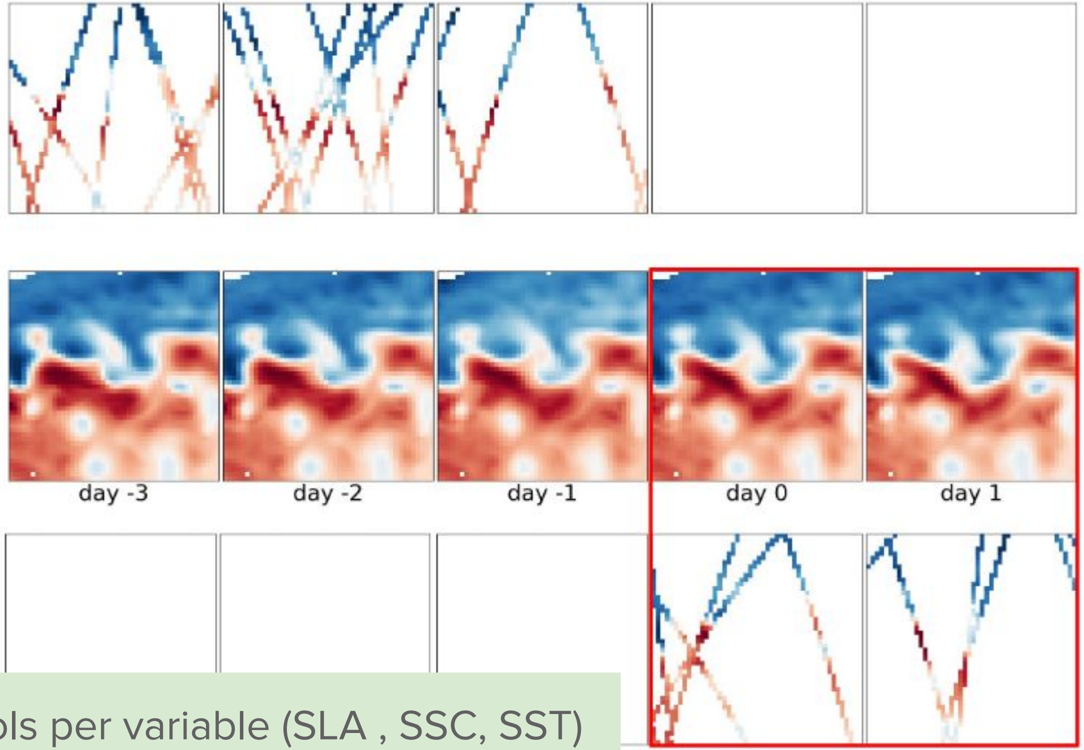
Evaluation procedure

Input: Nadirs / SWOT KaRIn

Trained neural forecast model

Output: gap-free future SLA

Reference: SARAL / AltiKa



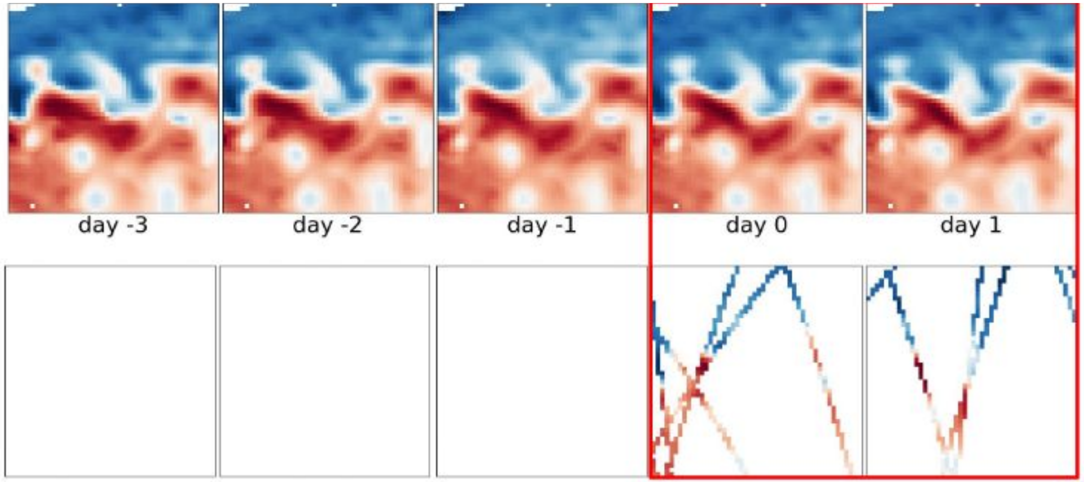
Metrics + Visualization tools per variable (SLA , SSC, SST)

Evaluation procedure

$$nRMSE\ score\ =\ 1\ -\ \frac{\sqrt{\frac{1}{N}\Sigma(\hat{y}_i - y_i)}}{\sqrt{\frac{1}{N}\Sigma y_i^2}}$$

,where $\hat{y}$ corresponds to predicted SLA values, y corresponds to real SLA values (from altimeters) and N is the number of samples.

Output: gap-free future SLA



Reference: SARAL / AltiKa

** and SSC-based scores, with drifters reference dataset*

2023, Baseline operational forecast & Neural emulators vs End-to-end neural forecasts

	nRMSE score, SLA related			% Vel. Magnitude, SLA related			% Vel. Angles, SLA related			Average effective resolution, km SLA related		
	0	3	5	0	3	5	0	3	5	0	3	5
GLO12	0.9119	0.9059	0.9016	72.0600	71.2900	70.8600	60.6300	56.9500	54.6200	326.1480	364.7801	397.1354
GloNet	0.9062	0.9133	0.9110	71.0800	71.3300	71.5000	58.3000	61.3700	59.6500	339.4132	310.3503	316.5479
XiHE	0.9119	0.8426	0.8423	72.0600	60.0400	60.3900	60.6300	34.5100	34.6700	326.1480	622.1721	607.0198
4dVarNet OSSE	0.9364	0.9333	0.9283	72.5300	71.9900	68.8700	69.4300	68.3200	64.8900	209.8130	219.3075	256.8074
4dVarNet OSE	0.9364	0.9339	0.9295	72.6200	71.1900	69.0000	71.5000	70.6600	66.6600	211.5023	224.5830	267.8910
Unet-17M OSSE	0.9316	0.9256	0.9216	71.8000	68.9200	66.8700	63.4900	61.7200	60.1200	240.3432	244.9968	253.6146
Unet-17M OSE	0.9384	0.9376	0.9382	70.5400	70.3600	70.3500	73.0400	72.9600	72.7700	210.7299	211.1087	210.6981

−4 −3 −2 −1 1 2 3 4
nRMSE score, SLA related

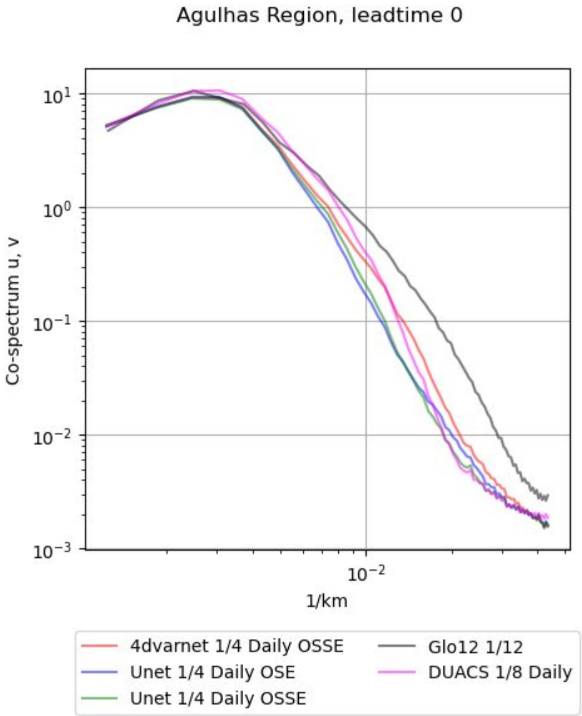
−12 −10 −2 −1 1 2 10 12
% Vel. Magnitude, SLA related

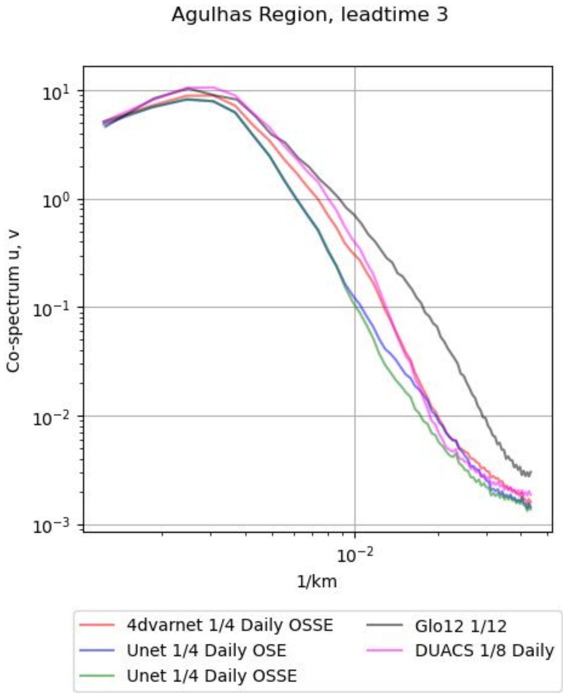
−30 −25 −20 −1 1 20 25 30
% Vel. Angles, SLA related

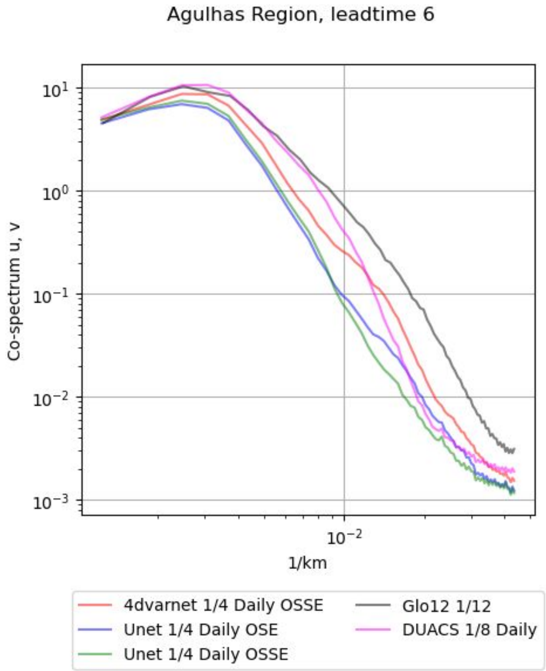
−40 −35 −25 −1 1 25 35 40
Average effective resolution, km SLA related

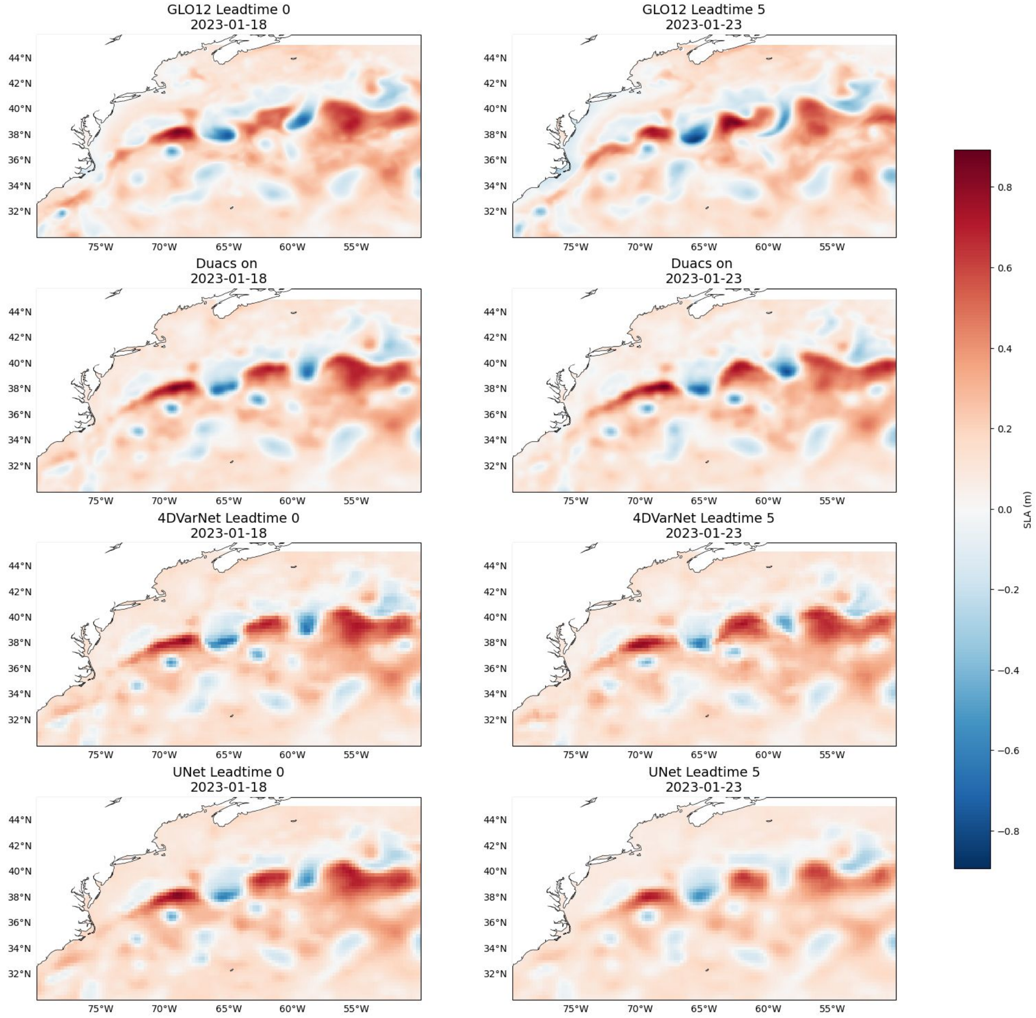
Leadtime → Leadtime → Leadtime → Leadtime →

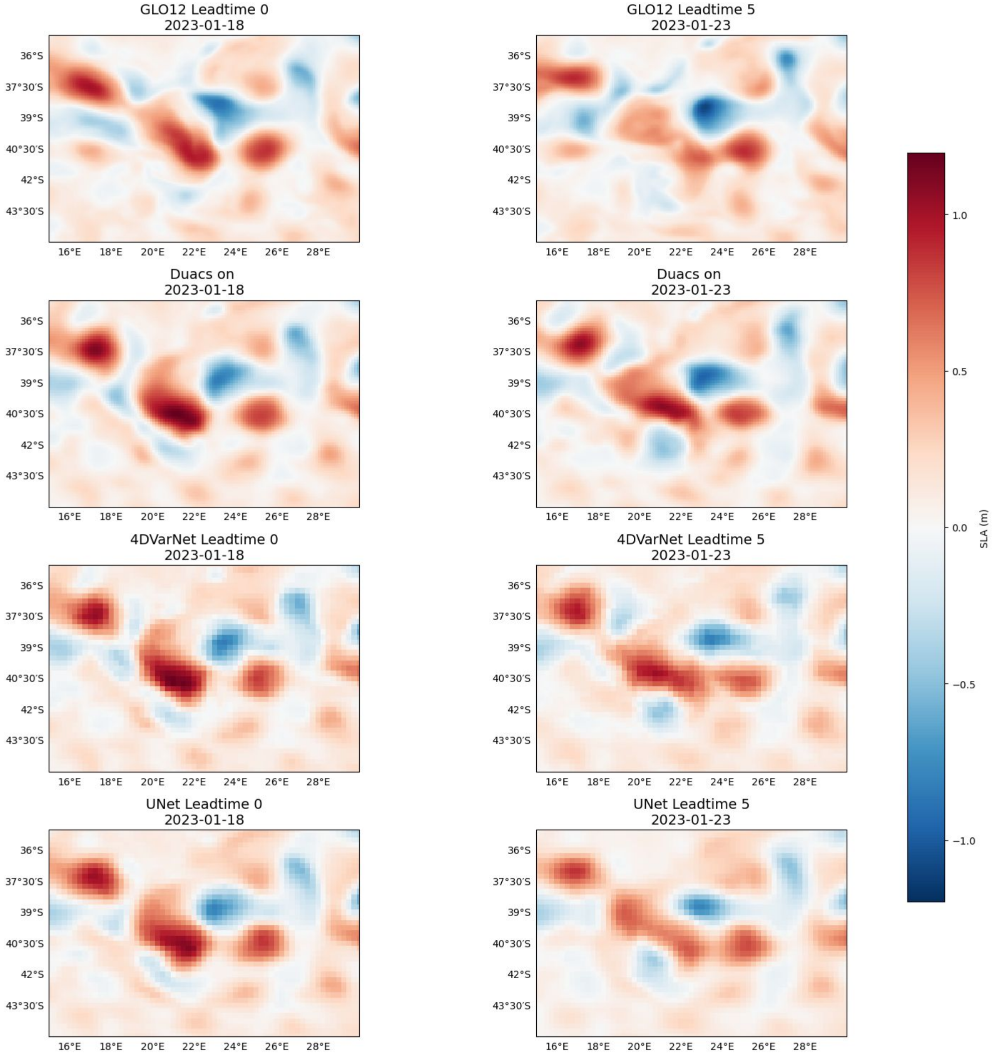
2023, Baseline operational forecast & Neural emulators vs End-to-end neural forecasts





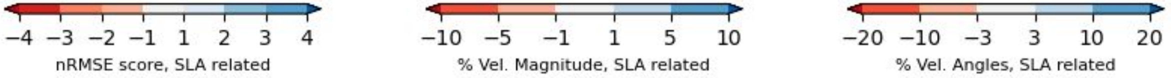



GLO12 Leadtime 0
2023-01-18
GLO12 Leadtime 5
2023-01-23
Duacs on
2023-01-18
Duacs on
2023-01-23
4DVarNet Leadtime 0
2023-01-18
4DVarNet Leadtime 5
2023-01-23
UNet Leadtime 0
2023-01-18
UNet Leadtime 5
2023-01-23
SLA (m)


GLO12 Leadtime 0
2023-01-18
GLO12 Leadtime 5
2023-01-23
Duacs on
2023-01-18
Duacs on
2023-01-23
4DVarNet Leadtime 0
2023-01-18
4DVarNet Leadtime 5
2023-01-23
UNet Leadtime 0
2023-01-18
UNet Leadtime 5
2023-01-23
36°S
37°30′S
39°S
40°30′S
42°S
43°30′S
16°E
18°E
20°E
22°E
24°E
26°E
28°E
SLA (m)
1.0
0.5
0.0
−0.5
−1.0

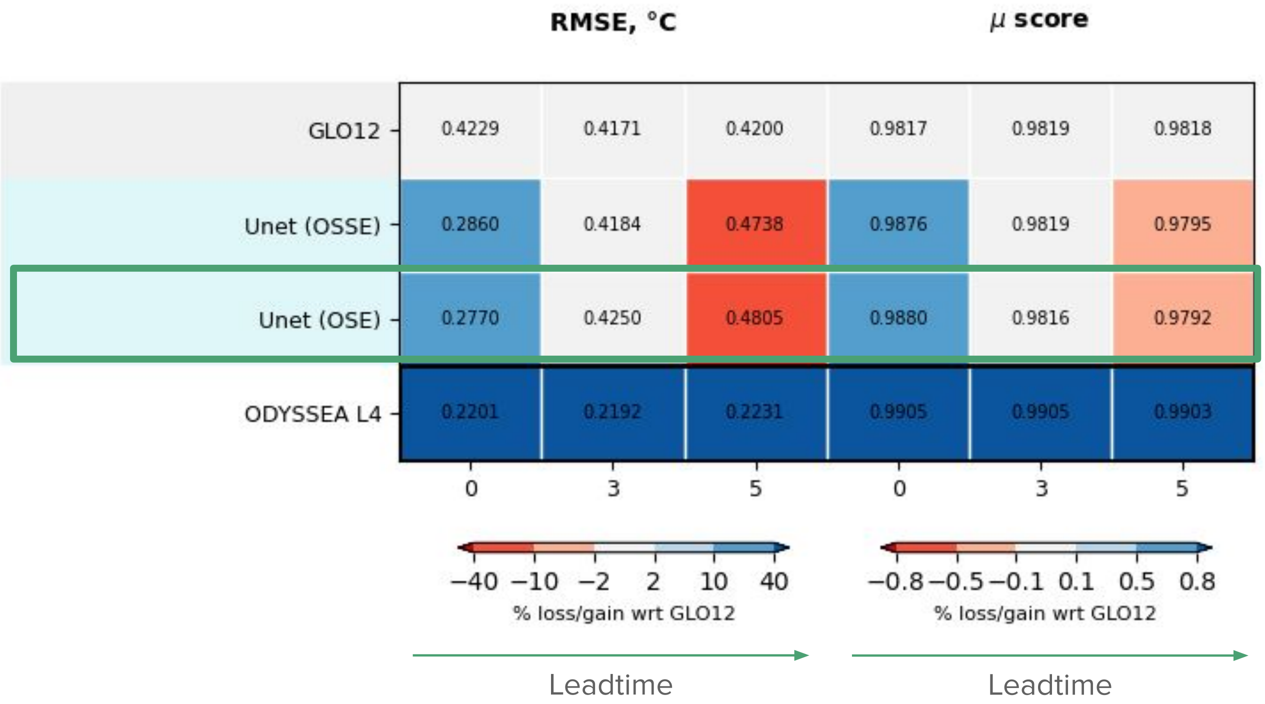
2024, Nadir VS Nadir+SWOT input

	nRMSE score, SLA related			% Vel. Magnitude, SLA related			% Vel. Angles, SLA related		
2023 reference GLO12	0.9119	0.9059	0.9016	72.0600	71.2900	70.8600	60.6300	56.9500	54.6200
WenHai	0.9132	0.9084	0.9013	69.2200	68.5200	68.1100	58.3500	56.0900	54.3700
4dVarNet OSSE (Nadir)	0.9358	0.9344	0.9291	69.4600	69.1000	66.0700	66.2100	64.9600	62.0100
Unet-17M OSSE (Nadir)	0.9343	0.9302	0.9255	69.7700	67.1800	65.0700	65.8000	64.7400	63.6000
4dVarNet OSSE (Nadir & SWOT)	0.9365	0.9357	0.9309	70.4100	70.3800	67.2100	67.7000	67.0400	64.0500
Unet-17M OSSE (Nadir & SWOT)	0.9339	0.9302	0.9255	70.7700	67.7800	65.3600	66.3500	64.8200	63.3600



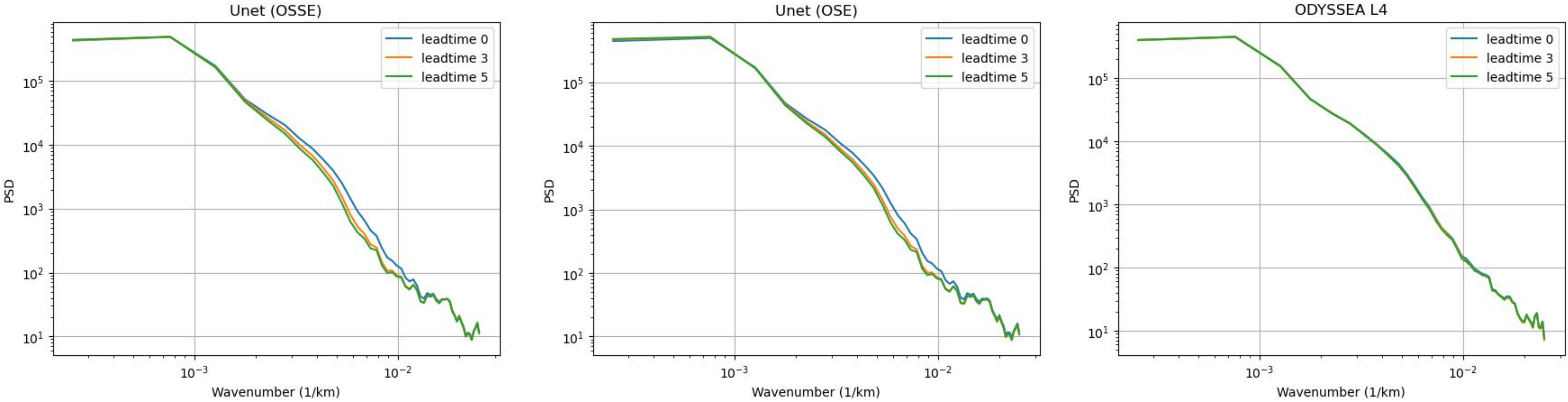
End-to-end SST forecasting

End-to-end SST forecasting



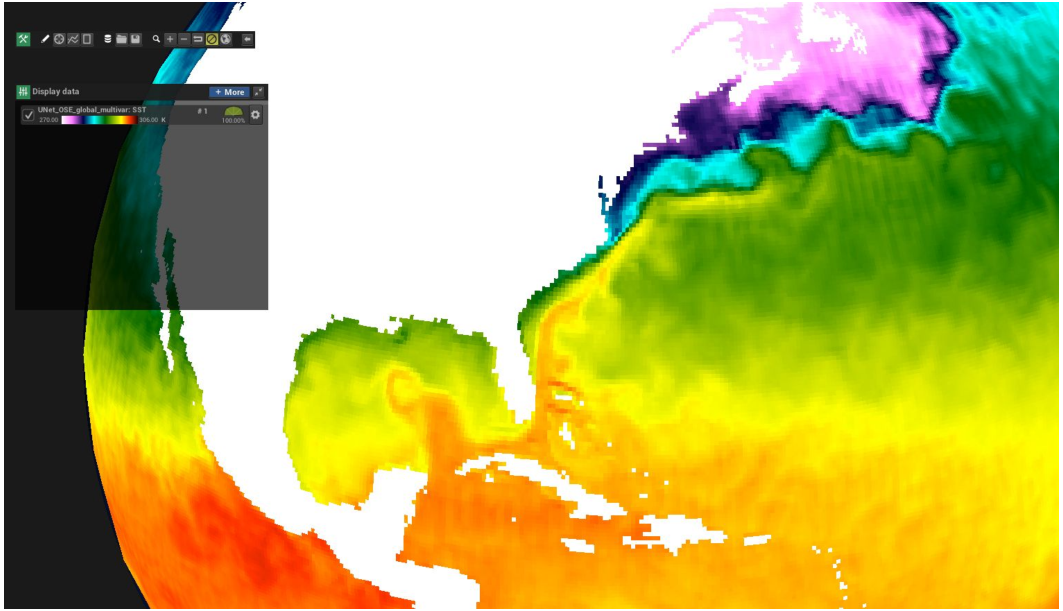
End-to-end SST forecasting

Gulf Stream, 2023

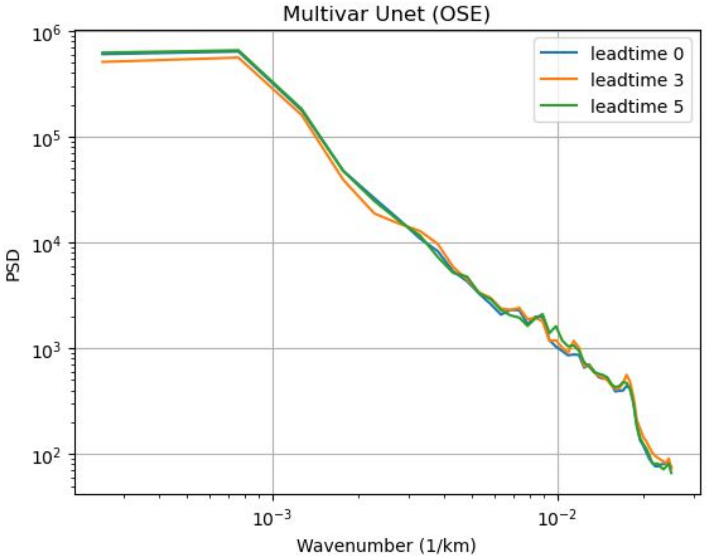


Towards end-to-end *multivariate* forecasting

Case-study: multivariate input globally (SLA + SST) & SST output.



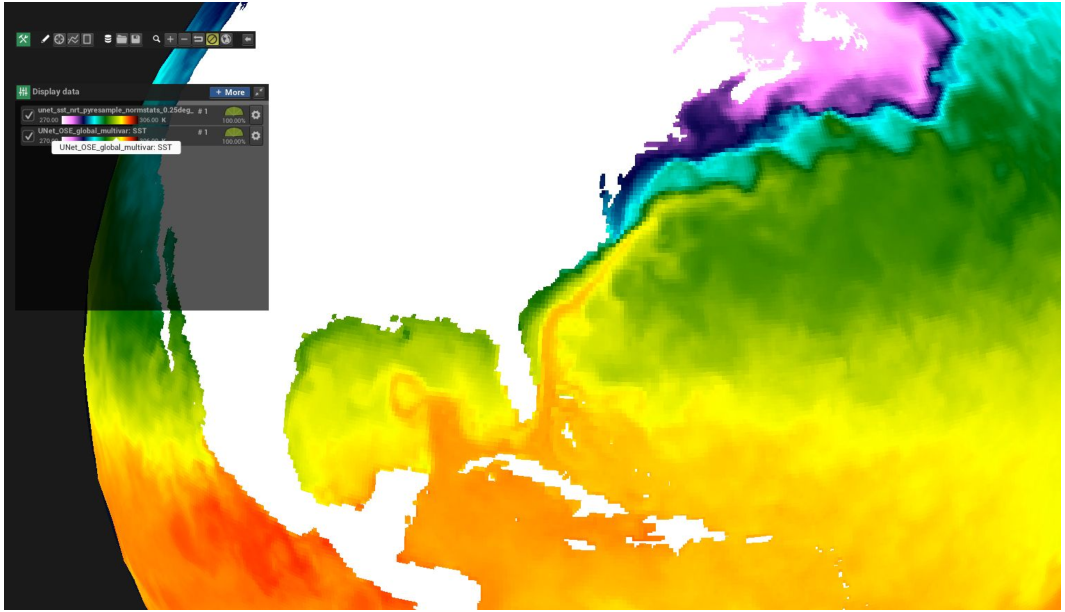
Predicted SST from SLA + SST inputs at leadtime 0, on 15 January 2023 in the Gulf Stream



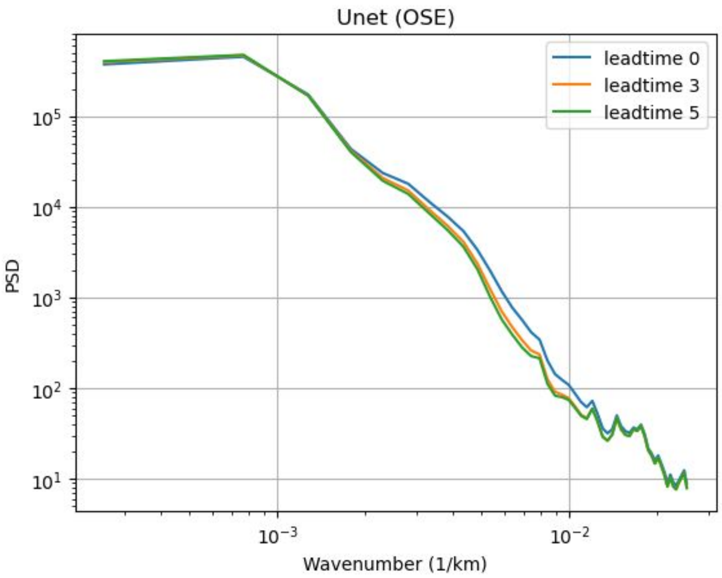
PSD of multivariate SST forecast across leadtimes, Gulf Stream, 2023.

Towards end-to-end *multivariate* forecasting

Case-study: multivariate input globally (SLA + SST) & SST output.



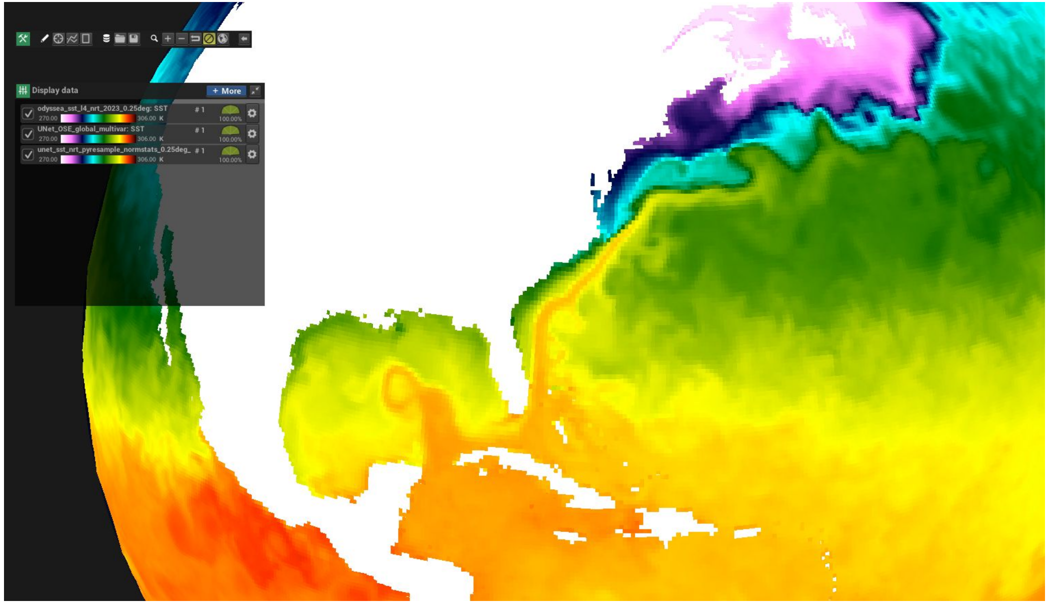
Predicted SST from SLA + SST inputs at leadtime 0 *vs predicted SST at the same leadtime using univariate SST as input, on 15 January 2023 in the Gulf Stream*



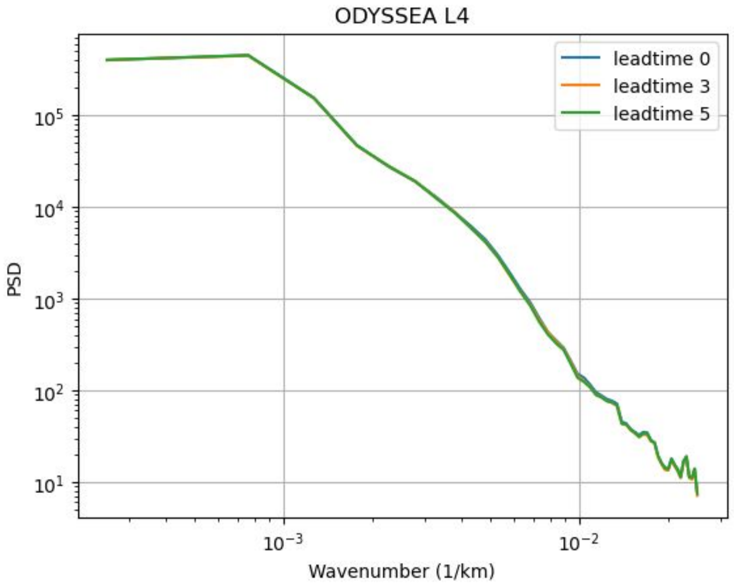
PSD of *univariate* SST forecast across leadtimes, Gulf Stream, 2023.

Towards end-to-end *multivariate* forecasting

Case-study: multivariate input globally (SLA + SST) & SST output.



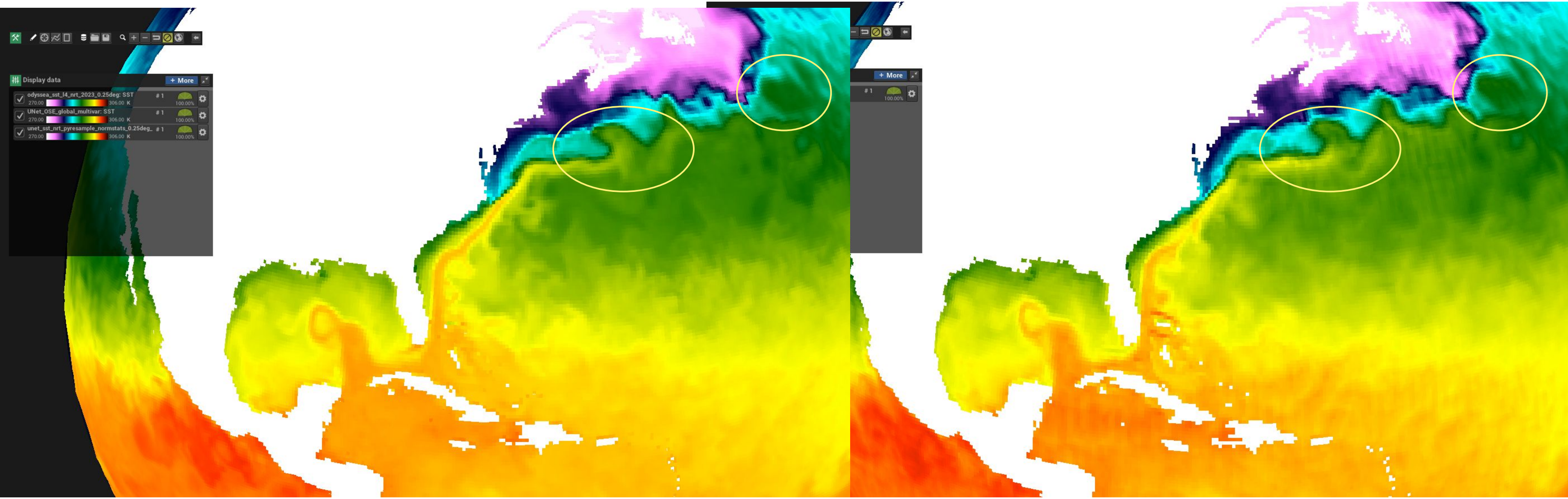
Predicted SST from SLA + SST inputs at leadtime 0
vs ODYSSEA L4,
on 15 January 2023 in the Gulf Stream



PSD of *ODYSSEA L4* 'forecast' across leadtimes, Gulf Stream, 2023.

Towards end-to-end *multivariate* forecasting

Case-study: multivariate input globally (SLA + SST) & SST output.



ODYSSEA L4 vs predicted SST from SLA + SST inputs at leadtime 0, *on 15 January 2023 in the Gulf Stream*

More details, Botvynko et al. 2025: https://arxiv.org/abs/2512.22152
(submitted at Env. Data Sciences)

arXiv > physics > arXiv:2512.22152

Search...
Hel

Physics > Atmospheric and Oceanic Physics

[Submitted on 15 Dec 2025]

Neural ocean forecasting from sparse satellite-derived observations: a case-study for SSH dynamics and altimetry data

Daria Botvynko (Lab-STICC_OSE, IMT Atlantique - MEE, IMT Atlantique), Pierre Haslée (Lab-STICC_OSE, IMT Atlantique - MEE, IMT Atlantique
Lucile Gaultier (ODL), Bertrand Chapron (LOPS), Clement de Boyer Montégut (LOPS), Anass El Aouni (MOi), Julien Le Sommer (IGE), Ronan F
(IMT Atlantique - MEE, Lab-STICC_OSE, ODYSSEY)

We present an end-to-end deep learning framework for short-term forecasting of global sea surface dynamics based on sparse satellite altimetry data. Building on two state-of-the-art architectures: U-Net and 4DVarNet, originally developed for image segmentation and spatiotemporal interpolation respectively, we adapt the models to forecast the sea level anomaly and sea surface currents over a 7-day horizon using sequences of sparse nadir altimeters observations. The model is trained on data from the GLORYS12 operational ocean reanalysis, with synthetic nadir sampling patterns applied to simulate realistic observational coverage. The forecasting task is formulated as a sequence-to-sequence mapping, with the input comprising partial sea level anomaly (SLA) snapshots and the target being the corresponding future full-field SLA maps. We evaluate model performance using (i) normalized root mean squared error (nRMSE), (ii) averaged effective resolution, (iii) percentage of correctly predicted velocities magnitudes and angles, and benchmark results against the operational Mercator Ocean forecast product. Results show that end-to-end neural forecasts outperform the baseline across all lead times, with particularly notable improvements in high variability regions. Our framework is developed within the OceanBench benchmarking initiative, promoting reproducibility and standardized evaluation in ocean machine learning. These results demonstrate the feasibility and potential of end-to-end neural forecasting models for operational oceanography, even in data-sparse conditions.

Learning-based methods for ocean modeling and prediction. (Review)

Durand et al., 2026 (submitted at ARMS)

Access the review:

Learning-Based Methods and the Future of Numerical Ocean and Sea Ice Modeling

Charlotte Durand[1], Daria Botvynko[2,3], Hugo Frezat[1], Davide Grande[4,5,6], Andrea Storto[4], David S. Greenberg[7], Ronan Fablet[2,3], Saïd Ouala[2,3], Julien Le Sommer[1]

[1]Univ. Grenoble Alpes, CNRS, INRAE, IRD, Grenoble INP, IGE, 38000 Grenoble, France, email : julien.lesommer@univ-grenoble-alpes.fr
[2]IMT Atlantique, UMR Lab-STICC, Brest, France
[3]Inria, Odyssey team, Brest, France
[4]National Research Council of Italy (CNR), Institute of Marine Sciences (ISMAR), Rome, 00133, Italy
[5]University School for Advanced Studies IUSS, Pavia, 27100, Italy
[6]Interdisciplinary Centre on Sustainability and Climate, Sant'Anna School of Advanced Studies, Pisa, 56127, Italy
[7]Model-Driven Machine Learning, Helmholtz-Zentrum Hereon, Institute of Coastal Systems, Geesthacht, Germany

Annual Review of Marine Science 2027. AA:1–26

https://doi.org/10.1146/((please add article doi))

Keywords

ocean modeling, sea-ice modeling, operational oceanography, machine learning, deep learning, artificial intelligence

Abstract

The field of operational oceanography is undergoing a significant evolution with the increasing integration of artificial intelligence (AI) methods, which are complementing and, in some cases, redefining traditional numerical modeling approaches. This review explores how AI methods—particularly model-based autoregressive emulators, hybrid modeling, and end-to-end model-free approaches—are reshaping the representation of ocean and sea-ice dynamics in operational systems. We focus on three key objects: sea-ice parameters, near-surface ocean properties, and the 3D ocean state, each characterized by distinct observational and dynamical challenges. While AI-driven innovations offer new opportunities for improved monitoring, forecasting, and uncertainty quantification, their long-term impact on operational systems remains uncertain, especially given the sparsity of subsurface observations and the complexity of ocean dynamics. By synthesizing recent advances and identifying open questions, this paper aims to guide the ocean modeling community toward a future where AI and physics-based approaches coexist synergistically.

ML emulators for ocean & sea-ice modeling and prediction. (Living Review)

Durand and Le Sommer, 2026

Machine-Learning-Based Emulators for Ocean and Sea Ice Modelling and Prediction : a living review

Scope & Context

This repository hosts a **living review** of machine-learning-based emulators for ocean and sea ice systems. The goal is to provide a curated, up-to-date list of peer-reviewed and preprint publications that advance the use of machine learning in emulating oceanic and sea ice dynamics from operational forecasts to climate modelling.

The review is a priori restricted to machine learning based emulators that can be run autoregressively, whether trained from observations, reanalyses or physics-based numerical models.

The review, which was initiated during the preparation of review article for ARMS in 2026, is community-driven and open to contributions.

How to Cite

If you use this review in your work, please cite it as:

> Durand and Le Sommer. (2026). *Living Review: Machine-Learning-Based Emulators for Ocean and Sea Ice Modelling and Prediction* (Version 1.0). Zenodo. DOI 10.5281/zenodo.18851354