Reconstructing ocean interior from surface observations and in-situ water column observations using data-driven approaches: a feasibility study based on model outputs

Aina Garcia Espriu, Cristina González Haro (ICM-CSIC), Fernando Aguilar (IFCA-CSIC)

IFCA Instituto de Física de Cantabria

Institut de Ciències del Mar

BEC Barcelona Expert Center

icat mar

Objective & Content

Table of Contents

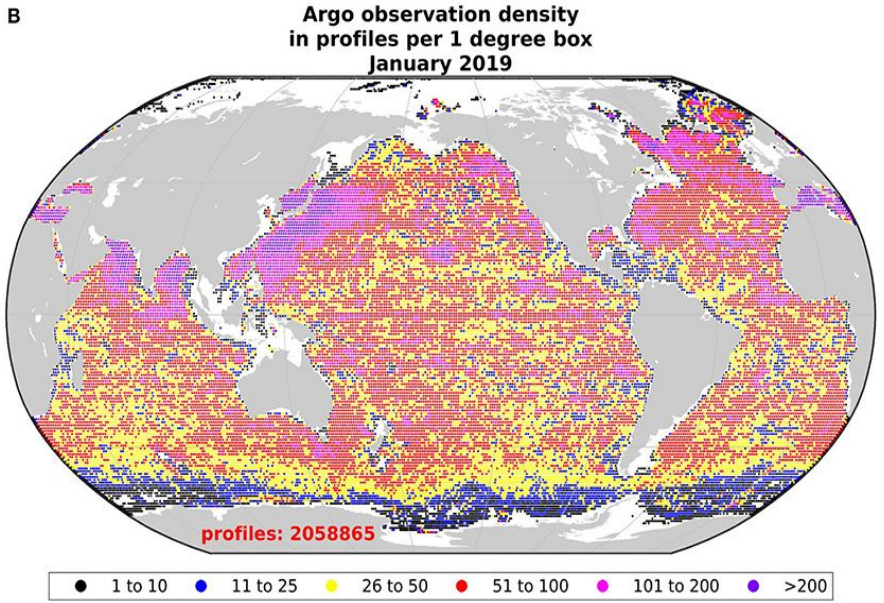
Main objective

To analyze the feasibility of a 4D reconstruction of the ocean from surface observations, ideally obtained from satellite sensors and vertical profiles of in-situ observations using a simulated observation system derived from numerical models.

Here, we present the Salinity reconstruction but a similar work was also done for Temperature.

How is the ocean measured? (simplified version)

- ★ The surface physical variables can be measured daily through satellite observation.
- ★ In depth measurements can be taken at specific points. It is expensive and can not cover the complete globe. We use buoys and profilers.
- ★ Time series longer than 10 years. Vertical and surface measurements. (Temperature, salinity, currents, SSH...)
- ★ Low daily in-situ resolution. If we aggregate the complete time series we have a better sampling of the ocean.

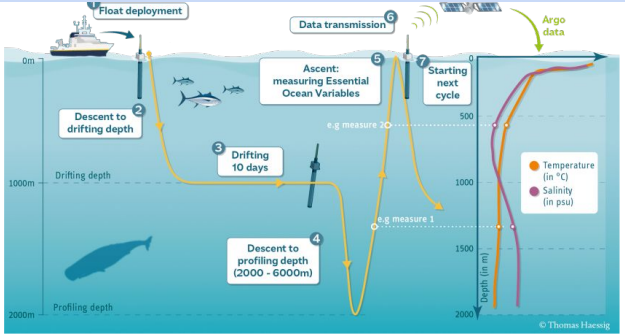


Wong, A. P. et al.. (2020). Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats. Frontiers in Marine Science, 7, 568494.

Available data

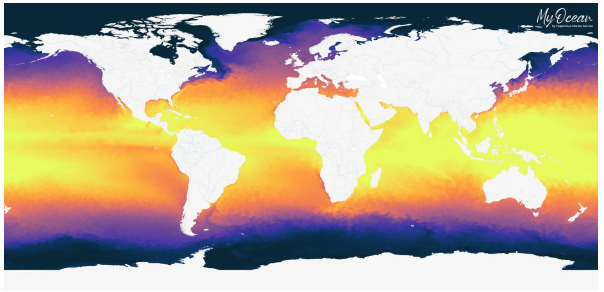
Argo profilers

- ~400 daily profiles (global)
- One measurement every 2m up to 2000m depth (most of them)
- Salinity/temperature
- 10-day cycle length
- From 2000 to current date



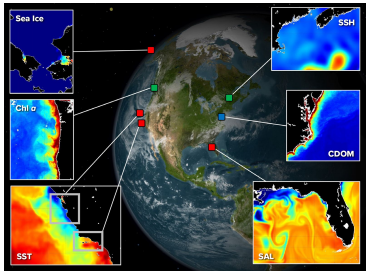
CMEMS reanalysis (OGCM)

- Up to 6 km depth
- From 1993 to 2022
- 0.25º x 0.25º resolution
- Salinity, temperature, currents, SSH, winds…
- Daily maps



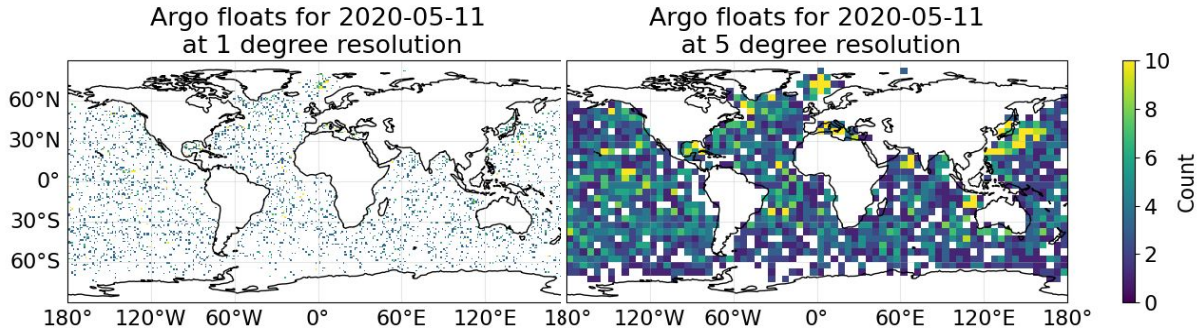
Satellite products

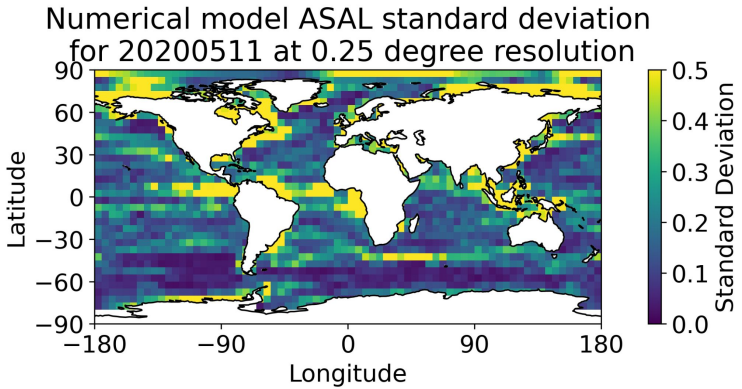
- Daily maps (aggregated)
- Resolution from 0.25ºx0.25º
- Different variables (SSS, SST, SSH, Currents, Ocean color…)

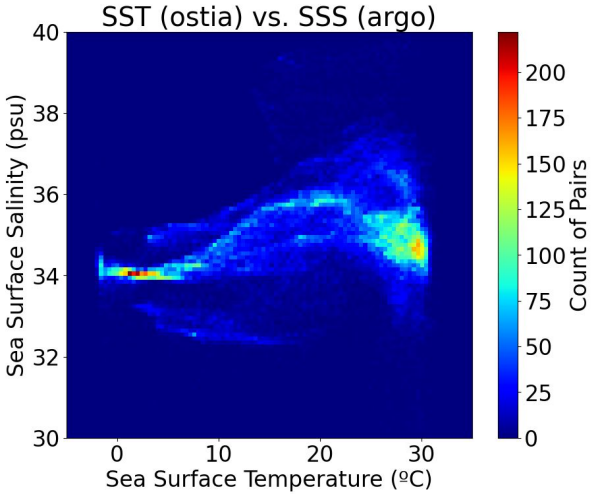


What kind of model do we need?

- ★ 5ºx5º resolution to have enough coverage of the globe.
- ★ Variability > 0.5 psu in salinity if we take 5ºx5ºboxes.
- ★ **Point-based methodologies to maintain the spatial resolution and smaller scale dynamics.**

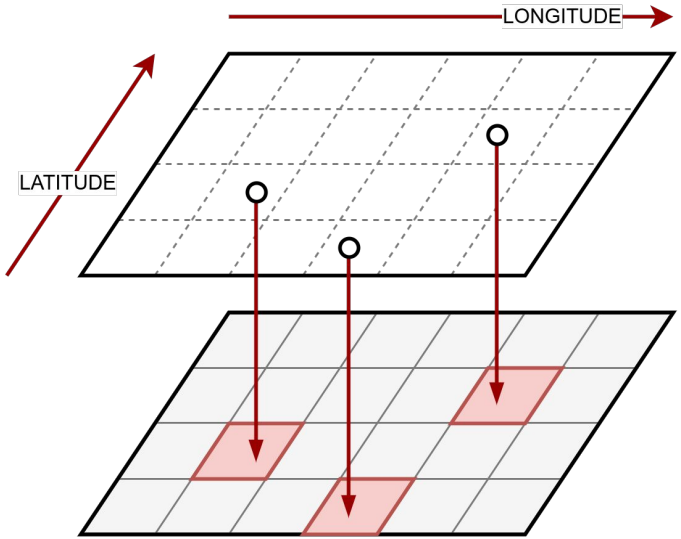






- ★ Water density formula combines salinity and temperature.
- ★ Different regions have different water density.
- ★ **Able to model non-linear relations.**

How did we test the feasibility? [Dataset]



ARGO measurements

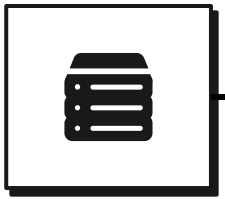
A: Points in reanalysis grid seen by ARGO

B: Non-sampled points

We use the data from the CMEMS Reanalysis to **simulate the current sampling** made by ARGO.

- ★ We use the **dataset A** to train/validate our models (train/test split)
- ★ We use the **dataset B** to validate predictions of non-sampled points
- ★ This approach allows us to validate that:
  - ○ The models do not overfit on points seen by in-situ (currents make them drift to certain positions)
  - ○ We can validate how the model extrapolates to a global scale

Training Structure



OGCM data at ARGO positions

For each position:
Salinity vertical profile (predicted variable)
+
surface data in that pixel (predictors)

Split train/test (daily)

Filter depths > 1000m

Filter latitudes < abs(60°)

Same filtering criteria for all models

Reshape & variable selection

More than 8 Random Forests & 6 LSTM configurations. For the validation, we chose 2 of each.

Training model

Different configurations, variables list etc.

Proposed models

RFR (Random Forest Regressor)

- ★ Simple & Fast
- ★ Can model non-linear relationships
- ★ Binary decisions, some artifacts can be induced due to the space division.

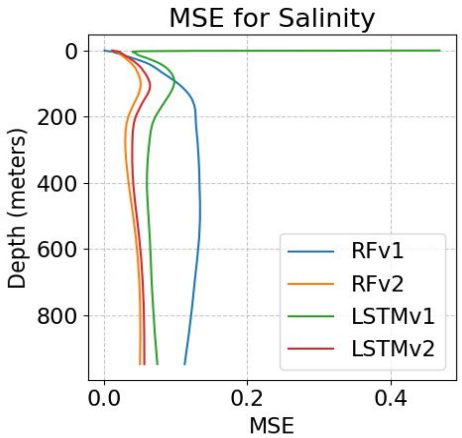
LSTM (long-short term memory)

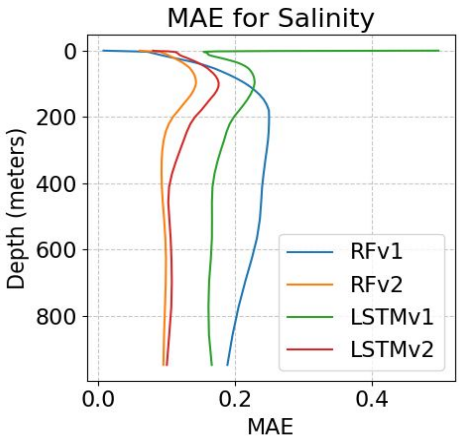
- ★ Efficient generating long-sequences
- ★ For this specific challenge, we can use the vertical profile as a sequence (instead of the timeseries)
- ★ Good results in the current literature
- ★ More complex than the Machine learning alternatives.

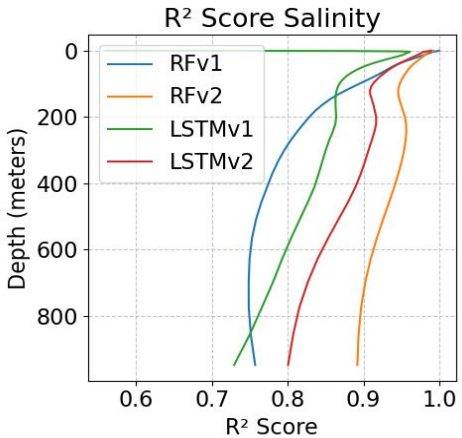
RFRv1	***Salinity and Temperature***
RFRv2	***Salinity, Temperature, Currents, MLD, SSH and latitude.***
LSTMv1	***Salinity, Temperature, Currents, MLD, SSH, day of the year, depth, longitude and latitude.***
	Same architecture as *Buongiorno Nardelli 2020*
LSTMv2	***Salinity, Temperature, Currents, MLD, SSH, day of the year, depth, longitude and latitude.***
	Optimized configuration for Salinity reconstruction.

Performance & Explainability

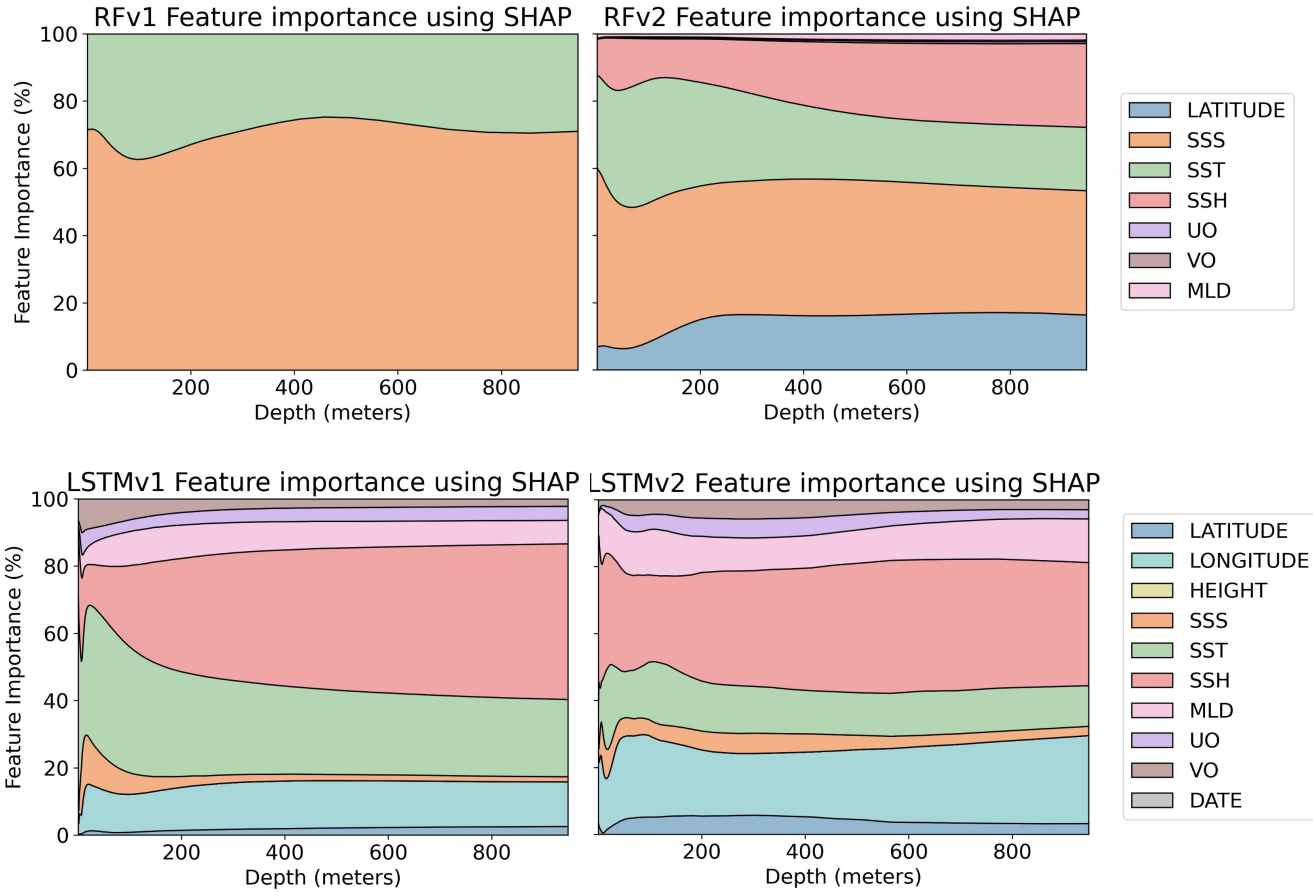
	R^2	MSE	MAE
RFRv1	0.88	0.08	0.17
RFRv2	0.95	**0.04**	**0.11**
LSTMv1	0.87	0.08	0.2
LSTMv2	**0.96**	**0.04**	0.13





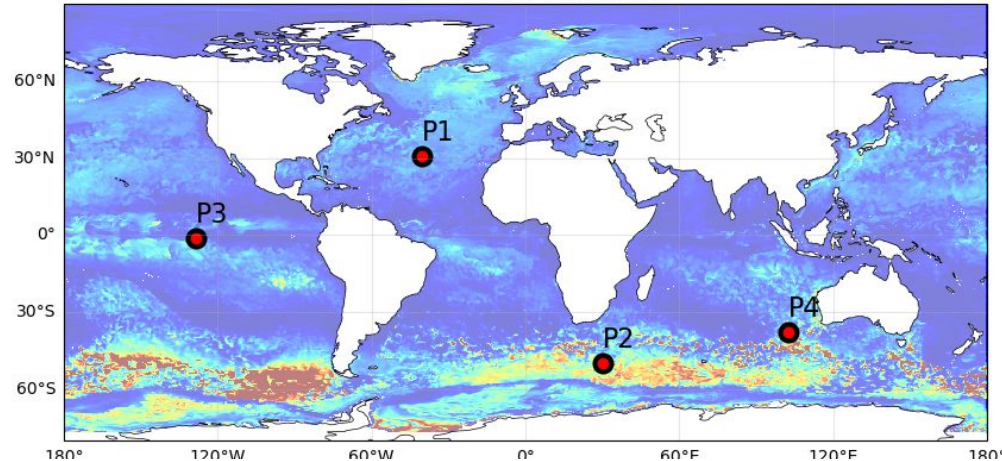


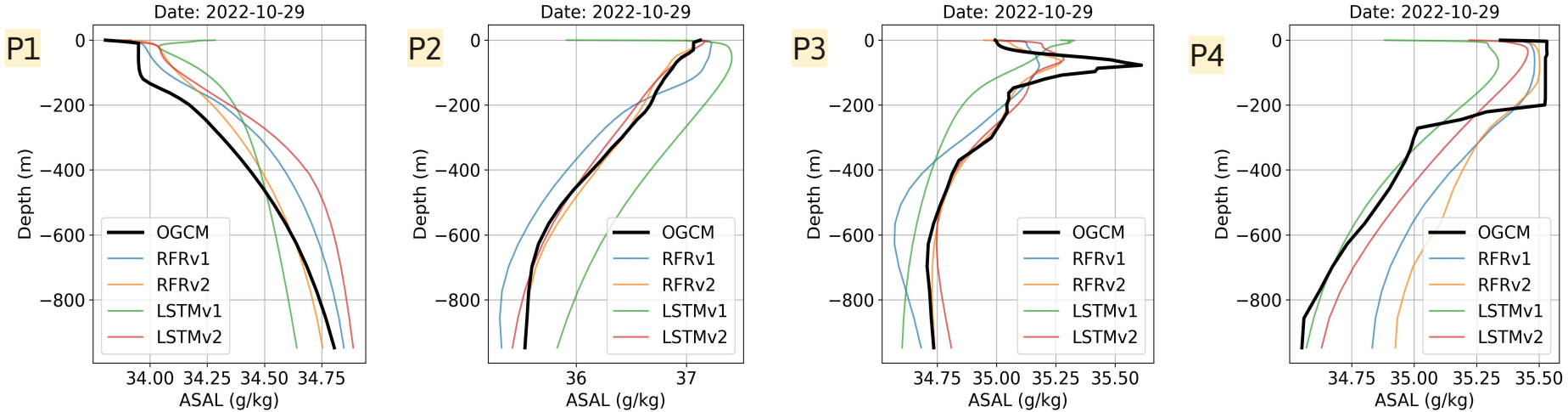
Performance & Explainability



Vertical profiles validation

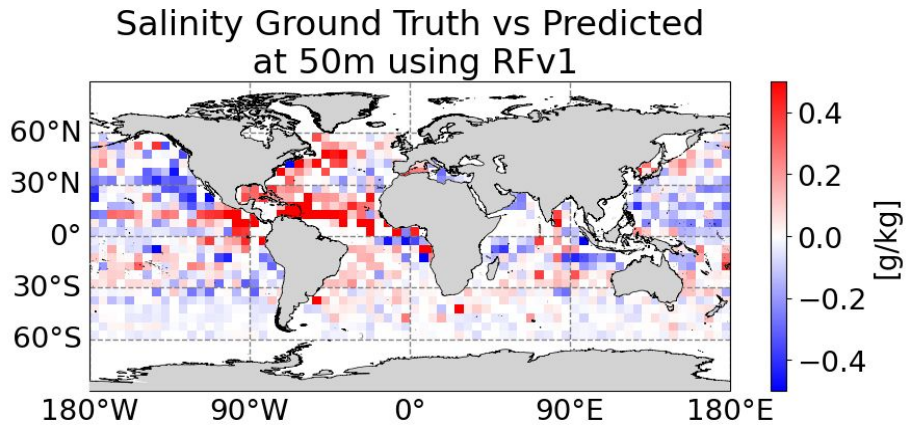
- ★ We chose 4 points with different dynamics
- ★ We compared the vertical profile as seen by the numerical model vs. the predicted ones.

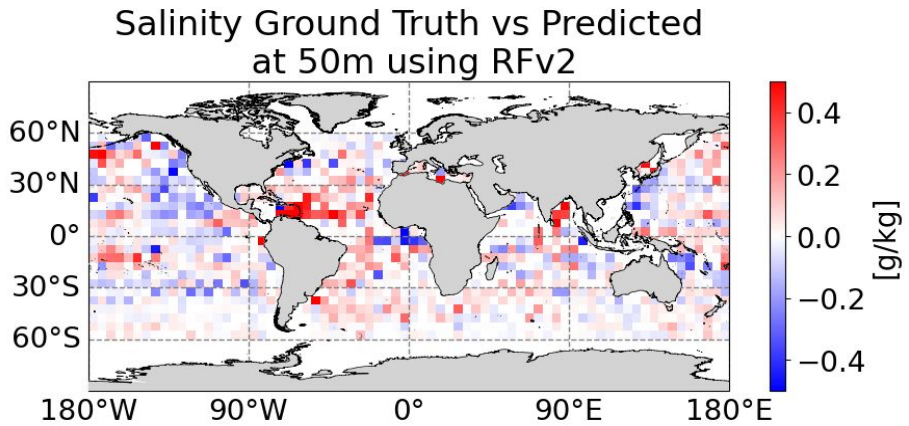


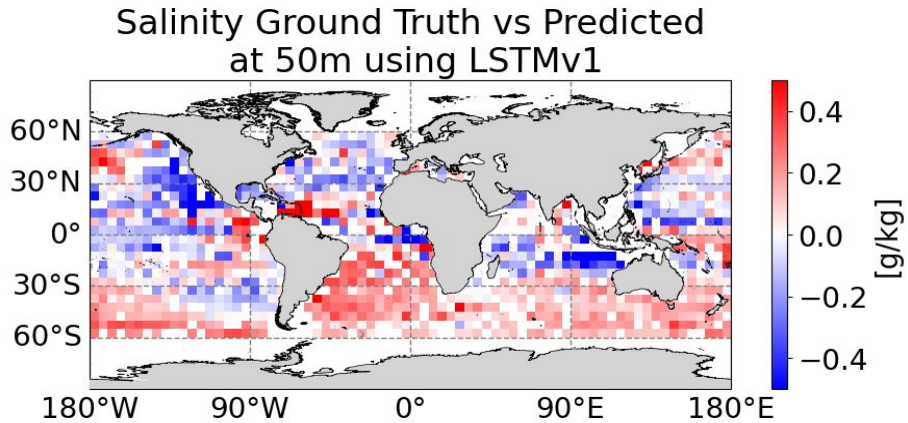


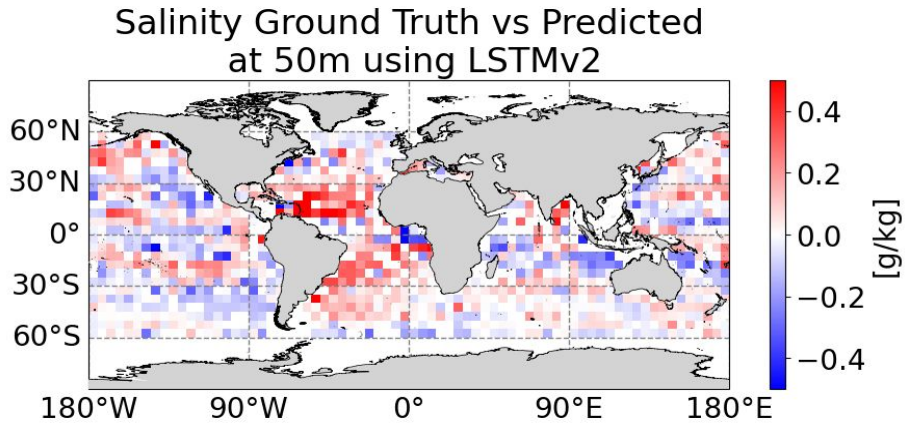
Spatial Analysis

Predicted points aggregated at 5° x 5° resolution, 20th october of 2022









Validation with the numerical model

RFRv2

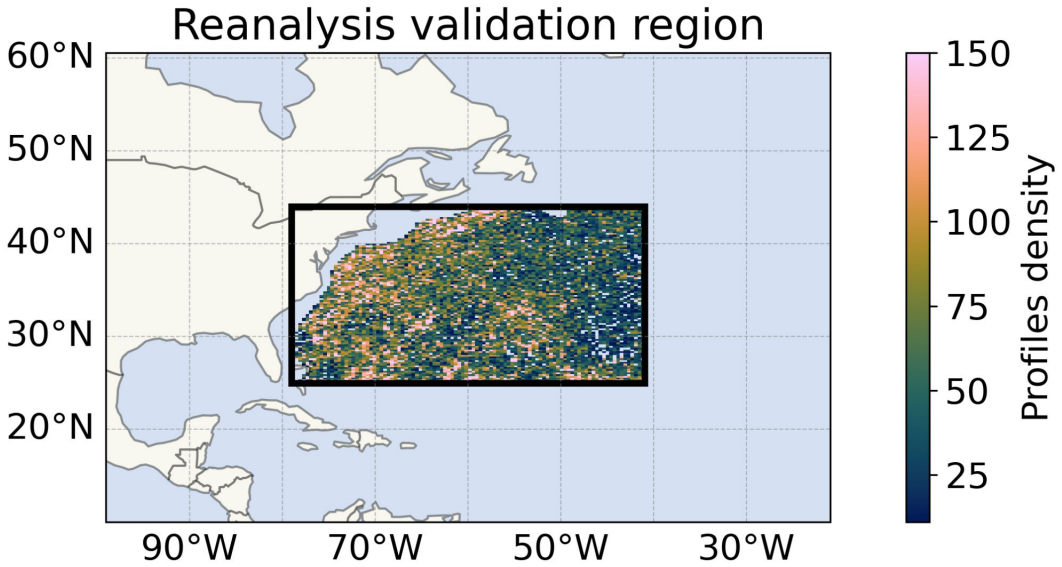
Salinity, Temperature, Currents, MLD, SSH and latitude.

LSTMv2

Salinity, Temperature, Currents, MLD, SSH, day of the year, depth, longitude and latitude.

Optimized configuration for Salinity reconstruction.

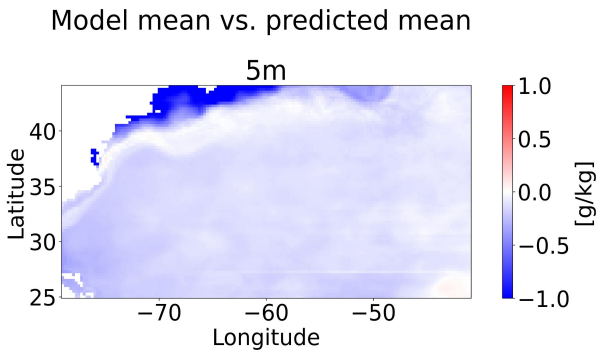
- ★ Bias of the 2-year mean map (2008/09)
- ★ Standard deviation map
- ★ MSE map
- ★ Temporal correlation map
- ★ Spatial correlation time series

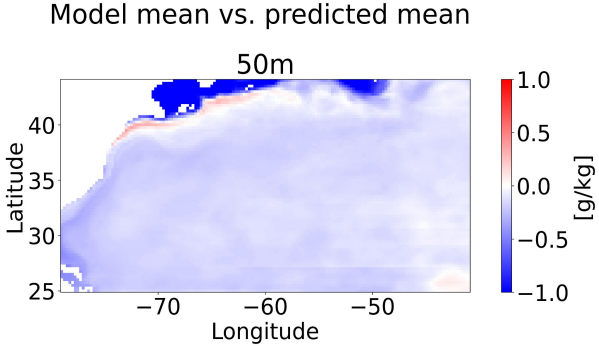


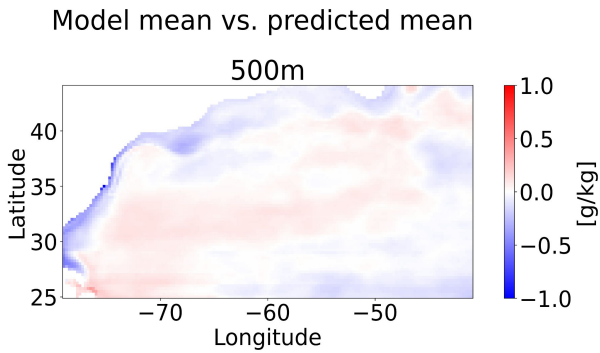
Validation I: Bias of the mean value

5m *50m* *500m*

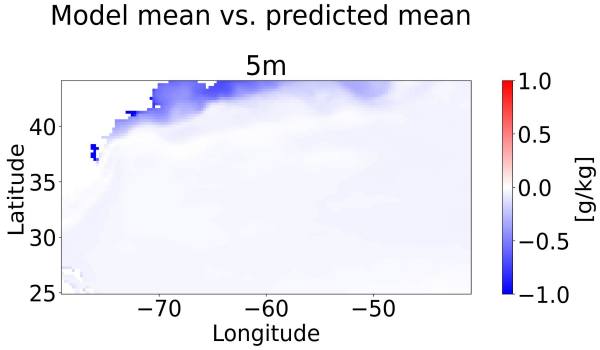
RFRv2

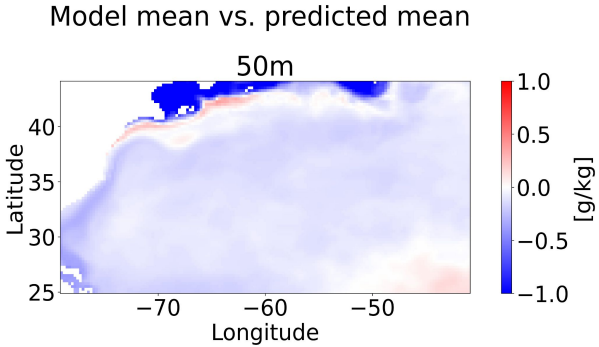


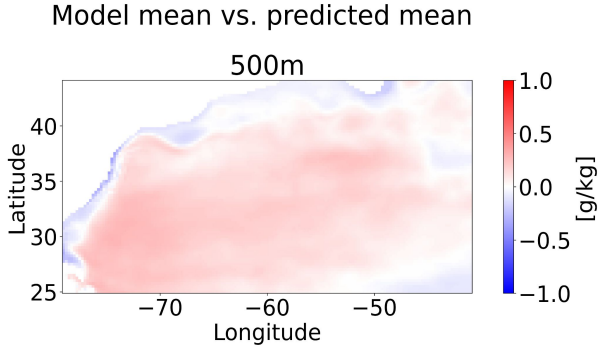




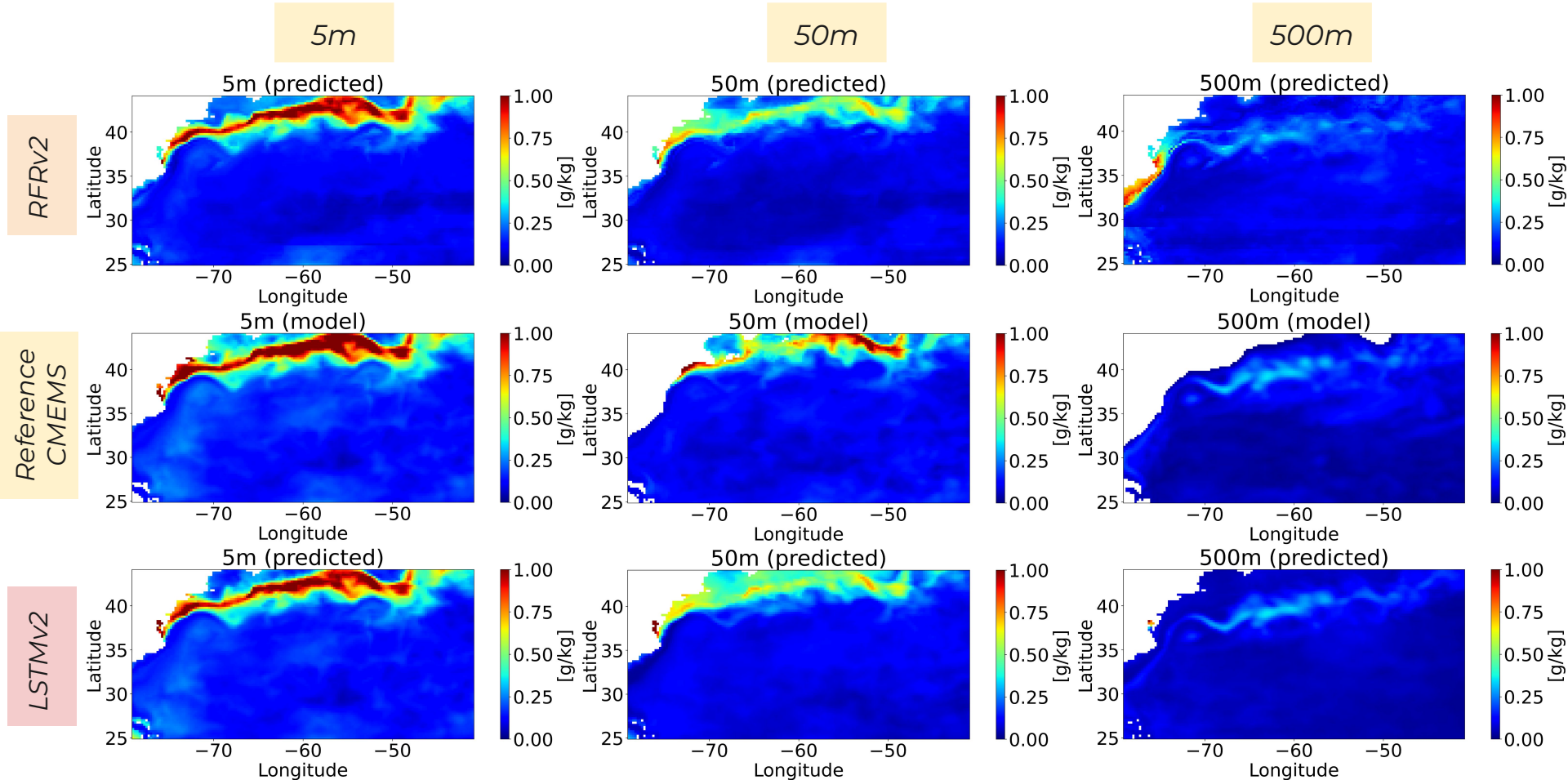
LSTMv2



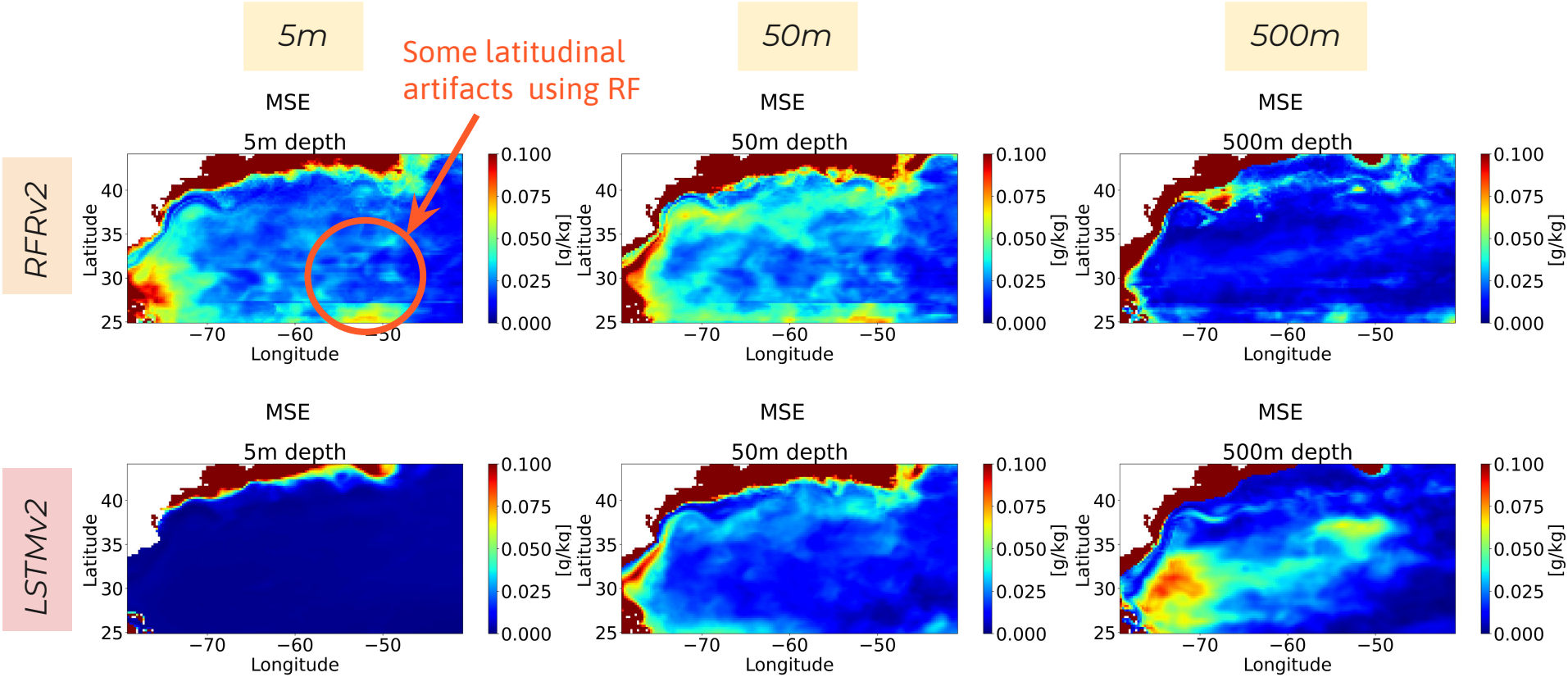




Validation II: Temporal Variability



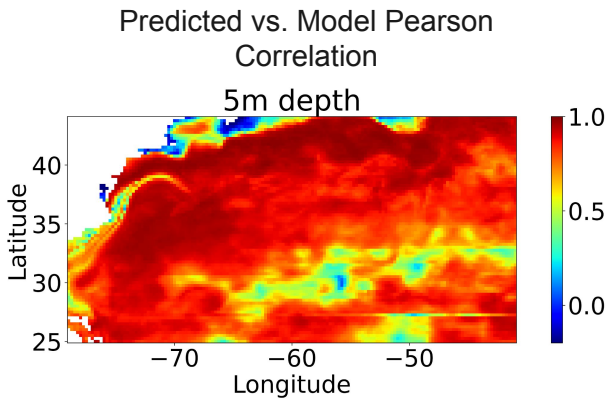
Validation III: MSE

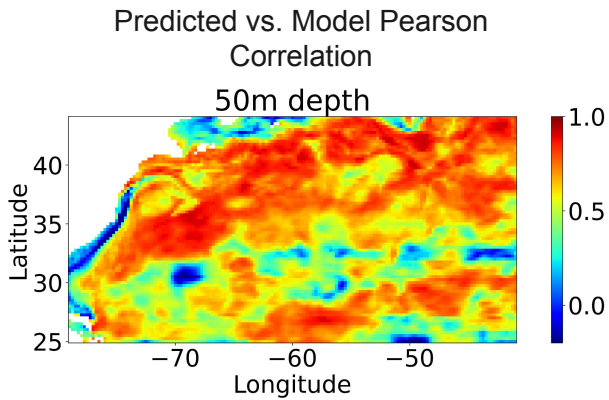


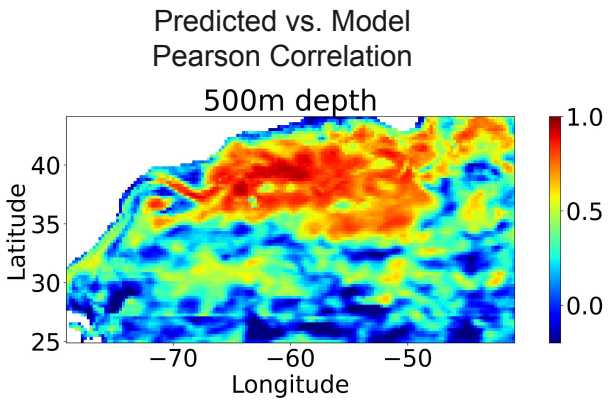
Validation IV: Pearson temporal correlation

5m *50m* *500m*

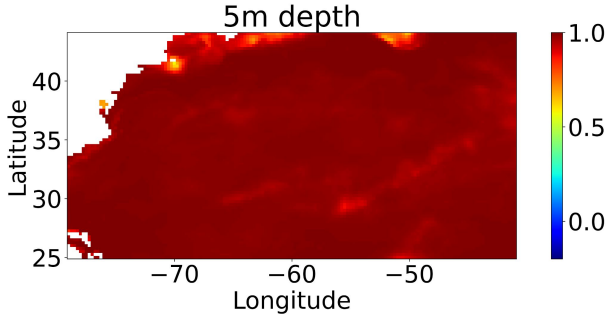
RFRv2

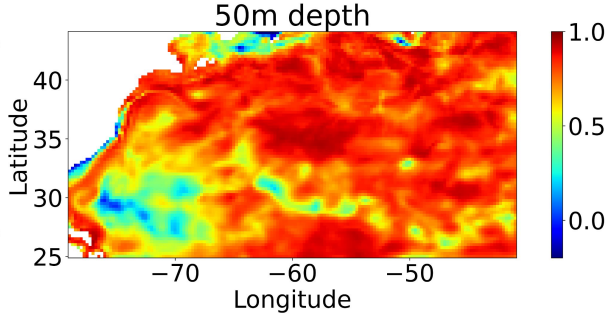


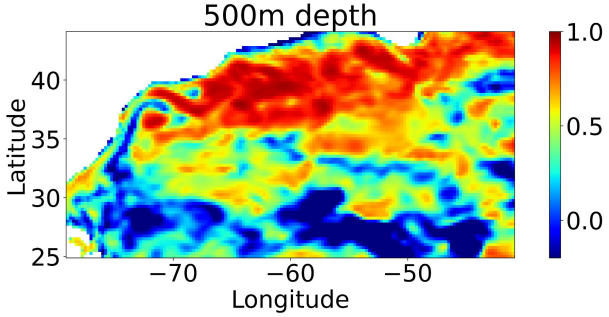




LSTMv2



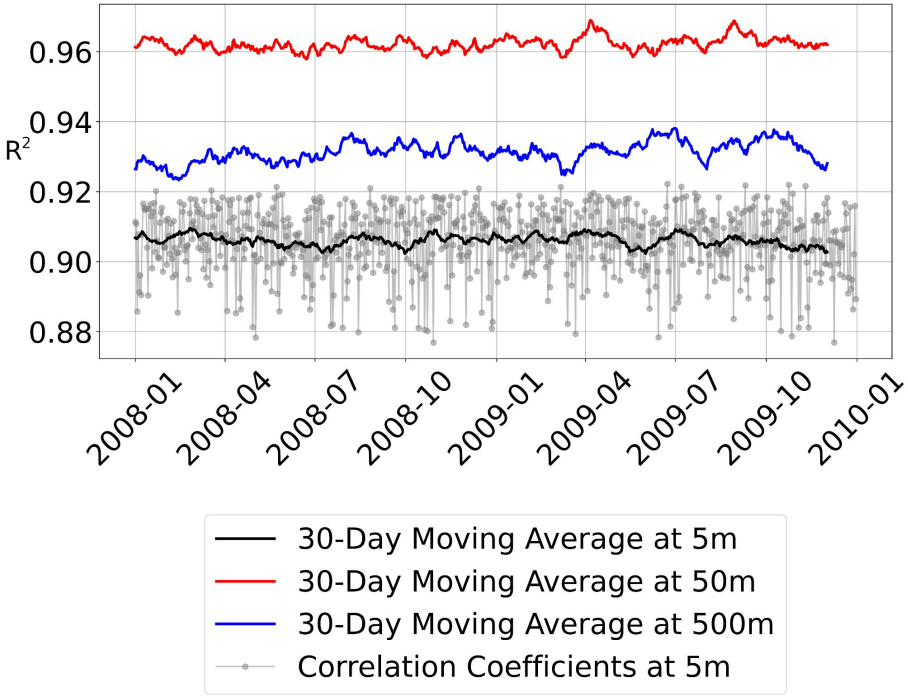




Validation V: Spatial Correlation

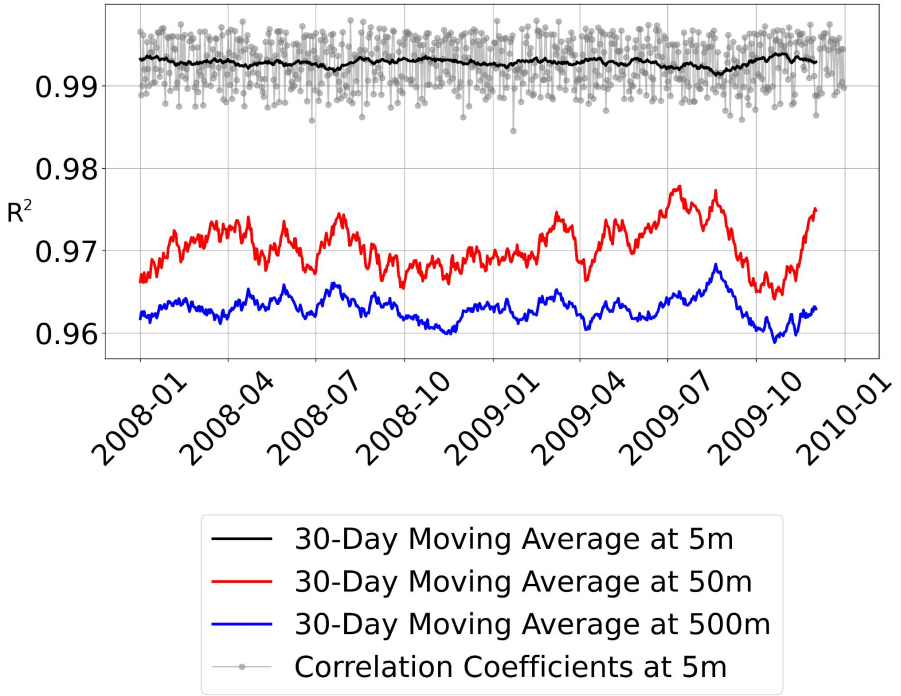
RFRv2

RF Salinity Correlation Coefficients



LSTMv2

LSTM Salinity Correlation Coefficients



Conclusions and future work

Conclusions

Best model	LSTMv2 with an R^2 score of 0.96. We want to remark that the RF offers similar results and is more simple and efficient. (artifacts to be removed)
Quality of the reconstruction	Salinity models capture the variability seen by the data, although we observe some systematic biases.
Spatial resolution of the reconstruction	Depends on the input surface resolution. The depths are fixed.

Future work

- ★ Study on the effective spatial resolution obtained by the model
- ★ Study of the systematic biases induced by the models.
- ★ Try different models based on images such as encoder and decoders, etc.
- ★ Test if the model can be used with real data using inference techniques. (Same weights but with real data)
- ★ Introduce activation functions which are coherent with the physical rules of the ocean.

Thank you for your attention

Contact: ainagarcia@icm.csic.es