Supplementary material, Neural network approaches for Sea Surface Height predictability using Sea Surface Temperature

L. Ollier, S. Thiria, M. Crépon, A. Charantonis, C. E. Mejia

July 10, 2024

1 Supplementary material

5 performances table are supplied to complete the article :

- Table 1 gives the results for different architecture on the less active zone
- Table 2 gives the results of different architecture on the dynamic zone
- Table 3 performs with the teacher forcing method for different inputs and outputs configurations (RMSE).
- Table 4 gives the results of different architecture on the dynamic zone with MAE.
- Table 5 performs with the teacher forcing method for different inputs and outputs configurations (MAE).

1.1 Hyper parameters

Most of the training parameters were determined through trial and error. We conducted a Bayesian optimisation [2] on various parameters including the starting learning rate, patience for the learning rate scheduler, reducing factor, batch size and network depth. The following settings were used for the optimization:

- Learning rate: We tested values ranging from 0.1 to 0.00001, using a learning rate scheduler that reduces the learning rate on a plateau.
- Batch size: We used a batch size of 32, which is typically large enough for networks with batch normalization. A power of 2 is a commonly recommended batch size, and 32 was a good trade-off between computational efficiency and size.
- Network depth: We explored the performance of the architecture with different depths, including 3, 4, and the original U-Net architecture with a depth of 5.
- Activation function: We used the SiLU activation function, which is similar to ReLU but has a global minimum and non-monotonic increase (see FIG 1). It can serve as a regularizer for excessively high weights.[1].
- Convolution parameters: We used a classic 3x3 kernel with a stride of 1 and without dilation. A one-pixel padding was applied on each side of the figure through each convolution. We found that zero padding led to too much error on the boundaries, so we switched to using replication padding (which pads using the replication of the boundary).

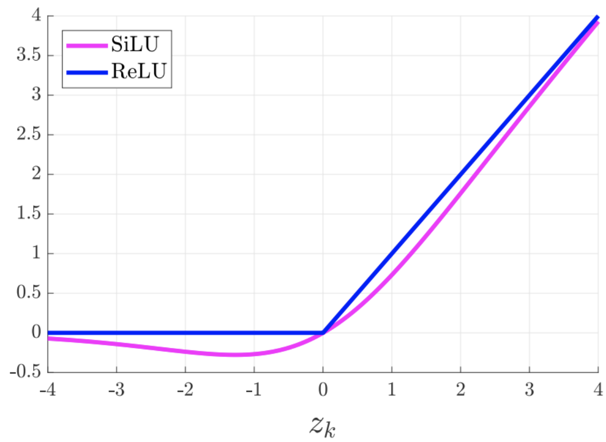

Figure 1: Sigmoid Linear Units activation function for neural networks compared to Rectified Linear Units. The activation of the SiLU is computed by the sigmoid function multiplied by its input.

Table 1: Comparison of architectures on the less active oceanic region (lat N from 26.5 to 33.7°)

Inputs/Predictions	Evaluate	n=5	n=10	n=15	n=20
Persistence	RMSE(cm)	5.4	8.5	10.54	11.97
Architecture : Unet					
SLA/SLA	RMSE(cm)	10.25	11.45	12.46	13.24
SLA-SST/SLA	RMSE(cm)	10.22	11.51	12.59	13.39
SLA-SST/SLA-SST	RMSE(cm)	12.08	13.29	14.3	15.03
Architecture : SLARESUnet					
SLA/SLA	RMSE(cm)	6.68	8.78	10.56	11.72
SLA-SST/SLA	RMSE(cm)	6.31	8.78	10.59	12.05
SLA-SST/SLA-SST	RMSE(cm)	6.24	8.44	9.78	10.66

Table 2: Comparison of architectures performances with RMSE on the dynamic oceanic region (lat N from 33.7 to 44.42°)

Inputs/Predictions	Evaluate	n=5	n=10	n=15	n=20
Persistence	RMSE(cm)	13.00	19.79	23.74	26.18
Architecture : Unet					
SLA/SLA	RMSE(cm)	13.59	18.19	21.74	23.94
SLA-SST/SLA	RMSE(cm)	15.41	19.54	22.25	23.81
SLA-SST/SLA-SST	RMSE(cm)	14.66	18.32	21.20	23.06
Architecture : SLARESUnet					
SLA/SLA	RMSE(cm)	13.27	17.86	21.73	23.66
SLA-SST/SLA	RMSE(cm)	13.13	18.53	22.07	24.59
SLA-SST/SLA-SST	RMSE(cm)	13.81	18.73	21.88	23.76

Table 3: RMSE on the dynamic oceanic region (lat N from 33.7 to 44.42°) with teacher forcing training method

Inputs/Predictions	Evaluate	n=5	n=10	n=15	n=20
DY SLA/SLA	RMSE(cm)	15.43	19.35	22.28	23.80
DY SLA-SST/SLA	RMSE(cm)	10.89	13.56	15.01	15.91
DY SLA-SST/SLA-SST	RMSE(cm)	12.23	16.79	19.13	20.50

Table 4: Comparison of architectures performances with MAE on the dynamic oceanic region (lat N from 33.7 to 44.42°)

Inputs/Predictions	Evaluate	n=5	n=10	n=15	n=20
Persistence	MAE(cm)	8.51	13.10	15.83	17.58
Architecture : Unet					
SLA/SLA	MAE(cm)	10.96	13.56	15.75	17.22
SLA-SST/SLA	MAE(cm)	11.49	14.14	15.96	17.04
SLA-SST/SLA-SST	MAE(cm)	11.50	13.83	15.90	17.43
Architecture : SLARESUnet					
SLA/SLA	MAE(cm)	9.22	12.38	15.37	16.89
SLA-SST/SLA	MAE(cm)	9.43	12.92	15.12	16.59
SLA-SST/SLA-SST	MAE(cm)	9.33	12.81	15.07	16.62

Table 5: MAE on the dynamic oceanic region (lat N from 33.7 to 44.42°) with teacher forcing training method

Inputs/Predictions	Evaluate	n=5	n=10	n=15	n=20
DY SLA/SLA	RMSE(cm)	10.40	13.08	15.06	16.15
DY SLA-SST/SLA	RMSE(cm)	7.90	9.70	10.70	11.25
DY SLA-SST/SLA-SST	RMSE(cm)	8.55	11.64	13.51	14.55

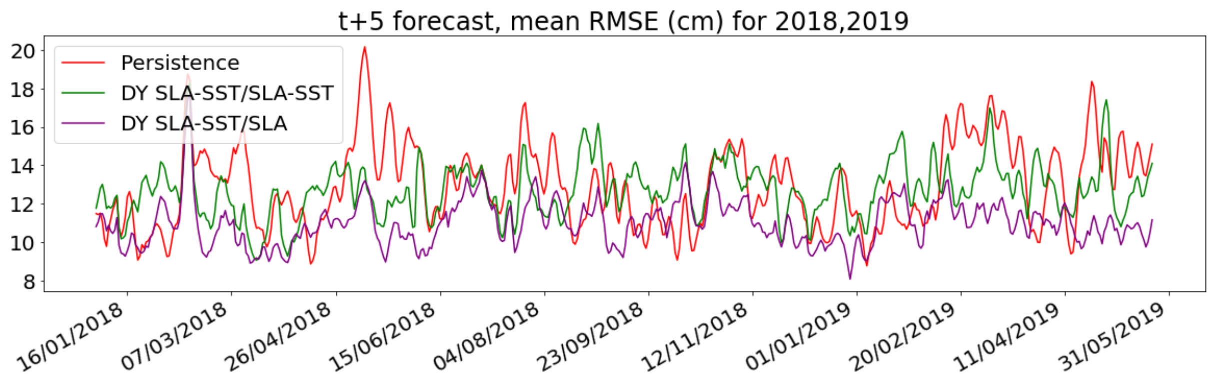

Figure 2: Temporal RMSE on the dynamic oceanic region (lat N from 33.7 to 44.42°) for persistence, DY SLA-SST/SLA and DY SLA-SST/SLA-SST.

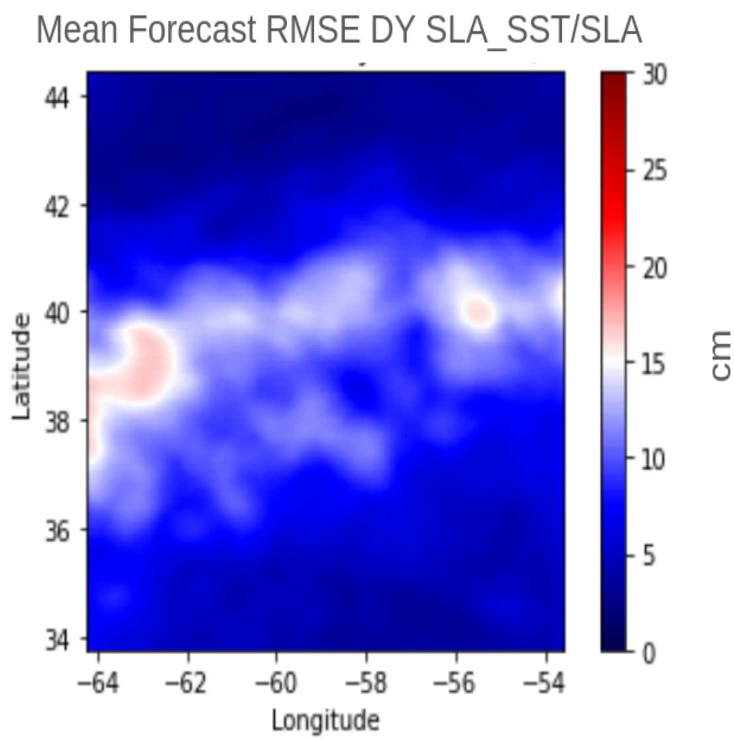

Figure 3: RMSE over the area for the DY SLA-SST/SLA architecture for a 5 day forecast. The white threshold of the colormap has been set to the mean standard deviation of the area

References

[1] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. *Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning*. 2017. arXiv: `1702.03118 [cs.LG]`.

[2] Peter I. Frazier. *A Tutorial on Bayesian Optimization*. 2018. arXiv: `1807.02811 [stat.ML]`.