Response To Referee 2
for 'Data-driven surrogate modeling of high-resolution sea-ice thickness in the Arctic'

Charlotte Durand, Tobias Sebastian Finn, Alban Farchi,
Marc Bocquet, and Einar Òlason

December 2023

RC: Reviewer Comment; AR: Author Response

RC: The paper presents a strong case of surrogate modeling by using neural networks to emulate the increase in sea ice thickness, however, the paper lacks clarity at several places in the manuscript and requires minor revisions:

AR: We deeply appreciate the reviewer's thorough and insightful review of our work. In the following, we will comment on the raised issues and the changes planned for our revised manuscript.

RC: 1. There is little information provided on the choice of atmospheric variables considered as forcings. Please provide more evidence from literature on this.

AR: We will add more references and comments in the manuscript to explain our choice of forcings. Firstly, let's note that for neXtSIM simulations, the atmospheric forcings consist of the 10 m wind velocities, the 2 m temperature, mixing ratio, mean sea level pressure, total precipitation and the snow fraction. We decided to limit ourselves to the first three. Plueddemann et al. (1998) and Kwok et al. (2013), for example, have shown that the sea-ice drift is strongly linked to wind velocities. There is a strong correlation between the atmosphere winds and the sea-ice motion, up to 0.8 in Central Arctic (Thorndike and Colony, 1982; Serreze et al., 1989; Zhang et al., 2000). Those forcings are a good proxy for the advection of the sea-ice, which is also necessary for emulating sea-ice thickness dynamics. Observational studies have shown that *interannual variability in sea ice conditions is caused by the variability in the large-scale atmospheric circulation which locally manifests itself as surface air temperature and wind* *anomalies*, (Deser et al., 2000; Prinsenberg et al., 1997). Experiments were originally conducted with additional forcings, including sea-surface temperature (SST); however, SST was later excluded because the simulation was coupled with the ocean in this version of neXtSIM.

RC: 2. If the neural network is designed for future forecasting, none of the input features should belong to the same timestep as the target. In case of this paper, all the atmospheric variables are of same timestep whereas like SIT, they should also be up till 't' timestep. You can justify through experiments how the current setting performs better than the one suggested.

AR: We chose to incorporate 'future forcings' based on the understanding that, in sea-ice modeling, the advection of sea ice is strongly influenced by the forecast atmospheric forcings. Experiments were also performed without those future forcings, up until t. The impact on forecast skills was non-negligible, as displayed in Fig. 1. Note that the results presented here are evaluated on the validation dataset. In the simulations on which our dataset is based on (Boutin et al., 2023), neXtSIM is uncoupled from an atmospheric model and uses just ERA5 forcings. In such settings, the atmospheric forcing can be given by forecasts, and, thus, known for the future. Consequently, using future forcings during training is nonrestrictive in terms of its potential operational capability.

RC: 3. There are some minor errors that should be corrected: UNet by definition is not a convolutional architecture but it is an encoder-decoder Neural Network architecture with skip-connections. There are several papers utilizing LSTM-based UNet or ConvLSTM-based UNet. Andersson et al. did not propose IceNet for SIC prediction. Their work targets SIP predictions which is slightly different from SIC.

AR: Thank you for your remarks. We will add references for more LSTM-based UNet or ConvLSTM-based UNet, correct the mention of the IceNet paper to SIP predictions, and the definition of UNet. We will define the UNet as an encoder-decoder-based CNN.

RC: 4. There are several other recent papers that utilize CNN, ConvLSTM and LSTM for SIC predictions. There is not enough convincing argument present on just relying on UNet for the surrogate model. Did the authors try a CNN or ConvLSTM based architecture for surrogate modeling?

AR: Several architectures have been tested on a coarse-grained dataset ($128 \times$

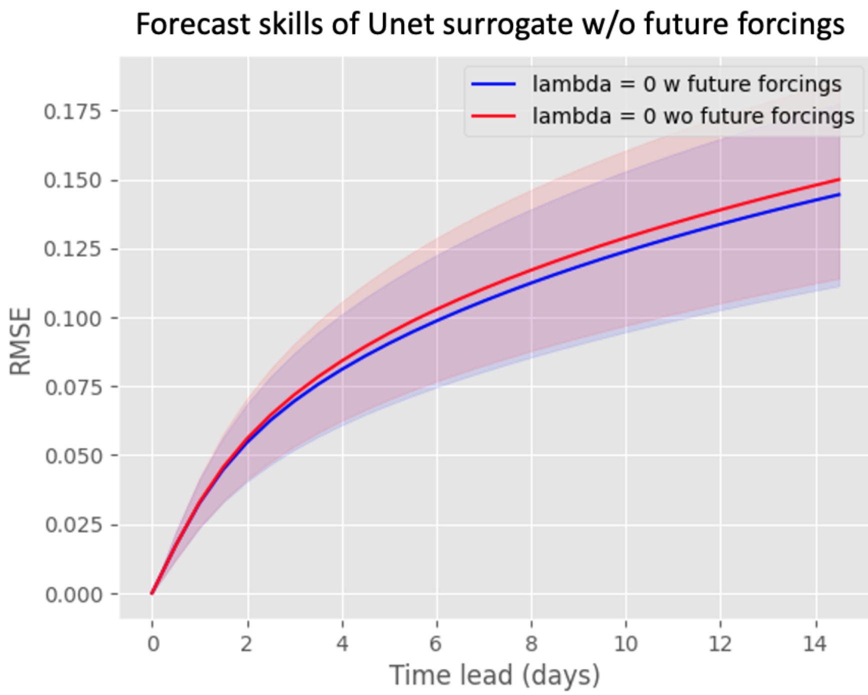

Figure 1: Forecast skills of surrogate models depending on the addition of future forcings. Two surrogate models are evaluated on the validation dataset: one with atmospheric forcings, up to t (red curve) and another with the addition of future atmospheric forcings, up to $t + 12$ hours (blue curve). Averages of the RMSE are shown with solid curves, and their associated standard deviations are outlined with transparency.

128 grid cells) to reduce computation costs. Both mentioned ResNet and ConvLSTM structures have been investigated, yielding quite similar results in terms of forecast skills on the validation dataset. From our experimentation, the specific structure of the convolutional neural network does not matter much. Our LSTM-based approach, with a lag of 48 hours, led to satisfactory forecast skills; caused by the high computational costs (441 s/epoch) compared to the UNet (108 s/epoch), we focus on the UNet structure when we moved to the high-resolution dataset. Regarding the ResNet architecture, also implemented with partial convolution, the results were also quite similar, despite higher computational costs (172 s/epoch). Forecasts skills results are presented in Fig. 2 for UNet, ResNet and ConvLSTM. As no extensive study for those different architectures have been conducted afterwards, and the results presented here are on a coarser resolution, not necessarily fully hyper-optimized, those experiments will not be presented in the manuscript.

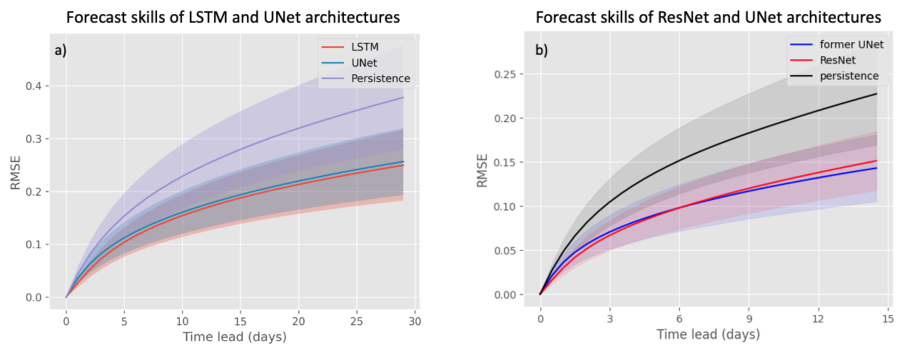

Figure 2: Forecast skills for different neural network architecture on a coarse-grained dataset (128×128 grid-cells) on the validation dataset. (a) Comparison between a ConvLSTM architecture (red) and a UNet architecture (blue). (b) Comparison between a ResNet architecture (red) and a UNet architecture (blue). The solid curves represent the averaged normalized root-mean-squared error (nRMSE) and in transparency is represented its associated standard deviation. The persistence baseline is also indicated in purple (a) and black (b).

RC: 5. How was 100 decided as the optimal value of lambda? Did you experiment with other values of lambda in calculating the global loss?

AR: This value was chosen after experimenting with several values. The impact of λ on the forecast skill is negligible, while having an important impact on the bias. The value of $\lambda = 100$ was chosen based on the evaluation on the validation dataset. Others values of λ were not evaluated on the test dataset. Please find the experiments on the validation dataset in Fig. 3. Selecting a value for λ that is excessively large could result in a loss of information at the fine scale. A value of $\lambda = 100$ seems to keep a good balance between fine-scale dynamics and global sea-ice thickness. We will add more explanation for the tuning of λ in our manuscript.

RC: 6. What is the timestep used in case of longterm forecasting?

AR: For the long-term forecast, the NN trained with one timestep is used. We will add this remark in the manuscript.

RC: 7. Did the authors consider using custom loss function instead of partial convolution to incorporate land-mask into the modeling?

AR: The loss function is already custom and takes the mask into account.

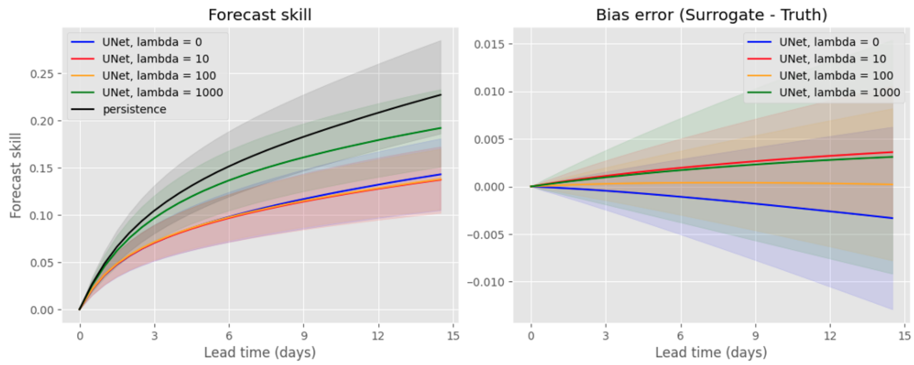

Figure 3: Impact of the choice of λ on the forecast skill and bias error on the validation dataset. On the left panel is presented the forecast skill, as defined on the manuscript. The solid curves represent the averaged normalized root-mean-squared error (nRMSE) and in transparency is represented its associated standard deviation. On the right panel is shown the bias error, as defined on the manuscript. The solid curves represent the averaged bias error and in transparency is represented its associated standard deviation. Results for several values of λ are shown for both panel : $\lambda = 0$ (blue), $\lambda = 10$ (red), $\lambda = 100$ (yellow), $\lambda = 1000$ (green). The persistence baseline forecast skill is also indicated in black.

Should we use normal convolutions instead of partial convolutions, we would zero-pad land pixels. The effect of the land masses would be then similar to effects of zero padding at image boundaries, which can lead to artifacts (Liu et al., 2018). Consequently, by taking the mask only for the loss function into account, we would possibly generate artifacts in regions with a lot of land masses. Additionally, without masking operations, during cycling of the neural network for longer lead times than 12 h, errors could rapidly accumulate on land and lead to physically inconsistent results.

RC: Ref: 1. Ebert-Uphoff, Imme, et al. "CIRA Guide to Custom Loss Functions for Neural Networks in Environmental Sciences–Version 1." arXiv preprint arXiv:2106.09757 (2021).
2. Ali, Sahara, and Jianwu Wang. "MT-IceNet-A Spatial and Multi-Temporal Deep Learning Model for Arctic Sea Ice Forecasting." 2022 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT). IEEE, 2022.
3. Kim, Eliot, et al. "Multi-task deep learning based spatiotemporal arctic sea ice forecasting." 2021 IEEE International Conference on Big Data (Big Data). IEEE, 2021.

AR: Thank you for the references.

References

Boutin, G., Ólason, E., Rampal, P., Regan, H., Lique, C., Talandier, C., Brodeau, L., and Ricker, R. (2023). Arctic sea ice mass balance in a new coupled ice–ocean model using a brittle rheology framework. *The Cryosphere*, 17(2):617–638.

Deser, C., Walsh, J. E., and Timlin, M. S. (2000). Arctic sea ice variability in the context of recent atmospheric circulation trends. *Journal of Climate*, 13(3):617–633.

Kwok, R., Spreen, G., and Pang, S. (2013). Arctic sea ice circulation and drift speed: Decadal trends and ocean currents. *Journal of Geophysical Research: Oceans*, 118(5):2408–2425.

Liu, G., Shih, K. J., Wang, T.-C., Reda, F. A., Sapra, K., Yu, Z., Tao, A., and Catanzaro, B. (2018). Partial convolution based padding. *arXiv preprint arXiv:1811.11718*.

Plueddemann, A. J., Krishfield, R., Takizawa, T., Hatakeyama, K., and Honjo, S. (1998). Upper ocean velocities in the beaufort gyre. *Geophysical Research Letters*, 25(2):183–186.

Prinsenberg, S. J., Peterson, I. K., Narayanan, S., and Umoh, J. U. (1997). Interaction between atmosphere, ice cover, and ocean off labrador and newfoundland from 1962 to 1992. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(S1):30–39.

Serreze, M. C., Barry, R. G., and McLaren, A. S. (1989). Seasonal variations in sea ice motion and effects on sea ice concentration in the canada basin. *Journal of Geophysical Research: Oceans*, 94(C8):10955–10970.

Thorndike, A. S. and Colony, R. (1982). Sea ice motion in response to geostrophic winds. *Journal of Geophysical Research: Oceans*, 87(C8):5845–5852.

Zhang, J., Rothrock, D., and Steele, M. (2000). Recent changes in arctic sea ice: The interplay between ice dynamics and thermodynamics. *Journal of Climate*, 13(17):3099–3114.