Response To Referee 1 for 'Four-dimensional variational data assimilation with a sea-ice thickness emulator'

Charlotte Durand, Tobias Sebastian Finn, Alban Farchi, Marc Bocquet, Julien Brajard and Laurent Bertino

April 2025

RC: Reviewer Comment; AR: Author Response

RC: The manuscript "Four-dimensional variational data assimilation with a sea-ice thickness emulator" by Durand et al presents an evaluation of a 4D-Var data assimilation framework using a data-driven sea ice thickness emulator of the neXtSIM sea ice model. The authors show how, through the emulator's back-propagation capabilities, sea ice thickness observations (both idealized and real) can be assimilated into the emulator using 4D-Var, effectively reducing the emulator bias. I very much enjoyed reading this manuscript and think it's a nice contribution to the literature. In fact, most of the comments I had noted down by the time of the discussion were then answered in the discussion, so thanks! My comments were overall minor, and I think the manuscript is almost ready for publication with a few small edits (see below).

AR: We deeply appreciate the reviewer's thorough and insightful review of our work. In the following, we respond to the comments and raised issues and point to the changes in our manuscript.

RC: General question regarding methodology

RC: Could you just clarify something about the methodology for me. Are the EOFs used for the background covariance static over the course of the DA simulation? My concern early on was the ability of the EOF approach to capture flow-dependent processes, given the strong seasonal cycle of sea ice (you do mention this

in the discussion). Is there some expectation that the minimization figures out which EOFs are most important and dynamically weights them (in time) according to w? I would be very interested to see how the approach compares to an Ensemble Kalman Filter (as you also say in the discussion).

AR: Thank you for your remark. Indeed we investigated the influence of the different EOFs weights and their seasonality although we did not present the results in this paper. They are presented in Durand (2024) in Chapter 7. Regarding the 4D–Var-EOF, we can assess the significance of the different weights associated with the EOFs and evaluate the time dependency of the predominant ones. The results are presented in Fig. 1 of the present document. Firstly, the weights associated with the largest amplitudes correspond to the first EOF coefficients. Secondly, tracking the time evolution of the first coefficient reveals a clear temporal dependency, in line with the annual evolution of the SIT. The second coefficient also exhibits a seasonal behavior, with an increase in amplitude around May. The 4D–Var-EOF approach captures the seasonal variability of the signal, which is expected to be the dominant source of temporal variability. However, the full flow-dependent covariance structure, dependent not only on the season but also on the specific realization of the forecast, cannot be represented by 4D–Var-EOF. Capturing this would require comparison with an ensemble-based method such as the EnKF, which would be interesting but is beyond the scope of this article. Note that a standard EnKF system does not rely on an EOF decomposition. However, if the ensemble were projected onto the same EOF basis, it would be possible to compute time-evolving weights w from the EnKF that vary with both location and time, making a direct comparison impractical.

RC: Comments

RC: L112: I suggest adding a citation to show an example of where observations are typically log-normal. E.g Landy et al 2020.

AR: Thank you very much for your suggestion, we will add this reference: "as more commonly encountered in sea-ice observations from satellites (Landy et al., 2020)"

RC: L117 and elsewhere: change "In average," to "On average,"

AR: Thank you for seeing this, we will correct it and check thoroughly the rest of the paper.

RC: L129 - L132: Somewhere in this section it might be worth high-

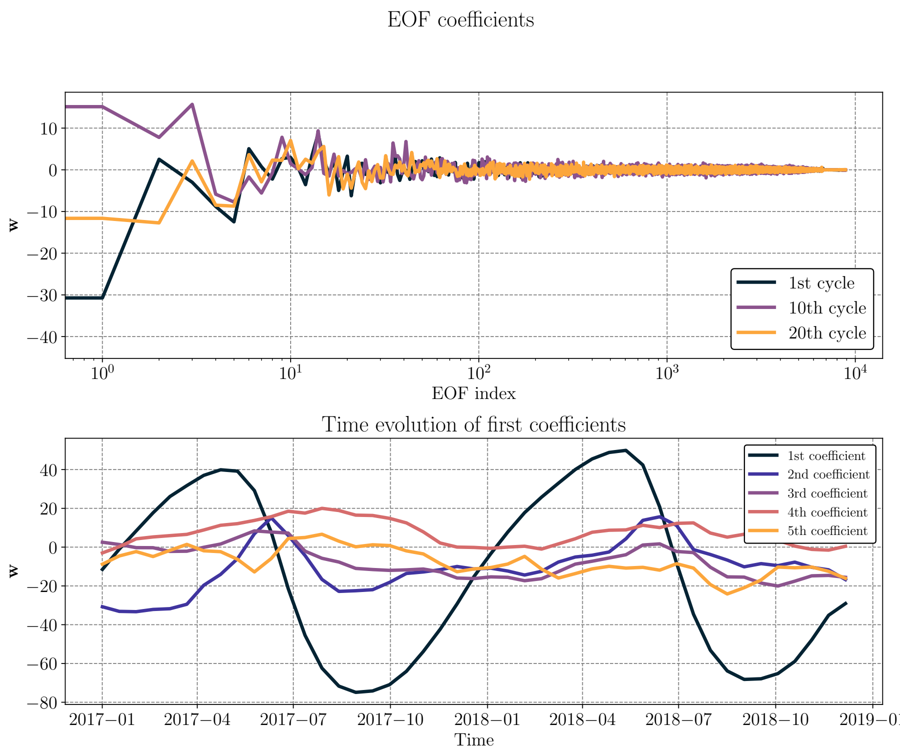

Figure 1: Upper panel: Values of the EOF weights for different assimilation cycles: the 1st cycle is represented by the blue line, the 10th cycle by the purple line, and the 20th cycle by orange line. Note that the x-axis is on a logarithmic scale to highlight the first weights. Bottom panel: Time evolution of the weights of $\mathbf{x}_\mathrm{a}$ associated with the first five EOFs for each cycle. The first coefficient is shown in dark blue, the second in indigo, the third in purple, the fourth in pink, and the fifth in orange.

lighting a recent paper (Nab et al. 2025) which quantified the effect on DA-derived analysis fields due to varying observational uncertainty on sea ice thickness measurements—Turns out to be quite sensitive.

AR: Thank you for this paper which we were not aware of. Indeed, this is worth mentioning. We will change the text in L132 to : "Note that Nab et al. (2025) showed that modifying SIT observation uncertainties introduces significant sensitivities during SIT assimilation."

RC: Figures 3 and 7: Missing text in all labels

AR: Thank you for noticing this, we will be careful to check if in the PDF the text is lisible.

RC: L258: Doesn't the RMSE in Fig 5 peak in July? I guess the bias error peaks just before May and then rises again in December? Maybe changing L258 to "from Fig. 5 top" to make it clear which panel in Fig 5 we are looking at

AR: Yes thank you for noticing that, the RMSE peak is indeed rather in July, we will change the sentence to "The analysis from Fig.5(top) reveals a strong seasonality in results, with the RMSE peaking in July."

RC: L306-310 : Can you borrow some info from data-driven NWP models which retain sharpness by augmenting loss function

AR: Yes, indeed! For the emulator, we are exploring alternative loss functions to better preserve sharpness. However, these often result in increased RMSEs, so this remains a work in progress and is beyond the scope of the current paper.

RC: L350 : Might be worth highlighting here that there are ongoing developments in this space. For example Chen et al and Gregory et al both show ML-based approaches for deriving complete daily sea ice and ocean fields from satellite altimetry at 5 km grid resolution. Both of these approaches model the spatio-temporal covariance of daily fields, rather than simply averaging through time. Although these studies show sea ice freeboard, it is conceivable that daily sea ice thickness observations are on the horizon.

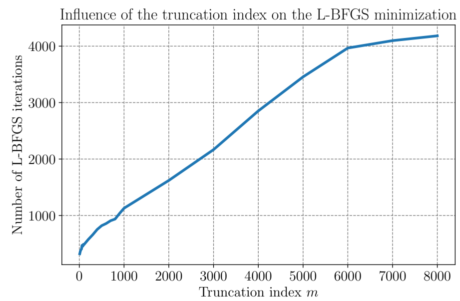

Figure 2: Evolution of the number of L-BFGS iterations as a function of the truncation index.

AR: Thank you for suggesting those references, we will add the sentence: "Ongoing works are proposing ML-based approaches to derive complete daily sea ice freeboard fields from satellite altimetry at fine spatial resolution (5 km), by modeling the spatio-temporal covariance of daily fields rather than relying solely on temporal averaging (Gregory et al., 2024; Chen et al., 2024)."

RC: L400: I thought neXtSIM-F was initialized through nudging and not EnKF (L274/275)?

AR: Thank you for seeing this error, neXtSIM-F is indeed initialized with nudging operationaly, we will correct the sentence: "With limited resources, such as emulating only sea-ice thickness and assimilating CS2SMOS observations, the developed 4D–Var system performs comparably to the operational neXtSIM-F system"

RC: Appendix B: Can you quantify the time change in the 4D-var minimization when increasing the truncation index m? For example, on L324 you say it's 155 seconds for m=7000. What is the time if m is halved to 3500? I guess I'm wondering what is the cost-accuracy tradeoff.

AR: We show in Fig. 2 of the present document the number of L-BFGS iterations as a function of the truncation index. The execution time to run the 4D–Var is proportional to the number of iterations. As you can see, the increase of the number of iterations is quite linear with the truncation index. When using this method in operations, the choice of the truncation index would be an important factor for the computational efficiency. Yet, to obtain the best RMSE values, based on our experiments, choosing a value above $m = 5000$ seems preferable.

References

Chen, W., Mahmood, A., Tsamados, M., and Takao, S. (2024). Deep random features for scalable interpolation of spatiotemporal data.

Durand, C. (2024). Deep learning, data assimilation and sea-ice dynamics.

Gregory, W., MacEachern, R., Takao, S., Lawrence, I. R., Nab, C., Deisenroth, M. P., and Tsamados, M. (2024). Scalable interpolation of satellite altimetry data with probabilistic machine learning. *Nature Communications*, 15(1).

Landy, J. C., Petty, A. A., Tsamados, M., and Stroeve, J. C. (2020). Sea ice roughness overlooked as a key source of uncertainty in cryosat-2 ice freeboard retrievals. *Journal of Geophysical Research: Oceans*, 125(5).

Nab, C., Mignac, D., Landy, J., Martin, M., Stroeve, J., and Tsamados, M. (2025). Sensitivity to sea ice thickness parameters in a coupled ice-ocean data assimilation system. *Journal of Advances in Modeling Earth Systems*, 17(3).