

Author's response to Referee 2  
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

August 2025

**RC:** Reviewer Comment; **AR:** Author Response

**RC:** Thank you for taking the time and effort to respond to the long list of comments in my previous review. I appreciate the improvement made to the manuscript and consider the science presented in the paper worthy of publication in *The Cryosphere*. However, I still think that the manuscript would benefit from a more thorough revision to improve its readability and accessibility. The majority of my Major Comments of the earlier review (with the notable exception of Major Comment 2) still stand. In particular, there isn't a clear "narrative" that brings the results together. Some of them could be pruned, whereas others could be elaborated or investigated into more thoroughly. In some cases, the authors' modification to the article's text has made the key ideas more obscure, which could be confusing to the reader (see e.g. Minor Comments 14 and 20 of this present review). The quality of the science needs to be matched with good presentation. Some minor comments are provided below, but I suggest that the authors focus on addressing the more important issues of readability and presentation first.

**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:** 1. Section 1, third and fourth paragraphs: I guess what you want to say in these two paragraphs is that the current work closes the gap in scientific knowledge: having a neXtSIM system (with its specific rheological model) with variational data assimilation,

a combination that hasn't been explored before. You need a "topical sentence" to give the writing some direction. As things stand, these two paragraphs consist of some not-so-related sentences put together, which aren't very useful to the reader.

AR: Thank you for your remark, we merged the two paragraphs and we added a sentence at the end to summarize the novelty of the work: "In this paper, we present the application of variational data assimilation methods in neXtSIM, enabled by a data-driven emulator."

RC: **2. Line 46: The wording "relies on" in this context is a bit odd. I suggest replacing it with "is achieved by running".**

AR: Thank you for your remark, we modified the sentence: "The propagation of the gradient information from the observational time backwards in time within the DAW is achieved by running on the model's adjoint during the cost function minimization."

RC: **3. Lines 47 – 48: A better wording for the sentence would be "Hence, the analysis increment at initialization time incorporates all observational information up to the end of the DAW."**

AR: Thank you for your proposition, we changed the sentence to this one.

RC: **4. Lines 98 – 99: You consider the variables you named, but for what purposes?**

AR: Thank you for your remark, we added two sentences to justify our choices: "Several atmospheric forcings are added as inputs to the neural network, as sea-ice thickness dynamics are largely driven by the atmosphere (Guemas et al., 2014). In particular, Arctic surface circulation and sea-ice movement are strongly influenced by atmospheric winds (Serreze et al., 1992), while surface temperature fluctuations also play a key role in sea-ice variability (Olonscheck et al., 2019)."

RC: **5. Lines 101 – 102: This sentence seems to be a bit out of place. Perhaps it's better to move it to the previous paragraph.**

AR: Thank you for your remark, we moved it to the previous paragraph. We had decided to put it after the forcings to emphasize that all the inputs of the neural network were normalized, which is the standard practice.

**RC: 6. Lines 112 – 115: These sentences are quite cryptic, at least to me. It is difficult to guess what they mean.**

AR: We modified the sentences: "Firstly, with  $f_\theta$ , we focus on 12 hours tendencies rather than predicting the full state directly, since small changes are harder to capture. The learning process is then split: we first train the emulator to reproduce SIT evolution, and subsequently use transfer learning so that  $g_\theta$  also respects the positivity constraint."

**RC: 7. Equations 4, A1 and A2: I don't see why you need a double-index and double-summation in the equation. A better way to simplify the equation is to use a single summation over the  $N_z = 8871$  valid grid points. Then you save the need to mention the rather cryptic sentence in lines 120 – 121.**

AR: We understand your point, but we cannot use a single summation across  $N_z = 8871$  since we are actually computing the MSE over  $N_x \times N_y = 128 \times 128$  grid cells and multiplying by the mask to only focus on the  $N_z$  pixels. We changed the sentence in line 120-121: "Note that, to simplify the equation, we did not include here the land-sea mask, which is applied in the numerical implementation to compute the loss solely on the  $N_z$  valid pixels." and kept the equations as they were.

**RC: 8. Lines 136 – 139: It is better to place the definitions of such notations in proximity to the equation that first uses them.**

AR: We modified the sentence and placed it after Eq (6b): The  $\mathbf{R}_k$ , defined by  $\mathbf{R}_k = \sigma_{\text{obs}}^2 \mathbf{I}$  with  $\mathbf{I}$  the identity matrix, are the observation error covariance matrices, and are all equals.

**RC: 9. Line 147: "considered" might be a better word than "evaluated".**

AR: Thank you for your remark, we changed the word 'evaluated' to 'considered'.

**RC: 10. Lines 158 – 159: I am not satisfied by the authors' explanation to Minor Comment 44 of my previous review. If it is a 3D-Var system then yes, inflating a diagonal B can be equivalently done by deflating R appropriately (when direct observations are used). However, this is not the case for 4D-Var because background-error covariances are implicitly propagated by the linearised model. Considering that this is only a minor point,**

perhaps it is best to remove this sentence if the authors continue to disagree with my comment.

AR: We indeed decided to remove the sentence.

RC: 11. Line 160: Please clarify that the observations are taken at each of the  $N_z = 8871$  grid points.

AR: Thank you for your remark, we added this point in the sentence: "Within the DAW of 16 days, with observations taken on each of the  $N_z$  grid points, and on every second day starting on day 2,..." "

RC: 12. Equation 9: What does the subscript  $\mathbf{F}_{k \times N_f \mapsto 0}$  mean? "

AR: Thank you for your remark, there was a small error we meant  $\mathbf{F}_{0 \rightarrow k \times N_f}$  to represent the forcings  $\mathbf{F}_0, \mathbf{F}_1, \dots, \mathbf{F}_{k \times N_f}$  corresponding to the  $k \times N_f$  iteration of the emulator. We corrected it.

RC: 13. Lines 181 – 182 and 286 – 287, on the choice of verification dataset for the real- observation experiments: I note the authors' response to Minor Comment 70 of my previous review, but I don't agree with their claim that there would be insufficient observations to assimilate [and to meaningfully constrain the sea-ice thickness field] when a subset of observations is taken away from the assimilation and used solely for verification. Verifying against assimilated observations could sometimes lead to misleading interpretations of results. A better and more conventional way of verification in the absence of a "truth" or independent model field is to compute so-called O-minus- B (observation minus background forecast) statistics, that is, comparing a short forecast (with a valid time after the end of the DAW) against yet-to-be-used observations (i.e. observations to be assimilated in a future assimilation cycle) in observation space. In this way, the verification dataset (the yet-to-be-used observations) is independent of the dataset to be verified. I suggest that the authors show at least some results from O-minus- B verification and, if they decide to retain the existing verification results (against used observations), emphasise more clearly (even multiple times) the caveats of verifying against used observations.

AR: Thank you for your remark. We would like to clarify that when assimilating real observations, and when comparing against neXtSIM-F (dashed lines in Figs. 8 and 9), we evaluate forecasts outside the DAW (up to 9 days). This ensures that the statistics are computed against observations

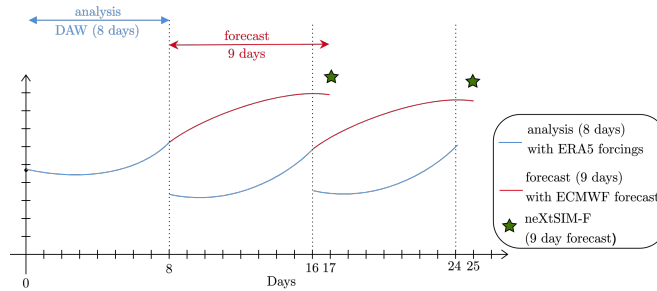


Figure 1: Setup for CS2SMOS assimilation. After 8 day DAW with daily assimilation of CS2SMOS, we continue with a 9 day forecast using HRES atmospheric forecast for the atmospheric forcings. The 9-day forecast is compared with neXtSIM-F. There is 21 assimilation cycles spanning from 2020-10-24 to 2021-04-02.

that have not yet been assimilated. To emphasize this point, please refer to the schematic in Fig. 1. We agree that this could be stated more clearly in the text. We removed the sentence in Line 285 "Note that ground truth is a quantity that is assimilated, and hence that the analysis score should not be over-interpreted" to avoid confusing readers, since that sentence discusses the analysis, while the subsequent results concern the forecast. We added a sentence in Line 285: 'We compare the forecasts from the 4D-Var analysis (after the 8-day window) with CS2SMOS observations that have not yet been assimilated, following the standard observation-minus-background (O-B) approach'. Additionally, we stressed the fact that the forecast RMSE (for which the results are discussed compare forecasts against yet-to-be-used observations in Line 183: "In the case of real observations, the forecast RMSE is computed by comparing our scheme with observations that have not yet been assimilated."

**RC: 14. Lines 208 – 213: Reading these few lines in combination with the authors' response to Minor Comment 23 (of my previous review) confuses me even more. If the noise is not properly log-normal, then perhaps it is best to get rid of the use of the term (throughout the article) but just use "noise type X" after defining it.**

**AR:** Thank you for your remark, we never use the term log normal for this noise in the manuscript. However, we forgot to mention its name 'cond-clipped' in the sentence to make it clearer. We changed the sentence to: "This noise definition is called cond – clipped in the following."

**RC: 15. Line 211:** The clause “based on...” is unclear.

AR: Thank you for your remark, for more clarity, we added a comma in the sentence : ‘based on  $SIT_{\min}$ , the corresponding  $0m$  thickness in the normalized space.’

**RC: 16. Line 224:** What do you want to convey here by using the term “single trajectory”? You can simply say something like “the 4D-Var experiment is run for  $N$  cycles from [date] to [date].”

AR: Thank you for your remark, we changed the sentence to: “The 4D-Var is run for  $N_{\text{cycle}} = 45$  cycles from January 1st 2017 to December 21th, 2018.”

**RC: 17. Line 231:** When you say the “non-diagonal terms in the B matrix”, do you refer to the “implied” B matrix, i.e.  $\varphi_m \varphi_m^\top$  ?

AR: Thank you for your remark, yes we refer to the ‘implied’ B. We added the word ‘implied’ in the sentence for clarity.

**RC: 18. Line 233:** Please mention that here you are converting the results into dimensional quantities (presumably for easier interpretation).

AR: All results are given in the physical space, indeed for easier interpretation. We will add a sentence in Sec.4.4 where the metrics are defined : “Note that the RMSE and the mRMSE are defined in the physical space, for an easier interpretation.”

**RC: 19. Line 239:** “largest corrections” à “largest negative corrections”?

AR: Thank your for your remark, we changed the sentence to “The largest negative corrections...”

**RC: 20. Line 247:** I note the authors’ response to Minor Comment 66 of my previous review. However, from the sentence “The analysis from Fig. 4(top)...” onwards, you are moving back to a general description of the results, so perhaps you could begin the sentence with “Overall,” or similar words for easier readability.

AR: Thank you for your remark, we added the word 'Overall' at the beginning of the sentence.

**RC: 21. Lines 274 – 275: I don't understand the clause "with the coefficient...". Also, I would like to see the graph – provided by the authors in response to Minor Comment 26 of my previous review – to be included in the article itself.**

AR: The clause 'with the coefficient...' corresponds to the units required in Minor Comment 26 of your previous review. We changed the word 'coefficient' to 'coefficients' and added the required graph.

**RC: 22. Lines 298 – 301: I struggle to understand this, even after reading the authors' response to Minor Comments 77 and 79 of the previous review. After all, the ERA5 dataset is only used here as a forcing, so it shouldn't matter whether the dataset is produced by assimilating observations or not.**

AR: One argument we may not have emphasized enough is that the neXtSIM-F forecast relies on weather forecasts, as it operates in a 'real' operational setting. To ensure a fair comparison between our method and neXtSIM-F, we also rely on ECMWF atmospheric forecasts. We have kept the arguments as written in the previous version of the manuscript.

**RC: 23. Lines 305 – 306: neXtSIM-F uses nudging instead of assimilation, so it is strange to refer to the "neXtSIM-F assimilation forecast".**

AR: We modified the sentence to remove the word 'assimilation': 'the assimilation of real data into our 4D-Var works and yields RMSEs similar to those of the neXtSIM-F forecast'.

**RC: 24. Lines 327 – 328: Isn't the IIEE measured in terms of areas (instead of percentages)?**

AR: Thank you for spotting those errors, we updated the manuscript with the results in km<sup>2</sup>: "The IIEE<sub>SIT</sub> serves as a reliable indicator of how accurately the MIZ is positioned. The IIEE of neXtSIM-F is 469km<sup>2</sup> and is slightly better than that of the assimilation run 522km<sup>2</sup>. At the end of each forecast, the assimilation run shows an improvement of 1400km<sup>2</sup> in IIEE compared to the free run, highlighting a significant enhancement in MIZ positioning achieved through the 4D-Var-EOF assimilation."

**RC: 25. Line 356: I don't understand why the emulator adjoint is evaluated as many as 8 times in each DAW.**

AR: Thank you for your remark, the sentence was incomplete, it is evaluated 8 times in the real observations case and 32 times in the twin experiment case. In any case, we removed the sentence.

**RC: 26. Line 376: Saying CryoSat-2 data are spatially sparse is somewhat misleading. It is dense along one dimension (along-track) and sparse in the other dimension.**

AR: Thank you for your remark, we changed the sentence to: 'However, current satellite data either provide temporally sparse observations with dense along-track but sparse cross-track coverage (like CryoSat-2) or smooth, time-averaged full coverage products, which introduces inherent correlations between the observations errors (like CS2SMOS).'

**RC: 27. Line 379: What does "average observation window" mean?**

AR: Thank you for your remark, the wording was wrong, we changed the sentence to: "It is important to note that our comparison with neXtSIM-F occurs 9 days after the last batch of assimilated observations, thus falling outside the DAW."

**RC: 28. Lines 436 – 437: It is a strange and abrupt ending to the main text of the article. What is the distinction between Sections 7 and 8? Perhaps it is better to combine them and end the article with a more general remark.**

AR: Thank you for your remark, indeed the end of the conclusion was a little abrupt. We moved the final paragraph of the discussion into the conclusion to ease the reading: 'These results are promising and demonstrate the potential for using model emulators in data assimilation, particularly with classical methods in real-world applications. Furthermore, it could be interesting to see the impact of assimilating several variables, like SIT and sea-ice concentration (SIC) onto an emulator. '

**RC: 29. Line 497: "background strategy"à"background-error strategy"**

AR: Thank you for your remark, we corrected the sentence.

**RC: 30. Appendix D is very short with seemingly something omitted, and Figure D1 is not discussed anywhere.**

AR: Thank you for your remark, note that results from Fig D1 are mentioned in the Discussion. Nonetheless, we added some comments about Fig.D1 in the appendix: "In this section, we present additional results showing the full two-year assimilation, including the one year spin-up cycles which are not included in the metrics computation. This shows the seasonality of the analysis RMSE, with increase during highly thermodynamics-driven period (spring melting and winter refreezing). The second part of the curve corresponds exactly to the RMSE shown in Fig.6a), where the spring RMSE increase is also evident in 2017, confirming this recurrent behavior during periods of intense thermodynamic change. In addition, we note a pronounced peak in November 2017, coinciding with the refreezing period."

**RC: 31. Line 552: I believe "with for" is a typo.**

AR: Thank you for your remark, a comma was missing, hindering the reading of the sentence: it has been changed to 'where  $\mathcal{J}_b^n$  corresponds to the background term of the cost function at cycle  $n$  and  $\mathcal{J}_o^n$  corresponds to the observation term of the cost function at cycle  $n$ ; with, for the initial cycle, a value of 1 for  $\lambda_{a,0}$ .'

**RC: 32. Line 563, the word "ratio": The ratio of what?**

AR: Thank you for your remark, we meant 'The ratio between the RMSE of  $x_{a,\text{inflation}}$  and  $x_{a,\text{noinflation}}$ ', it has been corrected: in which is plotted the ratio between the two analysis  $x_{a,\text{inflation}}$  and  $x_{a,\text{noinflation}}$ , with  $x_{a,\text{inflation}}$  the analysis with the adaptive inflation scheme with  $\lambda_m$  set to 3 and  $x_{a,\text{noinflation}}$  the analysis of the run with Gaussian noise without any inflation scheme, in 2018,...

## References

- Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., Fučkar, N. S., Germe, A., Hawkins, E., Keeley, S., Koenigk, T., y Mélia, D. S., and Tietsche, S. (2014). A review on arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 142(695):546–561.
- Olonscheck, D., Mauritsen, T., and Notz, D. (2019). Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations. *Nature Geoscience*, 12(6):430–434.
- Serreze, M. C., Maslanik, J. A., Barry, R. G., and Demaria, T. L. (1992). Winter atmospheric circulation in the arctic basin and possible relationships to the great salinity anomaly in the northern north atlantic. *Geophysical Research Letters*, 19(3):293–296.