

Review of “Four-dimensional variational data assimilation with a sea-ice thickness emulator”

by Charlotte Durand et al.

Version 2

Date of review: 30 July 2025

Summary and Major Comments

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.

Minor Comments

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.
2. Line 46: The wording “relies on” in this context is a bit odd. I suggest replacing it with “is achieved by running”.
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.”
4. Lines 98 – 99: You consider the variables you named, but for what purposes?
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.
6. Lines 112 – 115: These sentences are quite cryptic, at least to me. It is difficult to guess what they mean.

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.
8. Lines 136 – 139: It is better to place the definitions of such notations in proximity to the equation that first uses them.
9. Line 147: “considered” might be a better word than “evaluated”.
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 \mathbf{B} can be equivalently done by deflating \mathbf{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.
11. Line 160: Please clarify that the observations are taken at each of the $N_z = 8871$ grid points.
12. Equation 9: What does the subscript $k \times N_f \mapsto 0$ mean?
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.
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.
15. Line 211: The clause “based on...” is unclear.
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].”
17. Line 231: When you say the “non-diagonal terms in the \mathbf{B} matrix”, do you refer to the “implied” \mathbf{B} matrix, i.e. $\boldsymbol{\varphi}_m \boldsymbol{\varphi}_m^T$?

18. Line 233: Please mention that here you are converting the results into dimensional quantities (presumably for easier interpretation).
19. Line 239: “largest corrections” → “largest negative corrections”?
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.
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.
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.
23. Lines 305 – 306: neXtSIM-F uses nudging instead of assimilation, so it is strange to refer to the “neXtSIM-F assimilation forecast”.
24. Lines 327 – 328: Isn’t the IIEE measured in terms of areas (instead of percentages)?
25. Line 356: I don’t understand why the emulator adjoint is evaluated as many as 8 times in each DAW.
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.
27. Line 379: What does “average observation window” mean?
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.
29. Line 497: “background strategy” → “background-error strategy”
30. Appendix D is very short with seemingly something omitted, and Figure D1 is not discussed anywhere.
31. Line 552: I believe “with for” is a typo.
32. Line 563, the word “ratio”: The ratio of what?