Review of:
Title: Evolving beyond collapse: An adaptive particle batch smoother for cryospheric data assimilation
Author(s): Kristoffer Aalstad et al.
MS No.: egusphere-2026-831
MS type: Development and technical paper

Review by: Steve Margulis

Review criteria:

Principal criteria	Excellent (1)	Good (2)	Fair (3)	Poor (4)
Scientific significance: Does the manuscript represent a substantial contribution to modelling science within the scope of Geoscientific Model Development (substantial new concepts, ideas, or methods)?	**X**			
Scientific quality: Are the scientific approach and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)? Do the models, technical advances, and/or experiments described have the potential to perform calculations leading to significant scientific results?	**X**			
Scientific reproducibility: To what extent is the modelling science reproducible? Is the description sufficiently complete and precise to allow reproduction of the science by fellow scientists (traceability of results)?	**X**			
Presentation quality: Are the methods, results, and conclusions presented in a clear, concise, and well-structured way (number and quality of figures/tables, appropriate use of English language)?	**X**			

General comments: The manuscript is excellent in the areas of scientific significance, quality, reproducibility, and presentation quality. It makes a significant contribution to the literature on the development of cryospheric data assimilation methodologies. The substantive comments listed below are suggestions for improving the manuscript in terms of clarity or elaboration on results. They are enumerated in order of appearance in the manuscript. The minor text editing suggestions are to help remove any typographical or other errors in the final manuscript.

Substantive comments:
1. Figure 1 does not appear to be referenced anywhere in the manuscript text. It is worth adding a sentence or two that walks the reader through the schematic figure.

2. Section headers 3.2 and 3.3 respectively: Consider adding "Experiment 1: …" and "Experiment 2: …" to clarify to the reader there are two main experiments performed in this study.
3. Sections 3.2 and 3.3: In both sets of experiments, you should identify what the prescribed snow depth measurement errors were. This provides important context when looking at the results in terms of how well the snow depth estimates "should" fit the observations.
4. Sections 3.2 and 3.3: For the ES cases, it seems from the results you are updating/estimating parameters and not just states. If so, you should explain how this is done, e.g., via state augmentation or another approach.
5. Section 3.2, p. 22 when discussing the prior parameters: Presumably the perturbed parameters are uncorrelated. If so, it may be worth mentioning that explicitly.
6. Section 3.3: Was the "gold-standard" MCMC approach used as a reference for this experiment? If so, perhaps make that clear, but if not, explain why not. Perhaps related to this, is there a reason that the KLD metric is not used in these experiments? Is that because the MCMC is not applied?
7. Section 3.3: For the model parameters that were perturbed, can you provide more context for why they were chosen?
8. Figures 2-5, MCMC samples in the Forcing Correction Parameters subplots: The color mapping is hard to see and not explained. I presume it is related to the density of the points at a given location in the parameter space. I don't think that is so important (and hard to see) and would suggest just making those points a single color.
9. Figures 2-5, red dots in legend box: The use of red dots in the legend (as "observations") could be confused with the red dots used in the parameter space. I suggest using different colors.
10. Figures 2-5, model state space subplots: Consider whether an MCMC-derived trajectory should be included as the "gold-standard" reference.
11. Figures 2-5, model state space subplots: There seems to be an error in the date labeling convention being used on the time series plots. It is unclear and wraps so that "01/09" appears at both ends of the x (time)-axis.
12. Figure 3: Since ES is not technically an iterative method, I'd suggest using "Prior" and "Posterior" labels for clarity rather than "Iteration 0" and "Iteration 1".
13. Figure 6: Perhaps explain or reference how you estimate the "kernel density estimates".
14. p. 31, "… thousands of observations …": This comment is specific to the batch/smoothing you are doing in this paper. Can you comment here or elsewhere on how/why this would be different (more tractable) if using a filtering approach?
15. p. 31, "On average, AdaPBS required 8 iterations …": Given that the iterative/adaptive aspect of this method is novel and discussed as a benefit, is there any useful insight to be gained by presenting the interannual and inter-site number of iterations that were seen across site-years and discuss why those differences may have occurred? Perhaps a broader short discussion (not necessarily here) of when/why/where the number of iterations is expected to be small or large would benefit the reader.

16. p. 31, “The site with the largest run times …”: It would benefit the reader (here or later) to have some discussion on the scalability to spatially-distributed cases.
17. Figure 7: I would suggest either omitting or changing the naming of the “reference” curve. From what I gather it is a deterministic version of the open-loop and is therefore somewhat redundant to the open-loop curves, but “reference” could be misinterpreted as the target for which the estimates should be compared.
18. p. 33, “… more challenging high-dimensional setting”: Here, when you say “high-dimensional” are you mainly referring to the number of measurements, parameters, or both?
19. p. 33, results section: Given that this experiment involved a significantly larger number of parameters, I think it would be worth it to show results on how the parameter estimates differ (or are similar) between the ES-MDA and AdaPBS cases (e.g., perhaps along the lines of Figure 6).
20. Depending on how some of the comments above are handled, I could envision a paragraph in the conclusions discussing some of the implications around: scaling up to spatially-distributed applications, tradeoffs with filtering approaches, expected benefits of the adaptive nature of the AdaPBS method, etc. There are some instances of text throughout that touch on these issues, but having them in one place might help orient the reader when thinking about applications and extensions to this work.

Minor text editing suggestions:

1. On p. 2: “… remotely sensed products …”. Perhaps for clarity mention you are referring to passive microwave products.
2. On p. 3: “Ensembe” misspelled.
3. On p. 5: “Unfortunately … case , …” has extra space after “case”.
4. On p. 5: “ill-possednes” misspelled.
5. On p. 8: Two instances of “long enough” could be replaced by “enough iterations” to indicate the meaning is related to iteration rather than time.
6. On p. 23: “FMSM2;” has a typographical error.
7. Figure 2: “aditive” on y-axis is misspelled.
8. On p. 28: “In our experience …”. This sentence seems a bit out of place here. It sounds more like a conclusion statement.
9. On p. 32: “In a similar vein, …”. The word “cost” appears twice in sentence. Perhaps replace last instance with “expense”.
10. Table 3 caption: Replace “prior (Pri)” with just “Prior”.