Review of "Estimation of the state and parameters in ice sheet model using an ensemble Kalman filter and Observing System Simulation Experiments"

March 27, 2025

General comments

The manuscript by Choi et al. presents a data assimilation framework to improve the projection capabilities of ice sheet models. Specifically, the performance of an Ensemble Adjustment Kalman Filter in constraining the model state (ice thickness) and basal conditions (basal friction coefficient and bed topography) of a 2D plan-view ice model is assessed. Their results indicate that assimilating more observations generally increases the accuracy of model projections, with projections for up to 200 years in close agreement with the reference simulation. The performance of the data assimilation method is sensitive to the observational error as well as the cross-track spacing and grid resolution of surface elevation data.

I believe the science behind this study is sound and aligns with the focus of The Cryosphere (TC). However, the presentation of the methodology lacks clarity, at times adding avoidable confusion (e.g., the introduction of both acronyms EnKF and EAKF). This overall issue is addressed in more detail in the specific comments below, but I strongly suggest the addition of a flowchart outlining the methodology (ice sheet model and data assimilation) and experimental design (twin experiment and OSSEs). Furthermore, parts of the experimental design are currently placed within the results section and key aspects of the results are not addressed (e.g., why is the RMSE_C for Grid_20km_σ_{h}_20_σ_{v}_10 smaller than for the same grid resolution with smaller uncertainties as well as all 10 km grid resolution experiments?). Considering the performance of the data assimilation method is sensitive to the uncertainty in surface elevation observations, I believe also determining the effect of various uncertainties in the velocity data would add further value to the manuscript (perhaps as supplementary material). Finally, potential reasons/explanations for model results are often missing, e.g., why is the range of optimal localization radius (4 - 8 km) a lot smaller than suggested by previous studies (4 - 120 km)? I recommend the authors also take my specific comments listed below into account.

Specific comments

L25: This sentence is very similar to the second sentence in the introduction. Instead, consider opening with a sentence about the different DA methods (variational vs. methods leveraging time-varying observations). Then proceed to discuss advantages/disadvantages of each.

L28: Double brackets.

L33: Consider starting a new paragraph before *Alternatively*.

L37 – *assimilation period*: Readers unfamiliar with DA might not know what exactly you refer to here. It becomes a lot clearer later on, but it would be nice to have a brief definition here (similar for other DA-specific terms, e.g., *data denial experiments* in L61).

L40: Move further up to the rest of the discussion on variational methods.

L42: Consider introducing the term ensemble DA in general before describing the specific EnKF (e.g., ensemble DA vs. variational methods).

L54: Why are ensemble DA methods less commonly used in ice sheet modelling?

L63 – *(OSSEs)(OSSEs, ...)*.

L65 – *appropriate observation error distribution*: How do you determine if the distribution is appropriate or not?

L71: Although it is addressed in more detail in the next sentence, I believe adding *(ice thickness)* just after *model state* would add clarity.

L75 – *estimated state and parameters*: For consistency, I recommend using estimated *model* state and parameters throughout the manuscript.

L75 – *true reference values*: At this point in the manuscript, it is not clear what the *true reference values* are and how you obtain them.

L83-84: Remove this sentence and add the reference to the description of the specific sections in the text above.

L86-88: Repetition of the text just before the Methods section.

L101: I am not familiar with this specific method, but a standard deviation of 500 m seems quite large considering the bed varies only between $z_{b,deep} = -720$ m and ∼500 m (really difficult to see in Fig. 1a)

Fig. 1: I believe using two separate 2D plots instead of the 3D plot would make the identification of certain details and interpretation of the plot a lot easier. As you are already showing the bed topography in Fig. 2, I recommend combining Fig. 1 and 2 into a single plot with panels a: ice surface elevation, b: bed topography, and c: ice velocity. Note that the rainbow colour scheme is not in line with the journal guidelines. You can check all of your plots with the colour blindness simulator (https://www.color-blindness.com/coblis-color-blindness-simulator/). The fonts in panel b are too small and I recommend using a different colour for your contour lines.

Fig. 2: Why are you using such an asymmetrical (about y=40 km) bed topography compared to the commonly used symmetrical approach in idealized studies? The y-label in Fig. 2 indicates the domain ranges from 20 to 100 km, whereas in Fig. 1 it is 0 to 80 km. You also might want to remove the white margins at the top and bottom. How can the bed elevation be -1500 m when Eq. 1 limits the bed to $z_{b,deep} = -720$ m?

L103: Consider adding an additional panel (d) showing the triangular mesh to the new Fig. 1. Are you using adaptive mesh refinement, e.g., following the grounding line?

L104: I suggest adding another panel (e) for the basal friction coefficient to Fig. 1 or at least refer to Fig. 6a here.

Eq. 5: In case you are working in LaTeX, I recommend using *\left(* and *\right)* to get brackets of the correct size.

Eq. 6: Same as for Eq. 5

L114: Is C in Eq. 7 different from the one described in Eq. 4? If not, then remove *C is a friction coefficient*.

L117: Remove *equal*.

L124: Do you consider a melt rate of 200 m/yr realistic given that maximum present-day melt rates are around 100 m/yr?

L125-127: This belongs into results.

L129: Sec. 2.3 and 2.4 are referenced before 2.2.

L134: modified version of the *Ensemble* Kalman Filter? As I mentioned above, using EnKF and EAKF is confusing, especially since EAKF is introduced but only used within this paragraph.

L136-143: This is where I think a flowchart would really help the reader to follow the details of your method. Ideally, the flowchart should outline the details of the EAKF and how it relates to your specific study. For example: How do ensemble members differ? What exactly is your model forecasting? How is the observation window specified? Which ice sheet variables are considered in the state vector? What are the state variables?

L142: How would adding extra variables, like surface velocity, to the state vector affect your results?

L143-144: EnKFs or EAKFs? Does this challenge arise in your study? What is the ensemble size? What are the independently observed degrees of freedom in your case?

L146: stability of the *EnKF*?

L150: Add more detail about what exactly you mean by *sampling errors*.

L151: What localization and inflation parameters are you examining?

Sec. 2.3: You either need to embed this information into the previous section or clearly outline at the beginning of the methods section that you are first describing the EnKF in general and then how this general structure relates to your specific setup (with references to sections). Again, a flowchart linking the general structure to your experiments would be helpful.

L154: EnKF or EAKF? If EnKF, then why bother introducing EAKF?

L164: Why did you decide to use lower standard deviations? What is the plausible range?

L178-179: Are you assuming that the friction coefficient and bed topography are uncorrelated?

L184: What radii did you explore?

L186: Ensemble size of 30 to 100, but what steps exactly?

Fig. 3: The font size is too small. This is generally the case for a lot of plots and I will refrain from mentioning it again afterwards. Otherwise, I think this is a great figure supporting the description of your OSSEs.

L221 – *reference model mesh*: Do you mean the mesh used in the reference simulation?

L230-231: This information needs to come earlier.

L234: Can you provide any insight as to why these values lead to the minimum RMSEs?

L238: Why is that expected?

L240: You are using *inflation parameters* in the text but *inflation factor* in Fig.4.

L243-245: For ensemble size 100, the RMSE for friction coefficient does NOT continue to decrease steadily (increase at t=7a). The other two ensemble sizes also show a small increase just after 5 years. Similar peaks are also visible for the bed topography and ice thickness. What is causing this increase in RMSE?

L245: A larger ensemble size (100 vs. 50) actually leads to a larger RMSE after t=15 a for all panels in Fig. 5. Why do you think this is the case? And what does it mean for the design of future experiments?

L248: Why did you choose a localization radius of 4 km when bed topography and ice thickness showed a minimum RMSE for 6 km with a significant increase for smaller radii?

L250: Somewhere you should state explicitly that a localization radius of 4 km and an inflation parameter of 1.12 are your *optimal DA configuration.*

L254-256: You describe the results for the friction coefficient and bed topography, but what about the ice thickness?

L257-262: This is a description of your experimental design and should be in the methods section.

L264: Just to make sure I understood it correctly. The deterministic forecast uses the mean (across all ensemble members) basal friction coefficient and mean bed topography but is a single simulation (compared to running a simulation for all ensemble members and then calculating the ensemble mean).

L268-269: I think it is quite interesting that the deterministic forecast, which is based on the ensemble mean basal conditions, follows the reference simulation relatively closely while most of the individual ensemble member simulations show a much smaller ice volume change. I suspect this is due to non-linearities in the system, but it might be worth having a closer look at this.

L271-273: So assimilating more observations leads initially to a better agreement but increases the difference in ice volume change at the end of the forecast period. This should be addressed in more detail in the discussion.

L273-274: Discussing the increase in the rate of mass loss after 100 a and its implication for sea level rise projections over the next century (compared to 200 yr projections) could be a nice additional takeaway.

Fig. 4: Why are some experiments diverging? Why does the diverging area shift to smaller inflation factors as the ensemble size increases? What is the reason for the sharp increase in RMSE for localization radii smaller than the minimum RMSE? Use friction coefficient, bed elevation, and ice thickness as labels next to the colour bar.

Fig. 5: I suggest using RMSE_C, RMSE_B, and RMSE_H as y-labels or adding friction coefficient, bed elevation, and ice thickness as titles. The description of panel c is missing. The colour coding is somewhat confusing because you are using the same colours as in Fig. 4 but they do not represent the same thing (ensemble size vs. friction coefficient/bed elevation/ice thickness).

Fig. 6: Why did you choose a more or less symmetrical (about y=40 km) friction coefficient but a very asymmetrical bed topography? The units are missing in the colour bar label. You might want to increase the spacing between panels to make it clearer which text corresponds to which panel.

Fig. 7: What causes the sharp grounding line extent towards higher x values at y=10 km in the *no assimilation* panel? Units are missing. Increase spacing between panels.

Fig. 8: Why did you not show the difference in Fig. 6 and 7? What causes the *checkerboard* pattern?

Fig. 9: I suggest adding a legend for the different lines. Change *reference run* to *reference simulation*. Change *forecast simulation* to *deterministic forecast simulation* (all of them are forecast simulations).

Sec. 3.2.: Do not use abbreviations as section titles.

L285: Change *10 to 15 km* to *15 km*. Or did you also test, e.g., 13 km across-track spacing?

L286: I am not sure what you mean by *the performance declines due to suboptimal choices for inflation and localization parameters*. You are examining the effects of these parameters here, so shouldn't you be able to determine the optimal choices? Do you mean the optimal choices are outside your tested parameter ranges? If so, you need to show results supporting this claim.

L287: Start a new paragraph before *For the gridded*.

L288: Add the reference to Fig. 11 to the end of this sentence (currently at the end of the paragraph).

L289: range *of* localization

L292 – *we conducted*: Use present tense throughout the manuscript (e.g., same issue in L403 *presented*).

Fig. 10: In panels (f) and (i), why is the RMSE of localization radius 6 km and inflation factor 1.04 much larger than the surrounding values?

L297-299: Add reference to Table 2.

L301 – *RMSE values continue to decrease until the end*: Add reference to Fig. 12. Panels c, f, and i show an increase in RMSE at the end.

L301-303: Add reference to Table 3.

L304 – *marginal improvements ... after 10 - 15 years*: Add reference to Fig. 13. Again, RMSE actually increases in panels f and i. What do you think causes this increase?

L304-305: This is discussion.

L307: Add reference to Table 2. For RMSE_C, the smallest uncertainty leads to the second-largest RMSE of all tested uncertainties. So DA performance does not necessarily decrease as uncertainty increases!

L308-309: Add reference to corresponding panel in Fig. 12. For the 10 km across-track data, the maximum difference in RMSE_C is 15.43. For the 5 km data, it is only 8.65. Although the maximum difference is much smaller in the 5 km case, you argue that it shows a decrease in performance as uncertainty increases while the larger 10 km difference indicates a consistent performance across different uncertainty levels. I believe your statement is primarily based on the RMSE_B and RMSE_H results, but these details need to be spelled out!

L309-311: Adding to my previous comment, if you compare Track_15km_σ_h_10_σ_v_10 to Track_15km_σ_h_15_σ_v_10, and Track_15km_σ_h_20_σ_v_10, the DA performance increases as uncertainty increases for RMSE_C, RMSE_B, and RMSE_H. This needs to be stated clearly and discussed in detail!

L311: Add a reference to the corresponding panel in Fig. 12. They actually show an increase in RMSE after 15 to 20 years.

L311: Add new paragraph before *With the 1 km gridded*.

L312: Add a reference to Table 3. What about the friction coefficient? For the friction coefficient and Track_15km, the highest uncertainty level has the smallest RMSE and, therefore, the best performance. So DA performance does not necessarily decrease as uncertainty increases!

L313 – *DA performance does not vary significantly across different uncertainty levels*: I don't think I agree with this statement. Again, just looking at RMSE_C, the maximum difference across uncertainty levels at 1 km resolution is 2.84, while it is 10.28 at 10 km and 30.8 at 20 km. So if anything, the performance varies more significantly with coarser grid resolution. Additionally, the RMSE_C for the coarsest grid resolution (20 km) and highest uncertainty level is smaller than all other RMSE_C values at 20 km AND 10 km resolution! As the RMSE_C with the highest uncertainty and the coarsest across-track resolution is also smaller than all other values at this resolution, it seems unlikely that this is just a coincidence. So this really needs to be addressed.

Fig. 12: Use RMSE_C, RMSE_B, and RMSE_H in y-labels. Panel labels in the second and third rows are the same. Actually, even the subplots themselves look the same. The panel labels seem to be just a copy/paste issue in your Python code, not sure about the actual data.

Fig. 13: Same issues as for Fig. 12. What causes the increase in panel f after 20 years? Why is there such a rapid increase in RMSE in, e.g., panel d between 5 and 10 years? Why does this rapid increase get muted for coarser resolutions? A similar pattern occurs in Fig. 12.

L316 – *fast flowing regions*: You haven't mentioned fast-flowing regions before. Are you referring to *areas around the grounding line, where the signal-to-noise ratio of velocity is relatively high*? I'd argue that large differences also occur for y=70-80 km and x=450-640 km, which seems to be a relatively slow-flowing region.

L319 – *fast flowing regions*: Again, what about the region between y=70-80 km and x=450-640 km?

L322: *more* accurate.

L333 – *assumptions on the initial ensemble*: Be more specific.

L337 – *relatively small ensemble size*: Be precise.

L337-338 – *previous studies*: References are missing.

L342 – *a larger ensemble size could provide advantageous*: Or not. Fig. 5e shows a larger RMSE for 100 members than 50 members by t=30 years.

L343: What exactly should these studies investigate to identify the optimal approach?

L345 – *similar to values from earlier studies*: References are missing. What are these values?

L347: 4-120 km is a lot wider range than your 4-8 km. What causes these differences?

L351: ... assimilating more observations, *i.e. more assimilation years*, to estimate ...

L352 – *improves accuracy of model projections*: In general, yes, but the difference in ice volume change at t=200 years is larger in Fig. 9 panel b than panel a (between red and blue line).

L354 – *XX*: Add numbers.

L358 – *in this study that*: Replace with *here*.

L358 – *observations maintains*: Replace with *observations while maintaining*

L361: I suggest restating what OSSE means for readers quickly skimming through the manuscript.

L361: Remove *in this study*.

L361 – *the capabilities of OSSEs*: Replace with *their capabilities*.

L365 – *(Table 2 and 3)*: If you include references here then you should also include them in previous paragraphs of the discussion.

L365-366: As indicated previously, this is not the case for RMSE_C in Table 3.

L369-371: As mentioned above, additional data points can actually have a negative effect.

L373-374: I disagree. The friction coefficient has the largest differences in RMSE across uncertainty levels (for all resolutions in Table 2 and similar for Table 3), so it is actually the most sensitive.

L377: *will* or *should*?

L379-380: Your *future studies should address* sentences are spread out across the entire discussion. I recommend bundling all of them into one single paragraph at the end of the discussion.

L398-399: Again, this is not always the case.

L400: What does *great accuracy* mean? Be precise.

L404 – *Different levels of observational uncertainty*: Do you mean *smaller levels of uncertainty*?

L410: Will you also upload the data files?

I hope the authors find my comments helpful.

Sincerely, Kevin Hank