Estimation of the state and parameters in ice sheet model using an ensemble Kalman filter and Observing System Simulation Experiments
– Authors' response (RC1) –

Youngmin CHOI et al.

May 1, 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 topog- raphy) 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).

We thank the reviewer for reviewing the manuscript and constructive comments. We will thoroughly address specific comments and aim to clarify all points as clearly as possible. We appreciate the suggestion to include a flowchart outlining the methodology and will add it into the revised manuscript.

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_20 km_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.

We agree with the reviewer that the some of the results lack sufficient explanation. We will address this in the revised manuscript. Furthermore, we will conduct additional experiments with varying uncertainties in the velocity data and include the new results in the updated version.

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.

We will revise this sentence.

L28: Double brackets.

We will fix this.

L33: Consider starting a new paragraph before Alternatively.

We will make this a new separate paragraph as suggested.

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).

We will include brief definitions of several DA terms for clarity.

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

We will consider rewriting this paragraph to improve clarity.

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

We will add introductory sentences to clarify the ensemble DA.

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

Ensemble DA methods are less commonly used in ice sheet modeling primarily due to historical limitations in observational data and computational cost. Ensemble approaches rely on consistent, time-varying observations with well-characterized uncertainties. However, surface observations for ice sheets have often been sparse, noisy, or temporally inconsistent, which are less suitable for ensemble DA. The ice sheet modeling community has traditionally relied on adjoint-based (variational) inversion methods for parameter estimation using time-invariant mosaics or composites data (e.g., multi-year averaged surface velocity fields). Additionally, ensemble methods typically require multiple forward model runs, making them more computationally demanding than (static) variational approaches. These challenges led to the relatively limited adoption of ensemble-based DA in ice sheet modeling. We will revise this section in the manuscript accordingly for clarity.

L63 – (OSSEs)(OSSEs, ...).

We will fix this.

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

We meant a prescribed observation error distribution representative of real measurement uncertainties. We will revise this sentence to clarify our intent.

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

We will add that.

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

We will use consistent terminology 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.

We will consider rephrasing this sentence for clarity.

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

We will remove this sentence.

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

We will revise these sentences.

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 zb,deep = -720 m and ∼500 m (really difficult to see in Fig. 1a)

The midpoint displacement method generates a 2D surface by iteratively subdividing a grid, assigning random heights to corner points, and interpolating midpoints with added random displacement. The magnitude of the displacement is scaled by a standard deviation that decreases with each iteration as $2^{0.5H}$, where H is the roughness factor, set to 0.7 in this study. While the initial standard deviation of 500 m may seem large relative to the vertical range of the bed topography, it is used as a starting point in the midpoint displacement algorithm and is progressively reduced at each iteration based on the roughness factor. This results in a realistic, spatially correlated roughness pattern with limited high-amplitude variations. Additionally, the current value of $z_{b,deep} = 720m$ represents the base shape of the bed before roughness is added and the final bed elevation reaches depths of approximately -1,500 m. We will add these details to the text and revise the Fig. 1a as suggested below.

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.

We will revise the figure to address all of the reviewer's suggestions.

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?

Applying the midpoint displacement method results in an asymmetrical bed topography, which may better reflect realistic subglacial features, although we use an idealized twin experiment in this study. We will include this explanation and revise the figure as suggested.

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?

We use an adaptive mesh based on ice velocity and will include a new figure for the mesh in the revised manuscript.

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

We will refer to Fig. 6a to avoid repetition.

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

We will fix the bracket.

Eq. 6: Same as for Eq. 5

We will fix the bracket.

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

C is the friction coefficient, and C_x and C_y are the x and y components of C, respectively. We will add this to the text.

L117: Remove equal.

We will remove it.

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

We set the melt rate to 200 m/yr at a depth of 800 m, which results in an actual melt rate of approximately 170 m/yr beneath the ice shelf. We agree that this melt rate exceeds the maximum observed present-day basal melt rates. However, in this study, we chose this value to create a strong dynamic response in the model over a 200-year forecast period, ensuring that the effects of data assimilation could be clearly evaluated. The elevated melt rate is not meant to represent a realistic present-day climate, but rather to serve as a diagnostic tool in the context of a twin experiment. We will clarify this in the revised manuscript.

L125-127: This belongs into results.

This describes the process of creating the reference run for the twin experiment rather than presenting model results. We will revise the text to clarify it.

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

We will delete this sentence here.

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.

We will clarify the use of data assimilation terminology throughout the manuscript.

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?

We will include a detailed flowchart to clarify the methodology.

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

Surface velocity is an observation we assimilate in our study, but it is not part of the state vector. We will clarify this in the text.

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?

Here, we are explaining the general case for the ensemble Kalman filter. As mentioned above, we will revise this paragraph to improve clarity and better distinguish the general description from our specific implementation.

L146: stability of the EnKF?

Yes, we will change it.

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

In the revised manuscript, we will clarify that sampling errors in ensemble based data assimilation arise due to the limited number of ensemble members used to approximate the forecast error covariance.

L151: What localization and inflation parameters are you examining?

We will add details about the localization and inflation methods here, including the associated tuning parameters for each.

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.

We will clearly outline the description of the methods at the beginning of the Methods section.

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

EnKF is the general term, and EAKF is the specific approach we use in this study. We will clarify this in the text.

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

We selected lower standard deviation values (5 m for surface elevation and 10 m/yr for velocity) to provide a simple and conservative baseline for the twin experiment. While these values are lower than those used in Gillet-Chaulet (2020), they are still within the plausible observational uncertainty ranges reported in recent literature. For example, Dai and Howat (2017) report vertical elevation uncertainties below 5 m in well-constrained regions, and Mouginot et al. (2017) report horizontal velocity uncertainties ranging from 5–20 m/yr depending on the region. We chose values at the lower end of these ranges to isolate the performance of the DA framework under favorable conditions, and we explore sensitivity to larger uncertainties in the OSSEs presented in Section 3.2. We will clarify this in the revised manuscript.

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

They are correlated, as both parameters are sensitive to common observations. We will include this explanation in the revised manuscript.

L184: What radii did you explore?

The radii explored in the experiment ranged from 2 km to 20 km. We will add this information to the text.

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

We tested ensemble sizes of 30, 50, and 100, and will specify this information clearly in the text.

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.

We will increase the font size.

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

Yes, we will clarify this in the text.

L230-231: This information needs to come earlier.

We will move this information to the Method section.

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

The localization radius is determined through a set of sensitivity experiments and is based on the expected spatial correlation length scale of the parameters, which may depend on the size of flow features or stress balance regimes. We will add further discussion on this point in the revised discussion section.

L238: Why is that expected?

The smaller ensembles tend to underestimate the ensemble spread due to sampling errors, which can lead to filter divergence. To compensate for this underestimation, higher inflation factors are often required to maintain sufficient ensemble variance. As the ensemble size increases, the sampling error is reduced, leading to more accurate estimation of error covariances and therefore requiring less inflation. We will clarify this explanation in the revised manuscript.

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

We will clarify the term for consistency.

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?

We examined the small increase in RMSE in early assimilation years and found that it is likely due to a temporary mismatch between the model forecast and the observations during this period, potentially caused by transient model dynamics or nonlinearities in the response to assimilated observations. As the assimilation continues, the filter gradually corrects these discrepancies, which leads to a subsequent reduction in RMSE. These fluctuations are not uncommon in ensemble data assimilation systems, especially in complex, nonlinear models where localized error growth can temporarily degrade performance. We will discuss this behavior in the revised manuscript.

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?

While larger ensemble sizes generally improve the accuracy of error covariance estimates, they can also increase the sensitivity of the filter to model errors or sampling noise if not properly tuned. In our experiments, it is possible that the inflation and localization parameters used for the 100-member ensemble were not optimal for later assimilation periods, leading to slightly degraded performance after year 15. This suggests that filter performance does not necessarily scale linearly with ensemble size and that tuning DA parameters for each ensemble size is critical. It also emphasizes the importance of adaptive inflation/localization techniques or diagnostics for dynamically adjusting filter settings. We will revise the manuscript to reflect this finding and its implications for future ensemble DA experiment design.

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?

The minimum RMSE for bed topography and ice thickness occurs at a localization radius of 6 km, but the RMSEs for both parameters at 4 km are also low. We chose 4 km for illustrative purposes and will clarify this in the text.

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.

The optimal inflation factor and localization radius depend on the parameter being estimated. We chose 4 km and 1.12 for illustrative purposes and will clarify this in the text.

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

The pattern in the ice thickness results is very similar to that of bed topography. In our model setup, surface elevation is defined as the sum of ice thickness and bed topography (surface = thickness + bed). Therefore, as surface observations are assimilated, improvements in bed estimates are reflected in the estimated thickness field. We will clarify this relationship and summarize the results for ice thickness more explicitly in the revised manuscript.

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

We will move this to 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).

Yes. we will clarify it in the text.

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.

We agree with the reviewer's point. The observed behavior likely results from the nonlinearities of the model. While each ensemble member represents a physically plausible realization of the basal parameters, small deviations from the true field can lead to large differences in modeled ice volume due to non-linear feedbacks. However, the deterministic forecast, initialized with the ensemble mean of the basal fields, appears to capture the overall structure of the true conditions more effectively, reducing local extremes and yielding results that are more closer to the reference simulation. We will add this discussion to the revised manuscript.

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.

Assimilating more observations leads to better agreement throughout the forecast period, including at its end. We will include specific values for ice volume loss to support this comparison in the revised manuscript.

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.

We will add this to the discussion section.

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.

When the localization radius is too small, it overly restricts the influence of observations on the state update. This can lead to underestimation of error covariances and result in filter divergence. In our experiments, this is evident when the localization radius falls below the specific threshold of each variable (e.g., 4 km for friction and 6 km for bed topography). We will include this explanation in the revised manuscript. We will also revise the figure as suggested.

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).

We will revise the figure as suggested.

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.

We will revise the figure as suggested.

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.

The asymmetrical grounding line position is caused by the asymmetrical bed topography. We will revise the figure as suggested.

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

We chose to show the difference in ice thickness in Fig. 8 because changes in thickness are difficult to detect visually from the similar figure as Fig. 6 and 7. The artifacts observed in the ice thickness are the result of the conditional random fields generated using the Kriging method, which can produce "bull's eye" patterns commonly observed between observation points. We will clarify this in the revised manuscript.

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).

We will revise the figure as suggested.

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

We will change the title.

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

We will revise it as suggested.

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.

We will remove the phrase "suboptimal choices for parameters" and rephrase this paragraph.

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

We will revise it as suggested.

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

We will revise it as suggested.

L289: range of localization.

We will revise it.

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

We will revise the manuscript to consistently use the present tense.

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?

The RMSE values are calculated over the entire domain, and localized errors can increase the overall RMSE. We will include this discussion in the revised manuscript.

L297-299: Add reference to Table 2.

We will add the reference.

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.

We will add the reference.

L301-303: Add reference to Table 3.

We will add the reference.

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?

We will add the reference. We examined the small increase in RMSE near the end of the assimilation period and found that it is likely due to a temporary mismatch between the model forecast and the observations, potentially caused by transient model dynamics or nonlinearities in the response to assimilated observations. We will discuss this behavior in the revised manuscript.

L304-305: This is discussion.

We will move this to the Discussion section.

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!

We agree with the reviewer that this paragraph is not well presented. We will revise it to clearly reference Table 2 or 3 and Fig. 12 or 13 separately, in order to distinguish between the RMSE values at the end of the assimilation period and the changes in RMSE during the assimilation period.

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!

We are referring here to changes in RMSE values during the assimilation period. We will clarify this in the revised manuscript.

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!

Again, we are referring here to changes in RMSE values during the assimilation period. We will rephrase this paragraph to clearly separate the results shown in Table 2 and Figure 12.

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

We will add the reference. Again, we will add discussion on this increase in RMSE.

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

As suggested, we will separate the discussion of the 1 km results into a new paragraph.

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!

We will reference Table 3 and explain that the smallest uncertainty does not consistently produce the lowest RMSE, reflecting possible nonlinearities or localized sensitivities.

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.

We intended to refer to DA performance in estimating bed topography and thickness in this paragraph. We will rephrase it to incorporate all suggestions raised by the reviewer.

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.

We will revise the figure as suggested. The pattens for RMSE_B and RMSE_H over assimilation time is very similar to each other since surface elevation is defined as the sum of ice thickness and bed topography (surface = thickness + bed). The figures in the second and third lows look very similar but not the same figures (note that values in y axis).

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.

As mentioned above for other figures (e.g., Fig. 5), it is likely due to a temporary mismatch between the model forecast and the observations during this period, potentially caused by transient model dynamics or nonlinearities in the response to assimilated observations. Localized error growth can temporarily degrade performance. We will discuss this point in the revised manuscript.

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.

We will clarify the fast flowing regions in the revised manuscript.

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

We will clarify the fast and slow flowing regions in the revised manuscript.

L322: "more" accurate.

We will revise this as suggested.

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

We meant how the initial ensemble is generated. We will revise this.

L337: relatively small ensemble size: Be precise.

We will specify the ensemble size.

L337-338: previous studies: References are missing.

We will add the references.

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.

We agree that increasing ensemble size does not always guarantee improved DA performance. We will revise the discussion to acknowledge that larger ensemble sizes can improve performance in general but may also introduce challenges that must be carefully managed, particularly in long assimilation periods or highly nonlinear systems.

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

We will remove this sentence, as a rephrased paragraph discussing future studies will be included, as suggested by the reviewer.

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

We will add the references.

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

The differences in the optimal localization radius likely comes from the differences in model configuration, dimensionality, and spatial resolution. Our study uses a 2D unstructured mesh with relatively fine spatial resolution, whereas previous studies using flowline models (1D) with coarser grids may require broader localization to account for longer correlation length scales. We will clarify this in the revised manuscript.

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

We will revised it as suggested.

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).

We specified "up to 100 years". Up to 100 years, the determinist ice volume loss forecast in Fig. 9b shows better agreement with the reference simulation than in Fig. 9a.

L354: XX : Add numbers.

We will add missing values here.

L358: in this study that: Replace with here.

We will revise this as suggested.

L358: observations maintains: Replace with observations while maintaining

We will revise this sentence as suggested.

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

We will revise this as suggested.

L361: Remove "in this study".

We will revise this as suggested.

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

We will revise this as suggested.

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

We will remove these references here.

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

We will address this point as the previous mentioned in our response to the comments on Table 3.

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

We will address this point and related it to the temporal decline in DA performance.

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.

We intended to compare DA performance in estimating bed topography (and ice thickness) versus the friction coefficient. We will add the relevant results and revise the discussion in the revised manuscript.

L377: will or should?

We will use "should" instead of "will".

L379-380: Consolidate all future study suggestions into one paragraph.

We will combine this paragraph with the next one into a single paragraph.

L398-399: 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.

We will consolidate the mentions of future work into a single paragraph at the end of the discussion section to improve clarity.

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

Except for the temporary decreases in DA performance observed during the assimilation period, this statement generally holds true. We will clarify this point in the revised manuscript.

L400: What does great accuracy mean? Be precise.

We agree with the reviewer and will replace "great accuracy" with a quantitative assessment.

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

It is not necessarily smaller uncertainty levels; it also depends on the resolution or the track spacing of the data. We will clarify this in the revised manuscript.

L410: Will you also upload the data files?

Yes, we will upload the data files to the repository.