Estimation of the state and parameters in ice sheet model using an ensemble Kalman filter and Observing System Simulation Experiments
– Authors' response (RC2) –

Youngmin CHOI et al.

May 1, 2025

This is a review of "Estimation of the state and parameters in ice sheet model using an ensemble Kalman filter and Observing System Simulation Experiments" by Choi et al., submitted for publication to The Cryosphere. This manuscript describes the use of an ensemble-based data assimilation system, the Ensemble Kalman Filter (EnKF), to assimilate data into a 2D large-scale ice sheet models, for the purpose of better estimating parameter values and state variables during the historical period. It follows on other studies that have explored similar methods for 1D ice sheet models, and makes the crucial step of applying such methods to a model widely used for projections. This study also adds a novel "Observing System Simulation Experiment" in which different potential observing system configurations (resolution, track spacing, observational accuracy) are tested to determine their ability to improve accuracy in estimated parameters and state.

Overall, I think this is a pretty straightforward study using well-known tools in a new way with ice sheet models, advancing the state of the art in our field. My main suggestions are to further explore certain DA and modeling choices that are unexamined in the current version of the manuscript. I have detailed these suggestions and more minor ones below.

We thank the reviewer for reviewing the manuscript and constructive comments. We will revise the manuscript to include additional justification for key data assimilation and modeling choices, clarify methodological decisions, and expand the discussion on the implications and limitations of our approach. We address each specific comment in detail below and aim to improve the clarity.

1. The manuscript briefly describes what the EnKF is, and then indicates that the EAKF version is chosen for this study. There are multiple different flavors of the EnKF available in DART, so it *is unclear why EAKF is chosen and whether the results would be any different if another filter was chosen. My suggestion is to describe in some more detail what is done in an EnKF and how the EAKF is different from the standard EnKF. Additionally, either some justification for why the EAKF was chosen and some level of justification for why that is the preferred approach when others are available.*

Thank you for the suggestion; this was also raised by another reviewer. We will add more details about the EnKF and EAKF, and clearly describe the distinctions between the two approaches.

2. One thing that is unclear from your study design is the relative importance of assimilation window (e.g. 5 vs 15 vs 30 years) as compared to number of assimilation cycles. You don't change the frequency of observations, which may be sensible given than annual observations are reasonable for current observing platforms. However, it is then hard to understand as a reader whether there is something fundamental about having 20-30 years of observations related to the time scales of ice sheet response to adjustments, or whether it is having 20-30 assimilation cycles to improve. If the observations were more frequent (e.g. an IceSAT2-like 90 days) would it takes less time for the EnKF to improve to the level that you show here?

This is a great point. In this study, we aimed to estimate two constant-in-time parameter fields and the model state on an annual basis. Given the timescales associated with the model state and parameters, as well as the capabilities of current observational platforms, we chose to use annual observations for simplicity.

We agree that the distinction between the length of the assimilation window and the number of assimilation cycles needs further investigation. To address this, we will conduct additional experiments to explore the relative impact of the number of assimilation cycles versus the time period over which they are applied. We will include these results and a discussion in the revised manuscript.

3. A big difference between your perfect model design and a real scenario where DA might be applied is that only two constant-in-time parameter fields are unknown. In reality, (e.g.) ice viscosity and climate forcing are also likely to be poorly known (though at least climate forcing is directly observable), and climate forcing (and basal friction) may vary in time. Two possibilities that would be helpful to run some experiments to assess are:

(a) if you are mistaken about the values of other parameters, but still only estimate basal friction and topography, will the estimate of basal friction compensate for these other errors (particularly for ice viscosity which trades off quite directly with basal friction in a depth-integrated model) - there is some evidence of such compensation already happening in your estimates, see below

(b) could this DART-ISSM configuration be used to estimate multiple parameters at once (I don't see a reason why not, but the performance may not be the same as what is found for the single parameter estimation experiments explored currently).

I get that the design of these experiments are meant to mimic and compare directly to Gillet-Chaulet 2020, but it would be useful to also push beyond their design to get closer to a realistic case where DA might be used.

Thank you for this insightful comment. We agree that in real-world applications, other parameters such as ice viscosity and climate forcing are also poorly constrained and may vary in space and time. In this study, we intentionally limited the number of unknowns to two constant-in-time parameter fields (basal friction and bed topography) to enable direct comparison with Gillet-Chaulet et al. (2020) and to isolate the behavior of the EnKF framework in a controlled setting.

Regarding (a), we recognize that compensation effects between parameters (e.g., between basal friction and ice viscosity) may arise if other sources of uncertainty are not properly accounted for. We will add a discussion of this limitation and clarify that our experimental design assumes perfect knowledge of other parameters such as ice viscosity, which is not realistic. We also plan to explore the effect of misspecified background parameters (e.g., slightly incorrect viscosity fields) in future sensitivity experiments to assess the potential for parameter compensation.

Regarding (b), we believe that the DART–ISSM framework is capable of estimating multiple parameters simultaneously, and extending the current setup to include more parameters (e.g., ice viscosity, time-varying forcing) is a valuable direction for future work. We will clarify this in the revised manuscript and note that this study provides a foundational step toward that goal by demonstrating the feasibility and utility of ensemble DA with a simplified setup.

4. At more than one point it is suggested that using variational methods is more computationally intensive that ensemble-based DA. However, there is no real direct proof of this as you don't perform a direct comparison and to my knowledge this has not been done in the published literature. Given that ISSM has a variational DA option already implemented, it could be valuable to compare EnKF with ISSM to EnKF with ISSM in terms of core-hours for a simple standardized run. Short of that, it would be useful to have a sense for the DART overhead? If it is negligible then I would expect ensemble-based DA to have n times the computational expense of a conventional ISSM run where n is the number of ensemble members. Additionally, giving a sense for how this ensemble-based approach has (or can be) parallelized would be useful. In theory, ensemble DA is highly amenable to parallelization, but this depends on how covariance matrices are constructed and how shared memory parallelism is handled. More details on all the computational aspects of this new method would be very useful to include.

Thank you for the great suggestion. We agree that a computational comparison between the two data assimilation methods would be valuable. However, due to the fundamental differences in the core computational processes of variational and ensemble based approaches, a direct comparison of their computational costs is challenging and may be beyond the scope of this study.

Variational approaches using automatic differentiation (AD) tend to have higher memory demand, while ensemble DA methods primarily increase computational cost through the need to run multiple forward simulations. Additionally, the two approaches are implemented using different tools within ISSM – variational DA is built using the AD tool, while ensemble DA is integrated via DART – which further complicates direct comparison. We will include this limitation in the revised manuscript and mention the need for future work to systematically assess the computational trade-offs between these two methods.

It is also worth noting that DART supports parallel computing. We will also include this in the revised manuscript.

L16: less numerical model re-development

We will revise it as suggested.

L26: use a form of variational

We will revise it as suggested.

L28: realy on observational at a single time to

We will revise it as suggested.

L30: it often introduces nonphysical artifacts into the

We will revise it as suggested.

L35: The use of computational techniques such as automatic differentiation in ice sheet models

We will revise it as suggested.

L69: to my knowledge this is the first ice sheet modeling paper to apply OSSE, so I think you can be more direct about this sentence

We will revise this sentence as suggested.

L86: an ensemble...for ice sheet model initialization

We will revise it as suggested.

L88: on model initialization

We will revise it as suggested.

L92: simulation of ice sheets

We will revise it as suggested.

L100: explain what the random midpoint displacement method is

We will add the details with a reference.

L138: model simulations

We will revise it as suggested.

L143: I am confused here because you don't include velocity in the state vector, but later you say it is part of what is assimilated?

Velocity is an observation being assimilated, not the part of the state vector. We will clarify this in the text.

L184: does localization as implemented preserve covariance between different variables at the same location in space or does it simply localize along the diagonal of the covariance matrix? For example, there should be strong correlation between the ice thickness estimate and the bed topography estimate, and so you would be losing a significant amount of your ability to assimilate if covariances between these two variables at the same location in space where zeroed out by localization.

The localization is applied to the full ensemble covariance matrix, not just along the diagonal. Therefore, cross-variable correlations at a given location are preserved. We will clarify this point in the revised manuscript.

L205: what would happen if you had no velocity observations? How much of the performance is due to velocity observations vs thickness?

In our current experimental design, we include velocity observations based on the assumption that high quality annual velocity data are available for most glacier regions. These velocity observations provide strong constraints on basal friction, particularly in fast-flowing regions. If velocity observations were removed, we expect the performance of the data assimilation – especially the estimation of the friction coefficient – to degrade, as surface elevation alone provides weaker sensitivity to basal friction. However, the estimation of bed topography and ice thickness may still benefit from surface elevation observations.

We acknowledge that this trade-off between observation types is important for real-world applications. As similar points are raised by another reviewer, we will conduct additional experiments with varying uncertainties in the velocity data and include the new results in the updated version. Additionally, we will emphasize the need for future sensitivity studies to explicitly isolate and quantify the relative contributions of velocity and elevation observations to overall DA performance.

L243: I think this should refer to Fig. 5

Yes. We will change it to Fig. 5.

L259: mean to initialization the deterministics...full ensemble to initialization the ensemble

We will revise it as suggested.

Fig 4. In the caption you mention that highly localized experiments diverge. It would be helpful to speak to why these experiments diverge in the main text.

We will revise the main text to clarify that 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.

Fig. 7: There are artifacts in the bed topography and ice thickness estimates that correspond to the basal friction estimate. Can you speak to this? Is it related to how the localization is performed?

The artifacts in the bed topography and 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.

L276: can you quantify this change in spread? by eye it doesn't seem to change much between 20 and 30 years of assimilation

We will add the values for each spread in the revised manuscript.

Figure 9: can you add a legend and plot the ensemble mean as well?

We will add a legend and include the ensemble mean in the plot.

L290: It would help to discuss what this sentence means in practice. Is prediction accuracy degraded for this case? Or can you achieve similar results with different localization and inflation parameters?

We will compare these results with our reference experiments and clarify the differences in the text.

L325: initial estimates for the model parameters?

We will revise it as suggested.

L328: correlation between both parameters

We meant the “establish spatial correlation for each parameter”. We will revise the text.

L332: what do you mean by “initial ensembles”

We meant “initial model ensembles”. We will revise the text accordingly for clarity.

L354: need to fill in values for XX

We will add values.

L358: transient changes in model state but not in parameters

We will revise it as suggested.

L366: can you speak to the limitations on this? would having 100m resolution data be even better or proportionally so?

In principle, higher-resolution data (e.g., 100 m) could further improve data assimilation performance by providing finer spatial detail on surface features and more precise constraints on model parameters. However, the benefit of finer resolution may decrease beyond a certain threshold due to increased observational noise, modeling uncertainties, and the inherent spatial correlation scale of the parameters being estimated. We will clarify this in the revised manuscript and discuss the potential limitations of very high resolution observations in the context of our OSSE framework.

L370: this is a very important point that is worth highlighting in the abstract

We will include this finding in the abstract.