Reviewer 2

Review Felikson et al. (2022)

This study aims to constrain and improve projections of the Greenland ice sheet evolution by making use of changes in three different observed quantities (mass changes, velocity changes and dynamic thickness changes) that are a measure the Greenland ice sheet dynamics. So far, Bayesian calibration of Greenland ice sheet model projections have been based on a single observation type.

I think the study presents a very useful framework to improve ice sheet model projections based on observations. However, as also stated in the study, using three different observation types do not give a clear answer on what is the best observation type to constrain the ice sheet model behavior and there might be a better metric when different observation types are combined. Therefore I am wondering whether the study as presented in the present form is convincing enough that model projections have been improved by using Bayesian calibration using different observation types. In addition, I would like to propose some minor clarifications on the methodological choices that you can read below.

Author response: We greatly appreciate the reviewer's comments, and we agree that more work remains to be done to incorporate information from several observation types and, ultimately, improve the methodology of Bayesian calibration for ice sheet projections. In this present study, our goals are (1) to show that Bayesian calibration using any of the observation types reduces uncertainty in the projections and (2) to analyze the discrepancies in the resulting posterior distributions from calibrations that use different observation types. We show that model projections of ice sheet contribution to sea level are improved because the spread in the prior distribution of projections is reduced by using any of the observation types for calibration. Thus, using any one of the observation types improves the uncertainty (spread) in sea-level change projections. However, we also show that care must be taken when using the maximum a posteriori estimate of sea-level change or the cumulative probability of higher-end sea-level change because the specific choice of observation type may bias these. We leave the reconciliation between the three observation types to future studies, as there is a significant amount of analysis to be done to investigate how to combine the information from all three types, as well as to investigate the use of derived metrics (e.g., spatial extent of thinning) and other sources of observations (e.g., outlet glacier terminus positions). For these reasons, we feel that our current paper is convincing and leaves room for future studies to build on our results. We appreciate all feedback and we have addressed all comments below.

Main comments

You explain that you coarsen the resolution in order to remove the spatial correlation between the different observations. However, if you look at Figure 1, there is still a strong correlation between the observed velocity change and the dynamic thickness change, especially for the fast outlet glaciers. It looks like it is a workaround to justify the neglection of a covariance matrix. I am also wondering if you are not losing too much important, detailed information by coarsening the resolution only to justify that the different variables are not correlated. Especially velocity changes at 50 km resolution remove a lot of detail.

Author response: As noted by the reviewer, our approach for coarsening the observations allows us to assume that the residuals are uncorrelated. The method that we use is consistent with previous studies that have used Bayesian calibration for ice sheet ensembles (e.g., Edwards et al., 2014; Ruckert et al., 2017; Nias et al., 2019; Brinkerhoff et al., 2021). In this way, our paper provides an evaluation of existing methods, which have been used in the literature. Using a covariance matrix would, in fact, allow more information and detail in the observations be used in the calibration. This would be a good direction for future work, but we leave it out of the scope of our paper in order to focus on evaluating calibration methods that have been used in past work.

Reviewer 2

Did you investigate the posterior probability distribution for higher spatial resolutions?

Author response: Early on in our analysis, we experimented with higher spatial resolutions for averaging the dynamic thickness change and velocity change. However, we ultimately chose to use a 50 km grid to alleviate the complication of properly accounting for spatial correlation in the observational uncertainties. As noted above, this could be improved upon but it is consistent with previous literature.

The model is initialized to the year 2007 and observations are used until 2015. What is the influence of the length of the observation time on the results? Could it be useful to use a longer observational time series, for instance by initializing the model to the year 2000 to increase the observational length?

Author response: The impact of the calibration time span would indeed be useful to investigate. However, our method for initializing the model, described in Nias et al. (2023), prevents us from initializing before 2007 because this is the earliest year for which spatially complete surface elevation and velocity datasets are available. Our simulations are initialized using a data assimilation method to invert for the basal friction coefficient and this requires an ice-sheet-wide dataset for ice sheet surface elevation. The 2007 surface elevation mosaic (Howat et al., 2014) is the highest quality, spatially complete, observation-based map of Greenland Ice Sheet surface topography that is dated furthest back in time. An older datasets of Greenland surface topography exists (Bamber et al. (2001) dataset that was used in the SeaRISE ice modeling project (Bindschadler et al., 2013), and was a tremendous asset in ice-sheet-wide modeling at the time. However, this dataset is based on radar altimetry, which has high uncertainty around the ice sheet margin. The aerial photogrammetry digital elevation model (Korsgaard et al., 2016) provides surface topography of the margin with low uncertainty but is not spatially complete, lacking elevation observations in the interior of the ice sheet. There is on-going work to compile a surface elevation dataset further back in time and, once these new datasets become available, it would be valuable to investigate the influence of the length of the observational timespan on the calibration. We discuss this briefly on line 324 and we leave this for future work.

You refer in the methodology to the Nias et al. paper in review, but there is no preprint available (or no reference in the reference list) so it is impossible to acquire the right information. For instance, the model ensemble details can be found in Nias et al. (in review). It would be interesting to have them in this manuscript as well. Also, you mention on L153 that the modeled mass change is aggregated within the same drainage basin as used for aggregating the observed mass change. Can you explain what that means, why and how that is done?

Author response: Nias et al. (in review) was not publicly available as a preprint at the time of our submission and we were instructed by the publisher to provide it to the journal editor, who could distribute it to reviewers upon request. Since the time of our submission, that paper now been published as Nias et al. (2023), providing complete details about the model setup and experimental design for the ensemble. We provide a summary of the ensemble in Section 2.1 and we will update the Nias et al. (in review) reference to Nias et al. (2023).

We will provide some additional explanation for the motivation for aggregating mass change within drainage basins and, as suggested by Reviewer 2, we will add the drainage basin outlines to Figure 1.

The goal is to narrow the committed Greenland ice sheet change projections. However, you are changing not only model parameters but also the SMB, which you define as a forcing. Can you still speak of committed sea-level change if you change the SMB forcing?

Author response: We vary SMB forcing as part of the ensemble in order to account for uncertainty within the present-day SMB over the ice sheet. This way, our projections represent Greenland's committed sea-level contribution, given the uncertainty within the estimated present-day SMB, as well as the uncertainties within model parameters. This is explained in greater detail in Nias et al. (2023).

It would be nice to add some figures about the modelled Greenland ice sheet state. You show only 4 figures that basically focus only on the posterior probability distribution of committed GMSL rise due to melting of the Greenland ice sheet.

Author response: This is a very good suggestion, and we will add figures to show the modeled ice sheet state with panels to show velocity, thickness, and mass change for the heightest and lowest weighted simulations, similar to Figure 1 in Nias et al. (2023).

Minor comments

L61: It could be useful to give the values for the basal friction factor, the ice viscosity offset and the SMB offset for the 137 forward runs in the supplementary information. Could you add a justification for the offsets chosen?

Author response: The values for the ranges of these offsets, as well as a thorough justification for those chosen offsets, is given in Nias et al. (2023), including a supplementary file with the exact values for each ensemble member.

L151: The modeled quantities are regridded to 50 km x50 km. What is the original resolution of the model results?

Author response: We will add the original resolutions of each observation type, as well as the model results, to the appropriate subsections in Section 2.

L173-L175: The model uncertainty is a linear function of the observational uncertainty and you assign different multiplication factors for the different calibration to match the peak in the posterior distribution. Could you elaborate a bit more on the choice of these values? Because it looks contradictory to what you say on L185, that the median and maximum a posteriori GMSL are far apart.

Author response: This is an excellent question, and we will make this clearer in the text. We select the multiplication factor, *k*, for each observation type such that the value for the normalized probability of the peak in the posterior (i.e., the y-axis value) is approximately equal for all three calibrations. We did not make it clear in the text that it is the probability, and not the GMSL value at the peak probability, that we are aligning by varying the multiplication factor, *k*.

L207: That sounds like a logical consequence of the larger changes along the margin of the ice sheet.

Author response: This is a fair point, and we will note in the text that this is expected because we see larger observed changes around the margin.

L252: You discuss the firn thickness change as a potential bias. What is the modelled firn thickness change? Could you show a figure? Also you report a potential bias of 10 cm per year in the ice sheet interior. What is this number based on?

Author response: As suggested by the reviewer, we will add a supplementary figure to show thickness changes due to firn densification.

Reviewer 2

The 10 cm/yr bias was meant to represent a worst-case potential estimate of a potential bias in the firn thickness change. Our intention was to quantify the impact of such a bias on the calibration results. However, based on comments from both reviewers, we are going to remove this paragraph from the manuscript, as well as the associated plot, for two reasons:

1. A more thorough analysis should be done to investigate the impact of potential biases in the dynamic thickness change observations on the calibration, including from sources other than firn modeling.
2. Reviewer 2 commented: "I get the impression that you do not believe in the thickness change calibration because it is very dependent on the firn thickness change that is hard to constrain." While potential errors in the firn thickness change estimates are source of error in the dynamic ice thickness change observations, there are other sources of error that contribute. By focusing only on the impact of a potential bias in firn densification, we are unfairly presenting the importance of this source of error.

Typo's
L124: and and

Author response: This typo will be fixed.

L274: Open questions?

Author response: This typo will be fixed.

L284: correct 'to on the choice'

Author response: This typo will be fixed.

Figures
Figure 1: Something went wrong with the labels: please add/modify the labels (a), (b), (c), (d), (e) and (f). Please also adapt the colormap labels to make them uniform with the indication of the variable and the units between brackets. I would also add the resolution for the different gridded observations.

Author response: Thank you for noting these. We will make all the suggested improvements to this figure.

Figure 3: I do not see the blue curve with the Gaussian approximation.

Author response: We will add the blue curve – this was left out of the figure accidentally.

References:
Bindschadler, R. A., Nowicki, S., Abe-Ouchi, A., Aschwanden, A., Choi, H., Fastook, J., ... & Wang, W. L. (2013). Ice-sheet model sensitivities to environmental forcing and their use in projecting future sea level (the SeaRISE project). *Journal of Glaciology*, *59*(214), 195-224.

Reviewer 2

Bamber, J. L., Ekholm, S., & Krabill, W. B. (2001). A new, high-resolution digital elevation model of Greenland fully validated with airborne laser altimeter data. *Journal of Geophysical Research: Solid Earth*, *106*(B4), 6733-6745.

Brinkerhoff D, Aschwanden A, Fahnestock M (2021). Constraining subglacial processes from surface velocity observations using surrogate-based Bayesian inference. Journal of Glaciology 67(263), 385–403. https://doi.org/10.1017/jog.2020.112

Edwards, T. L.; Fettweis, X.; Gagliardini, O.; Gillet-Chaulet, F.; Goelzer, H.; Gregory, J. M.; Hoffman, M.; Huybrechts, P.; Payne, A. J.; Perego, M.; Price, S.; Quiquet, A. and Ritz, C. (2014). Probabilistic parameterisation of the surface mass balance–elevation feedback in regional climate model simulations of the Greenland ice sheet. The Cryosphere, 8(1) pp. 181–194.

Howat, I. M., Negrete, A., & Smith, B. E. (2014). The Greenland Ice Mapping Project (GIMP) land classification and surface elevation data sets. *The Cryosphere*, *8*(4), 1509-1518.

Korsgaard, N. J., Nuth, C., Khan, S. A., Kjeldsen, K. K., Bjørk, A. A., Schomacker, A., & Kjær, K. H. (2016). Digital elevation model and orthophotographs of Greenland based on aerial photographs from 1978–1987. *Scientific Data*, *3*(1), 1-15.

Nias, I. J., Cornford, S. L., Edwards, T. L., Gourmelen, N., & Payne, A. J. (2019). Assessing uncertainty in the dynamical ice response to ocean warming in the Amundsen Sea Embayment, West Antarctica. *Geophysical Research Letters*, 46(20), 11253-11260.

Nias, I. J., Nowicki, S., Felikson, D., & Loomis, B. (2023). Modeling the Greenland Ice Sheet's Committed Contribution to Sea Level During the 21st Century. *Journal of Geophysical Research: Earth Surface*, *128*(2), e2022JF006914.

Ruckert, K. L., Shaffer, G., Pollard, D., Guan, Y., Wong, T. E., Forest, C. E., & Keller, K. (2017). Assessing the impact of retreat mechanisms in a simple Antarctic ice sheet model using Bayesian calibration. *PLoS One*, *12*(1), e0170052.