Review of Reed et al. "Calibration of a coupled ice sheet-ocean model using observations of ice dynamics and basal melt in West Antarctica"

Reviewer: Xylar Asay-Davis

I wish my name to be relayed to the authors, as I feel I am always a better reviewer when I am not anonymous and I encourage others to consider reviewing non-anonymously whenever they feel able.

General Comments:

This manuscript presents a new calibration strategy for a coupled ice sheet–ocean model of the Amundsen Sea sector, using spatial observations of basal melt rates together with observed changes in ice speed and thickness over 2013–2017. The authors' stated aims are to produce the first historically calibrated coupled ice sheet–ocean model using both oceanic and glaciological observations, to provide evidence for the importance of accurately simulating basal melt near grounding lines, and to assess the implications of this calibration for projections through 2100. To do so, they introduce a modified melt parameterization, run an ensemble of 68 coupled hindcast simulations spanning a range of melt-parameter values, and compare a preferred "transient-coupled" calibration against a more standard "static-melt" approach. Their main findings are that the highest-scoring hindcast simulations are those that enhance melt near deep grounding lines, that including ice-dynamical observations leads to a much better reproduction of observed thinning and speed changes over the hindcast period, and that the preferred transient-coupled calibration yields substantially larger projected sea-level contribution by 2100 than the static-melt calibration, including a 14 mm larger contribution under RCP8.5.

The manuscript has a number of important strengths. First, it makes a forceful and, to my mind, persuasive case that observed ice-sheet thickness and velocity changes should be treated as key constraints on basal-melt calibration, rather than relying on melt-rate observations alone. The paper is especially effective in showing that calibrations based only on present-day melt estimates can miss the observed dynamical response of the ice sheet, whereas including ice-dynamical constraints materially improves agreement with the historical thinning and acceleration patterns. More broadly, I think the authors are right to emphasize that this type of multi-observational calibration remains underused in the wider community, despite the fact that coupled systems are meant to reproduce both the ocean forcing and the ice-sheet response to that forcing.

Second, the manuscript makes a strong case that the combination of an improved melt parameterization and a transient calibration strategy can greatly increase the model's ability to reproduce observations over a short hindcast relative to more standard techniques. The comparison with the static-melt/default-parameterization approach is compelling in showing that the standard setup tends to favor lower grounding-line melt, leading to poorer reproduction of the observed dynamical changes and to coupling shock. Relatedly, I was impressed by the rigor of the parameter-space exploration: running 68 coupled hindcast simulations for a model of this complexity is a substantial undertaking, and it gives the calibration exercise much more credibility than a lightly sampled tuning exercise would have. I also think the paper makes an important practical contribution by showing that, at least in the Amundsen Sea sector, a roughly 5-year calibration window is already long enough to extract meaningful constraints from observed thickness and velocity change. That is an important result in its own right, because it suggests that transient coupled calibrations of this sort may be computationally demanding but still tractable for broader use.

Finally, I found the paper to be very well written and well organized. The structure is clear, the motivation is easy to follow, and the figures and experiment design do a good job of guiding the reader through what is otherwise a fairly complex modeling framework. I also appreciated the discussion of limitations and future work. In particular, the manuscript does not present the new approach as an endpoint, but instead lays out several intriguing next steps: developing a more physically grounded treatment of grounding-line melt processes, moving toward joint calibration of ocean and ice-sheet parameters, testing alternative hindcast windows as new observations become available, and extending the calibration framework to additional forcings and regions. That forward-looking discussion makes the paper useful not only as a study in its own right, but also as a contribution that could help shape future work in this area more broadly.

My main concern is that the manuscript's central interpretation is not fully supported by the comparisons currently shown. The paper repeatedly emphasizes the importance of transient calibration of the coupled ice-sheet–ocean system using observations that constrain both components, but the main comparison between the preferred "transient-coupled" approach and the "standard" approach changes more than just the calibration strategy. As defined in the manuscript, the transient-coupled configuration combines three elements at once: a transient geometry, multiple observational constraints, and the new modified melt parameterization. By contrast, the static-melt comparison uses static geometry, melt observations alone, and the default melt parameterization. Because these factors are changed simultaneously, the relative importance of transient coupled calibration versus the new melt parameterization is not isolated by the results currently shown. This matters because the manuscript also shows that the highest-scoring hindcast simulations all use the modified parameterization, whereas the default-parameterization cases are among the poorest performers, and the low-scoring cases are explicitly linked to insufficient melt near the grounding line and the resulting coupling shock. In other words, the evidence presented seems at least equally consistent with the interpretation that much of the improvement arises from the added flexibility of the new parameterization, especially its ability to enhance grounding-line melt, rather than from the transient coupled calibration alone.

For that reason, I think the paper would be substantially stronger if the authors added simulations that explicitly disentangle these effects. Most importantly, I would like to see a calibration using the new melt parameterization but retaining the more standard setup of static ice-shelf geometry and melt-rate-only constraints. That experiment would directly test whether the modified parameterization alone can recover much of the apparent improvement now attributed to the transient coupled calibration. Relatedly, the manuscript already shows that when the calibration is restricted to melt observations alone but transient geometry is retained, the preferred solutions shift within the family of modified-parameterization cases, and Fig. B3 appears to suggest notably better performance for those runs than for the default-parameterization cases. That makes the missing control case especially important, because it raises the real possibility that a substantial fraction of the skill improvement could also be obtained without the transient coupled calibration, provided the new parameterization is used. I am typically quite hesitant to recommend new simulations as part of a review: You designed the experiments, not me, after all. But, if these additional simulations are not feasible, then I think the authors must at minimum soften the repeated claims about the specific importance of transient coupled calibration and more clearly state that the present study demonstrates the benefit of a bundled methodological package, not of transient calibration in isolation. This may substantially change the emphasis and character of the paper.

A separate concern is that the manuscript may understate the extent to which the calibrated melt parameters could be compensating for structural deficiencies in the ice-sheet model, rather than uniquely identifying the physically correct basal melt field. The paper notes that damage and calving are not represented and also acknowledges that only the ocean-side melt parameters are calibrated here, while a single set of ice-sheet parameters is held fixed. The limitations section further states that, in a fully calibrated ice-sheet model, calibrating against either ice-speed change or thickness change should converge on the same optimal basal-melt parameters, whereas in the present results these metrics favor different optima precisely because the ice-sheet model parameters have not themselves been historically calibrated. To me, this is an important caveat that extends beyond the limitations section. If missing or mis-specified ice-sheet physics affect the modeled dynamical response, then the melt calibration may partly be absorbing those errors. That possibility seems especially relevant here because the preferred parameter choices are those that increase melt near grounding lines, where the coupled system is most sensitive and where omitted processes such as damage, calving, or other ice-dynamical biases could plausibly matter. I therefore encourage the authors to discuss more explicitly throughout the manuscript that the optimized basal-melt field may in part be compensating for biases elsewhere in the coupled system, and that a future joint calibration of both ocean and ice-sheet parameters is needed before the inferred melt parameters can be interpreted as uniquely constrained.

As a result of these concerns, particularly about better acknowledging the fact that the current simulations do not untangle the relative importance of the new melt parameterization compared with the transient calibration, I feel I must recommend major revisions. My more specific comments are detailed below, with a few formatting suggestions under “technical corrections.”

Specific Comments:

l. 122-126: Thanks for detailing these changes! These problems in BedMachine (not fixed in v4 to my knowledge) make the dataset hard for coupled ice sheet-ocean modeling. It’s really helpful that you point out the problems and give other groups a sense of how the can be mitigated.

l. 190 Eq. (1): Could you say what Γ_{Turb} and $\Gamma^{T,S}_{Mole}$ are? Also, where is this choice of notation from? McPhee et al. (1987) uses Φ rather than Γ as far as I can tell and I am used to $\Gamma_{T,S}$ being a nondimensional heat or salt transfer coefficient in the numerator, rather than the denominator so that’s leading to a bit of confusion with Γ terms in the denominator here for me.

l. 225: “As opposed to the default melt parameterization, with C_d as its only tunable parameter…” I wonder why Γ_{Turb} is considered a fixed parameter and C_d is tunable. I think it is not unusual in ice-shelf-cavity modeling to consider both to be tunable.

Sec. 2.4.2: This is a section where I feel you could go further in acknowledging that the current approach runs the risk of tuning melt parameters to compensate for ice-sheet errors, e.g. inaccurate initialization, lack of calving or damage.

l. 330: “To avoid this, we adopt a spatially-uniform error of 5 m yr^{-1} for the melt dataset.” Could you provide a justification for choosing this value?

l. 387-389: “Some of this bias is likely attributable to ice damage and calving processes (Lhermitte et al., 2020; Joughin et al., 2021), which are not represented in the ice-sheet model.” This seems like a good opportunity to acknowledge that you are calibrating the melt model to help compensate for this missing process, possibly at the expense of more accurately representing the basal melt field itself.

l. 407-411: “When using only melt-rate observations in the calibration (while maintaining a transient ice-shelf geometry), the highest-scoring simulations shift towards those with velocity-independent melt over a wider region but lower γ_{T0} values (i.e., reduced turbulent exchange velocities and therefore lower melt rates) (Fig. B3). However, these same simulations are not the highest scoring when evaluated against changes in ice speed (Fig. B2).” I feel like this deserves more discussion. Here is the evidence that the new melt parameterization is lifting a lot of the weight compared with the “standard” calibration approach. It also feels like more should be said about why B2 and B3 may differ, and the likelihood that improvement in B2 come partly for the “right reasons” (better ice sheet response directly to melt) and partly for the “wrong reasons” (e.g. basal melting compensating for lack of damage and cavling or biases in ice-sheet thickness).

l 416-418: “Nevertheless, because this represents the top-ranked parameter set following the most common calibration approach in the literature, we extend this simulation to 2100 and compare it with the transient-coupled calibration in Sect. 4.” Again, how do you then know how much transient-coupled calibration really changes things, as opposed to the new parameterization?

l. 429-231: “Typically, tuning is done using depth-binned or spatially-integrated values of basal melt for a static ice-sheet geometry. This approach has two key limitations. First, spatial-integration does not provide information about the distribution of melt, which we show matters for reproducing ice dynamics.” I am not aware of depth-binning being a common practice in calibrating 3-equation parameterizations in ocean models with ice-shelf cavities. While some such models may use spatially averaged melt rates for calibration, I am also not aware of that being a standard method of calibration and it is not what is done in my own group – we compare the full melt field with an “observed” melt climatology wherever both have valid data. I am more aware of these types of metrics being used for analysis rather than calibration, and in parameterizations of basal melting in standalone ice-sheet models. I would suggest taking out or rewording your first point here.

l. 434-435: “This is evident in our low-scoring coupled model simulations, where low grounding-line melt rates cause upstream thickening and slowdown, in contrast to observations (Figs. 4–5).” I know I keep harping on the same points but this feels like yet another place to acknowledge that melt tuning may be compensating ice sheet model biases.

l. 436-437: “The modified melt-rate parameterization compensates for low vertical and horizontal resolution in the ocean model by generating enhanced melt near the grounding line…” Given the ad hoc nature of the parameterization, it might be best to be broader about what it might be compensating for in the ocean model. My feeling is that the boundary-layer physics under ice shelves is pretty far from being a solved problem or one what a single accepted parameterization at basically any practical resolution that ocean models with ice-shelf cavities can achieve. So the new parameterization is likely compensating for a variety of factors including low horizontal and vertical resolution as well as missing physical processes.

l. 450-458: I really appreciate this paragraph! It really drives home the value in the methods you are proposing here. Intuitively, I tend to feel like ice sheet models respond slowly to ocean forcing so it’s valuable to have it made starkly clear how clear changes in ice sheet thickness and velocity can be on these short time scales.

Fig. 6: I find the contents of the insets hard to see without zooming in an unreasonable amount of. Would it make sense to make this into two figures, or break the insets out into their own panels?

Sec. 5: To me, this is a really important and valuable section. You are showing with humility a lot of the important caveats of this work and where it can be taken in the future. You show that it has been far from a trivial exercise but it’s not outside the realm of possibility for other groups to do something similar. You point out the very important next step of calibrating the melt parameterization and the ice sheet together!

l. 629-630: “Importantly, by calibrating the ice-sheet model using ice-dynamics observations, the calibration of the ocean melt parameters would only require one type of these datasets along with melt observations.” I didn’t follow this last bit. Why would basal melt now only be calibrated with “one type of these datasets”? Do you mean it would be calibrated only with ice-sheet velocity or only with ice thickness changes? If so, why?

Sec. 5.3: This section is a useful acknowledgement of the limited scope of the calibration. It is not really phrased as a limitation (or in the context of future work) so it was not clear to me if it belongs in Sec. 5 or could be moved elsewhere.

l. 654-357: “Despite the limitations, our transient-coupled calibration approach represents a significant advance in capturing the complex interactions between ice dynamics and ocean melting. Our results demonstrate that calibrations using present-day melt estimates alone systematically underestimate ice-sheet response over the historical period. By incorporating ice dynamics constraints, this bias is removed entirely, resulting in more robust projections of future mass loss.” This paragraph brings up many of the concerns I have mentioned in my general comments. I do not believe your simulation results show that transient-coupled calibration is the significant advance here, as opposed to a new melt parameterization. Your results "demonstrate that calibrations using [only] present-day melt estimates..." **and MITgcm's existing parameterization of sub-ice-shelf melt** "...systematically underestimate ice-sheet response over the historical period." But you haven't shown that transient-coupled calibration is the reason that you get better results when you use transient-coupled calibration *and* a new parameterization.

l. 660-663: "Whereas previous calibrations have relied on melt estimates only and assumed static ice-shelf geometries, we have shown that including feedbacks between basal melting, evolving ice thickness, and cavity circulation in the calibration leads to a greatly improved representation of ice-sheet dynamics over the observational period."

Sec. 6 Conclusion: I found the conclusion more measured than some of the broader framing in the manuscript. In particular, the conclusion does not appear to claim that transient-coupled calibration alone is responsible for the improved reproduction of observed ice-sheet change; rather, it presents the results as demonstrating the value of a multi-observational, transient coupled approach using the modified parameterization. However, the conclusion would still benefit from a more explicit acknowledgment that some of the inferred melt-parameter changes may be compensating for unresolved or uncalibrated ice-sheet processes, a possibility that is recognized in the limitations section but not carried through to the final synthesis.

Technical Corrections:

Fig. 2: I found the pink transition zone and the bold black line for its center to make it very hard for me to compare the melt rates in panels e) and f) with those in a). My preference would be to put that transition zone in its own panel. It seems important to be able to visually see how abrupt or smooth the transition in melting is across the transition zone without the pink shading and black line obscuring it. But an alternative that still might help would be to use another color (green?) that is clearly not so close to the colormap for the transition zone.