

Response to review by anonymous referee #2

Review comments in italic, responses in regular font.

This is a manuscript with a very important message, well stated in the title, which I wholeheartedly agree with. Using a large ensemble of model simulations with a variety of parameter and forcing choices, it presents evidence that calibrating ice sheet models to reproduce snapshots of a present-day state is not sufficient for accurate predictions of future states. It will be a valuable paper, and certainly should be published.

The manuscript is thorough, with results well presented and the argument clearly made. I found myself agreeing with many well-made points throughout, and I don't have any major recommendations for changes to the current content. I would however be interested to see a little more discussion on a couple of points:

We are happy with the reviewer's enthusiasm.

1. *The assessment of model quality by matching a snapshot of data is the idea being argued against. However, with reliable observations over Antarctica being available for the past two to three decades, what are the authors' thoughts on using a match to the observational trend over a longer time period as a metric? I think this may have been one idea under consideration for the ISMIP7 analysis, using the historical runs required from models in that protocol.*

This is an interesting question, in our view without a conclusive answer so far. We will add the following lines to the Discussion section:

“There have also been suggestions to judge models based on how well they reproduce the observed mass loss trend, either in terms of the spatial pattern (along the lines of van den Akker et al., 2025), or in terms of in the integrated mass loss, i.e., the contribution to sea-level rise (along the lines of Otosaka et al., 2023). However, when compared to the existing uncertainties in ice-sheet models, the present-day mass loss trends are small. Extrapolating Antarctica's present-day sea-level contribution of about 0.4 mm/yr produces just over 10 cm of sea-level rise in 2300, a signal that vanishes beneath the uncertainty from the applied climate forcing and other uncertainties. Similarly, van den Akker et al. (2025) find that the acceleration away from the present-day trend, and the resulting committed collapse of the Amundsen Sea basin, in most of their simulations only initiates after 2300. These findings suggest that the bias that is introduced by not including the present-day trend, is obscured by the many other existing uncertainties in ice-sheet models.”

While most inversion processes are still based on single snapshots of data, this is quite widely recognised as an issue (although never directly addressed in this way

before, to my knowledge). In some cases (e.g. Goldberg & Holland, 2022), efforts are being made to introduce transient calibration, in which observational data from multiple timestamps is considered in the inversion process. This is seen to produce larger contributions to sea level in century-scale simulations. Related to the above point, do the authors think that matching multiple snapshots over time in this way could lead to an improved ability to judge the ‘quality’ of model setups from their initialisation, if the method were more widespread?

Yes, definitely. It is fundamentally impossible to derive both the sliding law and the bed roughness simultaneously from a snapshot observation. However, when time-varying observations are used, then it becomes possible to get a much clearer picture of the sliding process, and thereby improve the models, as e.g. in the work by Trevers et al. (2024) which we reference in the Discussion section. The work by Goldberg and Holland is indeed another good example, which we will also reference. Of course, this still requires a similarly detailed knowledge of other process affecting ice dynamics (basal hydrology, damage, and calving probably being the most important on those time scales), but it is definitely a step forward.

I have a few specific comments below.

Line 74-6: A very good point; I’ve always thought initialising to a steady state is bound to introduce errors, since the real system is not steady.

Line 105: I wouldn’t class 5km as being a particularly high resolution for accurately representing grounding line movement, although this perhaps depends on the scale of the problem. Specifying that this is for whole-of-Antarctica simulations would clarify.

We will clarify that we use 5 km for the grounding line in all Antarctica, but finer values for Thwaites and Pine Island (down to 1 km in the highest-resolution simulations).

Line 203: A brief explanation of ABUMIP is needed, as not all readers will immediately know what it is. My suggestion would be to omit the reference to it here and simply say “The experiments presented in Appendix A...”, then give more detail in the appendix.

We will move this explanation to the appendix.

Line 227-30: What is the justification for the choice of 40,000 years etc. for the initialisations? Do you have a specific criteria by which steady state is defined, that is met within that timeframe?

We aim for our initialisations to produce a sea-level drift below 1 cm per century. For most of the ones presented here, this was achieved in about half of the set duration, so that the

final value is smaller still; we deliberately over-engineered them for safety. We will mention this in the manuscript.

Line 348: I think it's important to note here that there is a significant increase in SLC at the upper end of the range covered by the high-resolution grounding lines, especially considering this high resolution is only applied to one region. It doesn't surprise me that there's little difference between 20km and 5km, but I expect if the 2km and 1km resolution were applied to the Siple coast and Ronne ice shelf where your model sees most retreat, the grounding line resolution would have a much larger effect. Convergence tests in another model have shown that mesh resolution should be under 2km to robustly simulate grounding line migration (Sun & Gudmundsson, 2026).

We agree that the resolution seems to have a stronger effect in the high-end cluster; we will mention this in the manuscript. However, we hesitate to quote exact numbers for the maximum resolution needed for accurate GL dynamics, as this is a detail that can vary strongly between different models.

Line 385-6: I would argue that inversions do nothing to compensate for the lack of calving in models, but that's a minor point of wording. The overall argument being made here is entirely right!

We will rephrase this to “The biases that these missing processes introduce into the present-day ice-sheet geometry and velocity are compensated for by the inversions for basal friction and sub-shelf melt”.

Line 406: This raises an interesting difference between the ‘nudging’ method and the other common approach of adjoint inversion. In the latter, observation errors are taken into account in the cost function. Many of the same points apply to both methods however, regarding the non-uniqueness of solutions and potentially reaching a ‘correct’ solution for the wrong reasons. While tangential to the main point of this paper (and I’m certainly not suggesting any further experiments!), it would be interesting to see some thoughts on whether the choice of inversion method would make any difference to your results.

It might be possible to introduce a similar weighting based on observational uncertainty into the nudging method. We will mention this in the manuscript. Also, implementing an adjoint-based inversion option in our model, and comparing it to the nudging method, is on our long-term wish list.

References:

Goldberg, D.N. and Holland, P.R., 2022. The relative impacts of initialization and climate forcing in coupled ice sheet-ocean modeling: Application to Pope, Smith,

and Kohler glaciers. Journal of Geophysical Research: Earth Surface, 127(5), p.e2021JF006570.

Sun, S. and Gudmundsson, G.H., 2026. The potential of marine ice cliff instability for Amundsen Sea glaciers revisited. Annals of Glaciology, 67, p.e15.