Review for Menthon et al. : Comparison of calibration methods of a PICO basal ice shelf melt module implemented in the GRISLI v2.0 ice sheet model

Submitted to Geoscientific Model Development
Reviewer: Clara Burgard

I do this review un-anonymously to clarify from which background and level of expertise the comments come from and because it makes the conversation during review more transparent on both sides.

> We would like to thank the reviewer, Clara Burgard, for taking the time to review the paper and provide detailed comments to improve the manuscript. We address all the points raised below, our responses to the comments are in blue.

Summary
Ocean-induced melt of Antarctic ice shelves and its representation in ice-sheet models is one of the main sources of uncertainty for ice-sheet projections. The parameterisations used to bridge the gap between ocean properties and the basal melt remain uncertain. The authors revisit this uncertainty and explore a new calibration approach to reduce it. To do so, they implement the PICO ice-shelf basal melt parameterisation into the ice-sheet model GRISLI and test different novel calibration approaches, resulting in a more robust parameter choice. They explore a variety of conditions that affect the calibration such as the metric, the resolution, the geographical specificities, and the forcing conditions.

The study is a timely contribution because the parameterisation of the melt at the base of Antarctic ice shelves remains one of the largest sources of uncertainty in projections of the future evolution of the Antarctic ice sheet and its contribution to sea-level rise. The approach explored by the authors is a useful addition to the community as it provides a set of parameters that seem to be more robust across conditions compared to previous sets. It could also be explored for other parameterisations in the future.

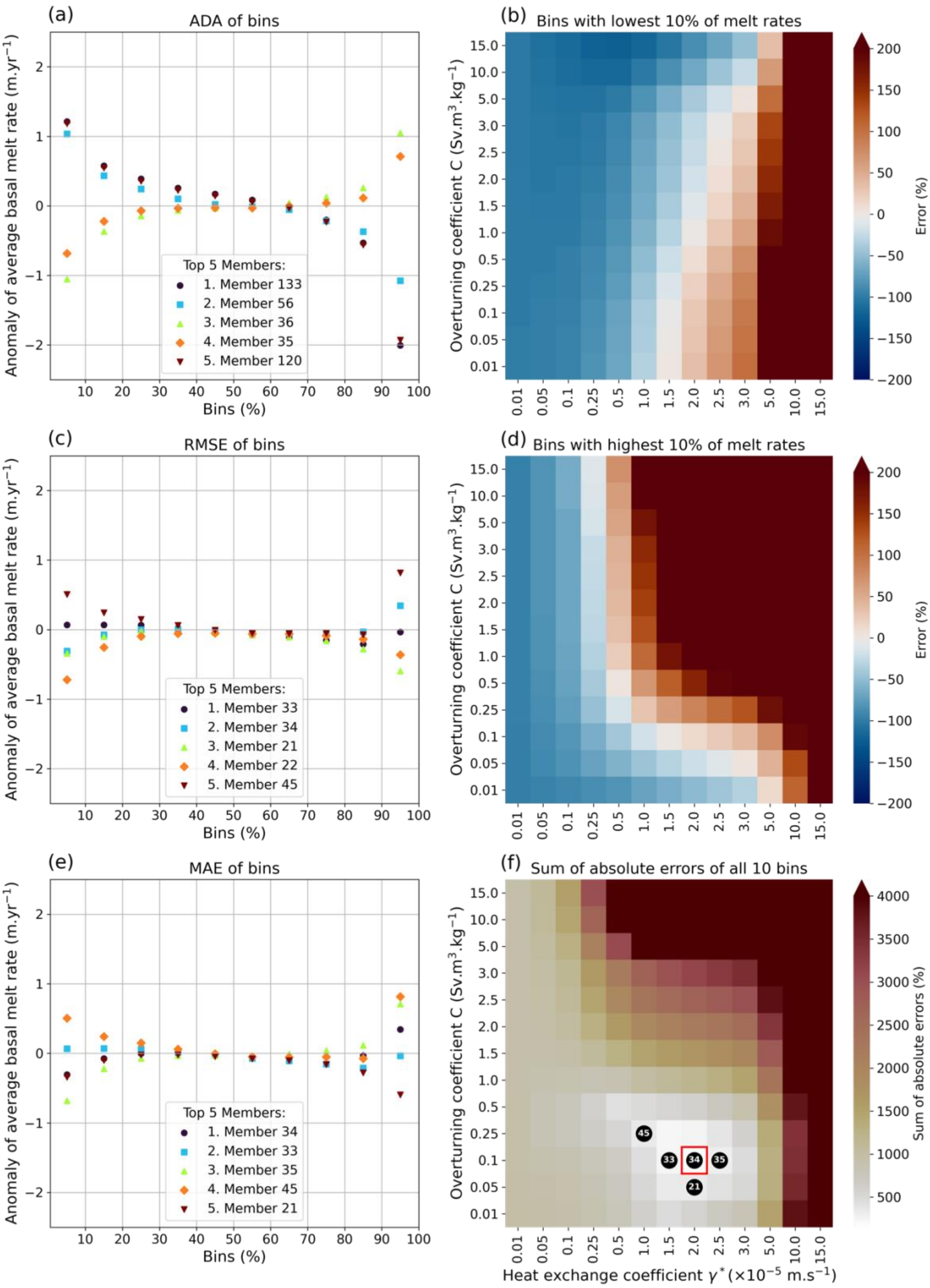
The manuscript is very clear and thorough. This thorough description is very well suited to inspire other researchers to explore the presented methods on other parameterisations and/or other ice-sheet models. Overall, I have a few larger comments as some aspects of the study appeared unclear to me and other more minor comments. I therefore suggest minor revisions.

GENERAL COMMENTS
I want to start by thanking the authors for making this manuscript very pleasant to as the manuscript and figures are very well done and clear.

#1 The results of this study are very encouraging. However, there is a strong limitation, which is not discussed much. In Figure 3, it becomes clear that using bins improves the melt rates for the "middle" bins but, in all cases, the spread between members remains high for the low and high-end bins. In particular, for the lowest melt, an anomaly of 0.5 m per yr could be quite high compared to the actual melt. For the highest melt, the order of magnitude is maybe lower compared to the actual melt. However, it is important to check this further because the points with the highest melt are also the points that influence the future of the ice sheet the most. Would it be possible to complete this by a metric that looks at the percentage formed by this anomaly compared to the melt value? Just to have a better idea of what this means exactly.

We agree, therefore we completed the figure with 3 panels that look specifically at the lowest bin (b), the highest bin (d) and the sum of of the absolute error of the 10 bins (f). All the results are expressed in percentages. We observe different responses between the lowest and highest bins and the new metric is in really good agreement with the results of calibration with the MAE of bins. We therefore added the following text. Also, this analysis has been done with the Adusumilli et al. 2020 datasets (shown in the main text) as well as the Paolo et al. 2023 datasets (added to the supplementary materials), both datasets leads to the same conclusions.

"The largest discrepancies between the binned values of modelled melt rates and those from observational datasets occur at the extremes, that are the bins with the lowest 10% and highest 10% of the values from the distribution. We therefore analysis further these two bins for all the PICO configurations in panels (b) and (d) of Figure 3, respectively. These panels reveal distinct sensitivities to the two PICO parameters. For example, at fixed values of C, increasing $\gamma * T$ consistently raises the bin values in both the lowest (panel (b)) and highest (panel (d)) deciles. This implies a reduction in error when the model underestimates melt (negative error, in blue), or an amplification of error when it overestimates melt (positive error, in red). In contrast, at fixed $\gamma * T$, varying C can produce divergent effects between the lowest and highest bins. For instance, at $\gamma * T = 2.0 \times 10-5$ m,s−1, increasing C leads to a decrease of bin error values in the lower 10% bin (b) and an increase in bin error values in the upper 10% bin (d). Finally, by computing the sum of the absolute values of the errors for all 10 bins we can find the combinations of PICO parameters that minimize this error the most. The panel (f) of Figure 3 shows the results with superposed the best members according to the MAE of bins methods shown on panel (e). We find that the two metrics, MAE of bins and sum of absolute errors of the bins, leads to a similar selection of the best ensemble members."



(a)
ADA of bins
Anomaly of average basal melt rate (m.yr^{-1})
Bins (%)
Top 5 Members:
1. Member 133
2. Member 56
3. Member 36
4. Member 35
5. Member 120
(b)
Bins with lowest 10% of melt rates
Overturning coefficient C (Sv.m^3.kg^{-1})
Error (%)
(c)
RMSE of bins
Top 5 Members:
1. Member 33
2. Member 34
3. Member 21
4. Member 22
5. Member 45
(d)
Bins with highest 10% of melt rates
(e)
MAE of bins
Top 5 Members:
1. Member 34
2. Member 33
3. Member 35
4. Member 45
5. Member 21
(f)
Sum of absolute errors of all 10 bins
Sum of absolute errors (%)
Heat exchange coefficient γ^* ($\times 10^{-5}$ m.s^{-1})

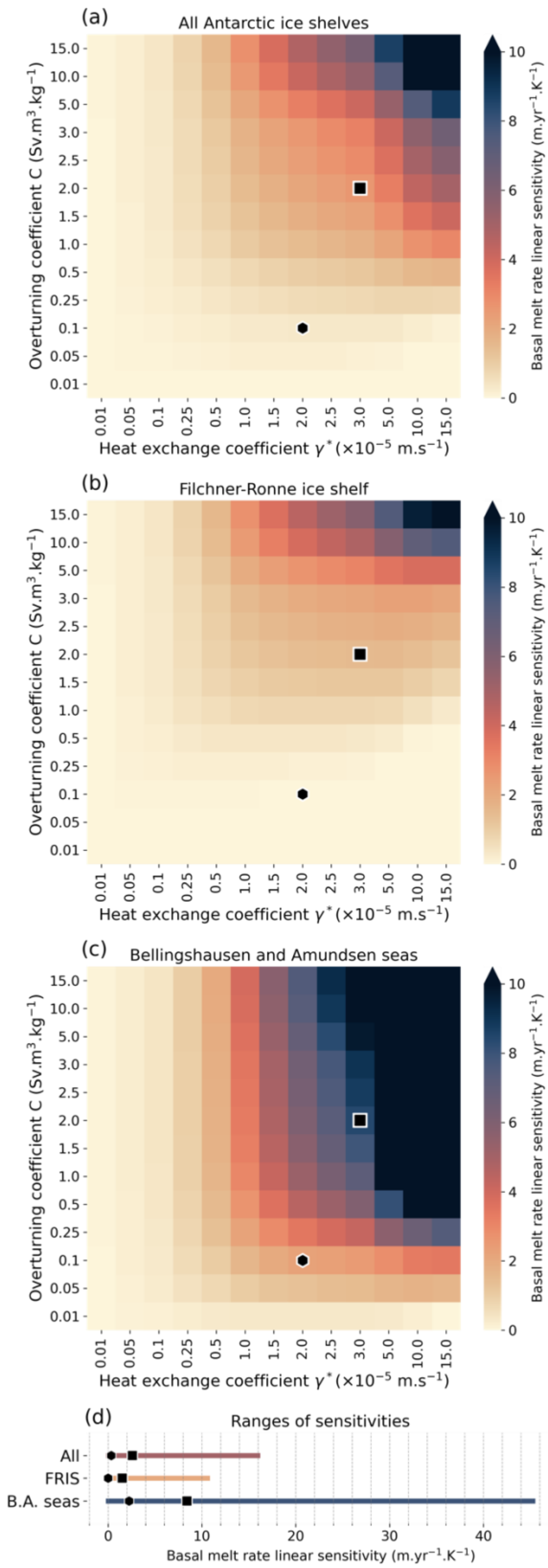
#2 I am not completely convinced by section 4.2. I do not understand how the authors can calibrate on different input temperatures (e.g. 1K warmer) but same target melt and make conclusions from this. If the parameters do not change much, does that not mean that they are not sensitive enough to temperature changes? Ideally, you would expect higher melt for higher temperatures if you use the same set of parameters, no? This makes me unsure about the whole approach. Does that mean that the parameterisation cannot react to changes in forcing and that the parameters are too strongly set? This would not be useful for projections and is what could be interpreted from the low sealevel rise contribution in Fig. 7c. To reassure the reader, I recommend that the authors clarify the implications of this result, discuss them more in detail or reformulate to avoid misunderstanding.

We agree with this concern, to answer and quantify it we decided to add a new subsection (with a new text and a new figure) about "What is the sensitivity of PICO?". In the new figure we quantify the sensitivity of PICO thanks to the ensembles with + 0 K and + 1 K forcings, and we separate the sensitivity of the 3 areas of interests used earlier.

The following text is added at the end of the subsection 3.4
"Since the best parameters do not vary much (about - 0.5 for the $\gamma*T$ and - 0.05 for C) between the + 0 K and the + 1 K forcings, we analyse in the next subsection what is the sensitivity of PICO to understand better what would be the response of PICO to warmer than present-day conditions."

And the following text and the Figure with sensitivities values is the subsection 3.5:

"Thanks to the ensembles with + 0 K and + 1 K ocean forcings from the previous subsection we can determine the sensitivity of PICO for all the ensemble members. We are here assuming a linear sensitivity and therefore compute it as the difference between the + 1 K experiment minus the + 0 K experiment. The results are shown in Figure 8 where we also differentiate the sensitivity in the 3 areas defined for the analysis shown in subsection 3.2. First of all, we see that in most cases the sensitivity of PICO increases when the value of either of the two parameters is increased. Second, we see that the sensitivity varies between areas. For instance, 1 K of warming with the calibration from PISM-PICO (squares on the Figure 8) would lead to an increase of 1.5 $m.yr^{-1}.K^{-1}$ in Filchner-Ronne ice shelf, whereas it would be 8.4 $m.yr^{-1}.K^{-1}$ in the Bellingshausen and Amundsen seas ice shelves. We also see that the range of possible sensitivities (panel (d)) is about four times larger in the B.A. seas than in the FRIS. These results quantify how much the sensitivity of the basal melt rate would change, globally and regionally, by changing the values of the two PICO parameters. In all the cases, the best calibration with the MAE of bins (black hexagons) is in the low range of all the tested combinations of parameter values. These values are also lower than the range of Antarctic ice shelves sensitivity estimates from some previous studies (Levermann et al., 2020; van der Linden et al., 2023), but closer to the PICO sensitivity obtained by Reese et al. 2023; Lambert and Burgard (2024) when optimising parameters for present-day melting. The methodologies in the assessments of the sensitivities are however different in each study. Nonetheless, based on the results show in Figure 8 we can expect a low to moderate response of ice shelves in this calibrated version of PICO to future projections scenarios."


(a)
All Antarctic ice shelves
Overturning coefficient C (Sv.m^3.kg^{-1})
15.0
10.0
5.0
3.0
2.5
2.0
1.5
1.0
0.5
0.25
0.1
0.05
0.01
Heat exchange coefficient γ^* ($\times 10^{-5}$ m.s^{-1})
Basal melt rate linear sensitivity (m.yr^{-1}.K^{-1})
0
2
4
6
8
10
(b)
Filchner-Ronne ice shelf
(c)
Bellingshausen and Amundsen seas
(d)
Ranges of sensitivities
All
FRIS
B.A. seas
0
10
20
30
40
Basal melt rate linear sensitivity (m.yr^{-1}.K^{-1})

DETAILED CONTENT COMMENTS
L3 : It is not only about limited computational resources but also about a lack of observational data. I suggest completing "limited observational data and computational resources".

> Now the sentence reads as follow: "As a result of limited scientific understanding of these ice-ocean interactions, poor observational data, existing biases in earth system models, as well as technical challenges related to computational resources and implementation, these interactions are parametrized rather than explicitly resolved in most ice sheet models."

L31: Not sure if the formulations only differ in the complexity of the "melt physics" themselves. In most cases, assumptions differ on the simplification of the ocean circulation in the cavity. Can the authors reformulate to clarify?

> Now the sentence reads as follow: "Over the last decade, several basal melt parameterisations have been developed and implemented in ice sheet models with different complexities in the simplification of the ocean circulation beneath the ice-shelves and its physical interaction with the ice."

L39-40: This is great! Calibrations that do not need regional corrections are what are currently needed!

> Thank you.

L104-108: This paragraph is unclear in regard to the authors' own contribution vs what has been done in PICO-PISM before. I suggest reformulating it to clarify.

> Clarification were made as follow, with the new parts in bold: **"The grounding line can be defined in different ways and therefore can lead to different PICO boxes geometries. Here, in GRISLI-PICO,** we consider as grounding line any ice points that is surrounded by grounded ice and not grounded ice, and as ice front any ice point that is floating and adjacent to ocean. **Whereas, in** PISM-PICO (Reese et al., 2018) they did not include the grounding line of ice rises and also excluded holes in ice shelves as ice shelves front when identifying PICO boxes. The grounding lines of ice rises are defined as not being directly connected to the main grounded part of the ice sheet which is identified by the size of the connected grounded region. Thus, **in PISM-PICO it** is possible to have ice shelves without grounding line connected to the main ice sheet, where PICO cannot define a box geometry. In these places, the parametrisation of Beckmann and Goosse (2002) **was used in PISM** to have a rough estimate of the basal melt rates."

L114: It is unclear where the value for the combined factor comes from. As far as I know K_T is a parameter that needs calibration. So how has it been calibrated in this case? I suggest the authors add the source or briefly explain the calibration.

> No additional calibration have been done for this K_T, therefore we clarified the sources:
> The combined factor KT ρw Cw ρiLf equals to 0.224 m.yr−1.◦C−2 as we keep the same KT value of 15.77 m.yr−1.C−1 from DeConto and Pollard (2016) and Pollard and Deconto (2012).

L122-124: It remains unclear to me why the authors use the coupled GRISLI-PICO setup if the geometry is kept constant anyway. Could the authors clarify why they do not use a standalone version of PICO then?

> It could have been done in a stand-alone PICO indeed. This choice has been made because the goal is to use PICO coupled with GRISLI as quickly as possible. Doing the coupling PICO with GRISLI leads to some changes (reading variables differently etc.), therefore doing the stand-alone and the coupled would have lead to more work for very similar results. We added the following in the text:
> "Doing the calibration of PICO in a coupled GRISLI-PICO with fixed geometry enables to facilitate the transition between the PICO calibration and the GRISLI-PICO transient experiments (see section \ref{m-future}) without impacting the results."

Table 1: Very useful table!

> Thank you.

Table 1: Maybe I did not think this completely through but aren't ADA and ADA of bins the same metric. Taking the mean of a mean should, I think, result in the same as taking the mean of the whole ensemble directly. This would also explain why the results are so similar between the two. Can the authors comment on that?

> We agree with your point. The very small differences observed between the two methods is due to the sampling of the bins, in other words how many grid cells values are per bins, and despite this the weight given to each bin will remain the same.
> For instance:
>
> - (1+2+3+4+5+6+7+8+9+10)/10=5.5
> - Bin 1: (1+2+3)/3=2
> - Bin 2: (4+5+6)/3=5
> - Bin 3: (7+8+9+10)/4=8.5
> - Average of the bins=5.16
>
> We therefore clarified this with the following text in the paper:
>
> "This method do not leads to exactly the same results as the ADA without binning because some bins contain slightly different numbers of grid cells than others, meaning that they cover slightly different areas but still get exactly the same weighting in the binning approach. As we intend to compare methodologies, we consider it relevant to also test this method, even if it gives results very close to ADA without binning, to quantify how much they differ."

L218-219: I am not 100% sure that a narrow set of parameters is a guarantee for "better". I agree it is useful for modellers but I would be careful with such kind of statements. If a large range of parameters are possible, this could also be linked to the formulation of the parameterisation. Still, in L294-296, the authors explain the advantages of having a narrow set of parameters. Maybe this could already be mentioned here?

> We argue that a narrow selection of best parameter sets proves that the results will be less dependent on the sampling choices for the ensemble, therefore the calibration methodology is better. However, we agree that it does not necessarily means that the

values of the parameter defined by the methodology will be the right ones for the rights reasons, the parametrisation remains an (imperfect) simplification of the physical processes. On the other hand, if a parametrization works equally good or bad for any parameter value, then the parametrization isn't useful, so in that sense a parametrization with a narrow range of parameter values is better.

L241: It is unclear to me what is the difference between the distribution curve and the magnitude of the spatial patterns. Is one the shape of the curve and the other the actual number?

No, the distribution curve is what is shown on figure 2 panels a, b, c g, h and i (the compilation of all the values regardless to their geographical coordinates); whereas the magnitude of the spatial patterns is shown on figure 4 (the values at their geographical coordinates). We therefore help the reader by referencing to the figures:

"[…] i) better able to match the distribution curve from the target (Figure 2, panels (a), (b), (c), (g), (h), (i)), ii) systematically give best values in the same small range of values (Figure 2, panels (d), (e), (f), (j), (k), (l)), and iii) the magnitude of the spatial patterns is similar to the target (Figure 4)."

Figure 8: Just a remark: The 2D results are interesting and give the feeling that C does not really play a role in the calibration. Out of curiosity, have the authors thought about what this could mean?

Good question, to which we do not have a definitive answer. One idea could be that for higher gamma, PICO extracts heat from the available ocean forcing quite efficiently, and the limitation on melting is how much heat is supplied (which is constrained by C). For lower gamma values, less heat is extracted from the ocean reservoir, and hence the overturning does play less of a role.

L330: Agreed that Joughin et al. (2021) is a good study to refer to here but it should not be forgotten that enough other studies (e.g. Reese et al. 2018) show that localised melt has a strong effect on buttressing. I suggest that the authors reformulate a little more carefully.

We added explicitly some nuance in our statement as follow: "However, other studies suggest that localized sub-ice-shelf melt can have a strong impact on the buttressing or that in more strongly buttressed areas sub-ice-shelf melt would have outsized effect (Gudmundsson, 2013; Reese et al., 2018b)"

L350: This is not completely true. The quadratic term of the quadratic parameterisation is there to mimic the effect of the overturning circulation in a very simple way.

We revised this point as follow:

"The overturning circulation under the ice shelves, which tends to reduce the basal melt rate, is computed differently in the parameterizations PICO and QuadNL. In PICO the overturning fluxes is computed with the overturning circulation coefficient C and the difference of densities (see equation 2). Whereas in the QuadNL it is in the product involving the thermal forcings which results in stronger overturning from warmer conditions (see equation (1) in Jourdain et al. (2020)).”

L360-362: Yes, it can be clearly seen in the sensitivities to warmer forcing in Burgard et al. (2023) and Lambert and Burgard (in press) that the quadratic parameterisation is an outlier towards high melt sensitivity. However, as we do not know what is the “right” sensitivity, this is not enough to say that one is better than another.

We completed the argument, the whole point reads as follow:

“The QuadNL calibration from Jourdain et al. (2020) implemented in the analysed simulations does not match the refreezing part of the observations whereas PICO does (see figure 6 panel (b)), therefore this QuadNL calibration could overestimate its sensitivity to oceanic forcings, which is in agreement with recent studies (Burgard et al., 2022; Lambert and Burgard, 2024). However, even if the QuadNL parametrization is on the high-end in terms of sensitivity compared to other parametrizations, neither PICO nor QuadNL sensitivities (which also depend on their calibration) can be rule-out as we do not know what the right sensitivity is.”

L363-366: This is not very clear. I suggest that the authors reformulate to clarify.

DETAILED WRITING COMMENTS

L20: Replace “warmth” by “heat”

Done as suggested

L22: “on the other hand” does not really work in this sentence. I suggest leaving it out.

Done as suggested

L34: Leave out “However”, it is confusing.

Done as suggested

L41-49: This could be shortened.

It has been shortened as follow:

“The article presents first the methodology in section 2, then the results in details in section 3 including: calibrations methods comparison, sensitivity estimates and future projections. The section 4 discusses the limitations and possible improvements for next studies. Finally, section 5 concludes and gives some perspectives.”

L54: Replace “are dependent” by “depend”

Done as suggested

L79 and 84: For the results, I suggest to stay consistent with present instead of past tense: “made” => “make”, “combined” => “combine”

Done as suggested

L83: Correct the citation format (\citet{} instead of \citep{})

Done as suggested

L98: Rephrase to “refreezing in some areas”

Done as suggested

L142: Replace “but also” by “and”

Done as suggested

L148: To improve reproducibility, I suggest that the authors add the information if the results are in m ice per year or in m w.e. per year.

We added: “(m of ice equivalent per year)”

L158: I suggest that the authors reformulate, the formulation is very unclear.

It has been rephrased as : "By applying the three first ranking methods, the ranking metrics do not enable to pick systematically the ensemble members with the best fit to the distribution of values of the observational dataset."

Figure 1 caption, last sentence: Replace "is" by "are"

Done as suggested

L190: Can the authors clarify if they are writing about the sensitivities of PICO and QuadNL to ocean warming or to something else?

This rephrased as follow for clarity: "the difference of sensitivity between PICO and QuadNL"

L209 and later: "side-by-side" sounds awkward. I suggest reformulating, maybe with something like "close" or "similar".

We disagree, close or similar would mean that the points are close but not necessarily "touching" each other, which is what we are trying to say here.

L220: Replace "explain" by "explained"

Done as suggested

L224-225: I suggest reformulating as this is not a complete sentence.

We suggest the following:
"Whereas, the RMSE of bins and the MAE of bins have systematically smaller anomaly values and do not allow for compensating effect."

L232: "led" => "lead"

Done as suggested

L235: remove "is"

Done as suggested

L236: missing "methods" in front of "without"

Indeed, done as suggested

L269: "Elemer" => "Elmer"

Corrected as suggested

Figure 6 and later: Would it be possible to replace CalibXX by an indication of the metric it was calibrated with? That would clarify the legend.

We understand this suggestion and we tried it. However, it makes the text a lot less clear (need to mention two values instead of one every time we want to refer to one simulation). Also, we believe it is important that the reader understand where the calibration is located in the parameter space and therefore it might be good to force the reader to come back to Figure 10 (a).

L284: I suggest reformulating "we further discuss the ISMIP 2300 …"

Changed to: "Lastly, we discuss the ISMIP 2300 results obtained with PICO further."

L359: "rates" => "rate"

The sentence has been changed to: "PICO tends to have more smoothed out melt rates and does not show significantly higher melt rates at grounding lines, as seen in satellite-derived fields."

L401-402: This is not a complete sentence.

We believe it is a complete sentence.

Supplementary material: There is A LOT of material and the captions are sometimes very short. I wonder if it would be possible to reduce the amount of figures or add one sentence explaining the core of the figure or set of figures when appropriate?

We strongly reduced the number of figures by removing all the figures showing results for the methods: 2D RMSE, ADA of bins and RMSE of bins. We argue that they show similar results to 2D MAE, ADA and MAE of bins (respectively), we therefore kept only the three later onces.

Moreover, we added an introduction to the supplementary to justify this choice and help the reader through the supplementary materials.
However, please note that in order to reply to the review thouroughly, two figures have been added to the supplementary materials: Figure S1 showing the PICO boxes and the drainage basins division, and Figure S29 showing the analysis of the bins but with target the datasets from Paolo et al. 2023 instead of Adusumilli et al. 2020.

References
- Reese et al. 2018 : https://doi.org/10.1038/s41558-017-0020-x
- Burgard et al. 2023: https://doi.org/10.1029/2023MS003829
- Lambert and Burgard, in press: https://doi.org/10.5194/egusphere-2024-2358