

We would like to thank the editor and the two anonymous reviewers for their effort to review our manuscript and greatly appreciate their helpful comments for improving our study. We will address the raised points during the revision and provide detailed replies to all comments below, with our responses indicated in blue.

Comment on tc-2021-348

Anonymous Referee #1

Referee comment on "Calibration of basal melt on past ice discharge lowers projections of Antarctica's sea level contribution" by Eveline C. van der Linden et al., The Cryosphere Discuss., <https://doi.org/10.5194/tc-2021-348-RC1>, 2021

In the manuscript, van den Linden et al. use linear response functions from Levermann et al. (2020) which were derived from ice sheet model responses in sea-level contribution to perturbations by uniform sub-shelf melt rate increases. Sea-level projections are updated by using CMIP6 instead of CMIP5 models and by recalibrating the sensitivity of melt rates to ocean temperature changes using observed mass changes. The authors conclude that with the new calibration, sea-level projections are lowered in comparison to LARMIP2 (Levermann et al., 2020) and ISMIP6 (Seroussi et al. 2020; Edwards et al. 2021).

First of all, I want to thank the authors for this well written manuscript which is easy to understand and follow. Unfortunately, I think that the approach presented in the manuscript cannot be applied this way. However, the approach could be interesting and informing further studies, so I suggest two possible modifications that would make it applicable in a methodologically correct way.

Major comment:

The central issue is that the calibration factor γ , which relates ocean temperature changes at depth to sub-shelf melt rates changes, is fitted over the historic period in the Weddell Sea, Ross Sea and in East Antarctica, all three regions where very little mass gains or losses have been observed (see Rignot et al. 2019; Figure A1 in the manuscript), in particular in comparison to the overall volume of the regions (if the methodology can be applied to the Antarctic Peninsula should also be checked). This makes a sound calibration with changes in ocean temperatures from CMIP6 models over the historic period basically impossible, due to a number of issues: (1) the changes in mass in the respective region might not even be causally linked to ocean forcing but explainable through, e.g., surface mass balance changes; (2) changes in ocean temperatures in CMIP6 models show a wide spread and how close they are to real changes and if they can actually capture subtleties in the historic record that can be linked to the small changes in ice discharge, is questionable.

This means that the calibration factor is fitted between two numbers that are zero or quite small, but have large uncertainties, so that in the end the calibration factor is not really constrained. And this is physically correct, because if there is no enhanced ice discharge due to changes in ocean forcing in the historic record, as for example in the Weddell Sea, the melt sensitivity to ocean forcing cannot be deduced from observations. This problem shows for example in the result that most calibration parameters are zero in many icesheet-

ocean-model combinations (section 3.1). And that, even if the parameters are fitted to represent past discharge, they largely underestimate the observed mass loss (Fig. 6).

Two suggestions to avoid this problem are:

1) focus the study on the Amundsen Sea (potentially also the AP), where actually enhanced ice discharge has been documented extensively and linked to enhanced ocean driven melting. This would allow you to derive a sound fit for that region. Then you could compare the projections for the Amundsen Sea with ISMIP6 and LARMIP2 for that region. Alternatively,

2) if you want to include the whole of Antarctica, you could use your proposed method to fit gamma in the Amundsen Sea (and AP), where substantial discharge occurred, and assume an uncertainty distribution of gamma (based on LARMIP, your fit, ISMIP6 calibration,..) for the other regions.

We would like to thank the reviewer for this constructive criticism. We agree that our results show that the calibration factor gamma cannot be well constrained in regions where historical ice discharge is small. As we discuss in the text and as shown in the titles of the panels in Figure 5, the gamma values for the Ross, Weddell and Peninsula regions are positive for less than half of the model pairs. In contrast, for the Amundsen and EAIS regions, around 80% of the model pairs have a positive calibrated gamma value. In these two regions, the historical period shows a warming in the multi-model mean subsurface ocean temperature (Figure 3) and an increased ice discharge in the observations (Figure A1). Since only for the Amundsen region past ice discharge has been linked to ocean warming, we agree that regional calibration on the Amundsen Sea would be most sound. In the first version of the manuscript, we applied a model selection to circumvent this issue of models that could not be calibrated, but we agree that the proposed solution by the reviewer has a more solid physical basis. Therefore, we will only apply the regional calibration to the Amundsen Sea region. In this way, we are certain that ocean forcing is the main driver and that ice discharge signal is large enough to obtain a sound calibration. We will include projections for the Amundsen region in the body of the manuscript and include a comparison with LARMIP and ISMIP6 for the Amundsen region.

Following the second suggestion of the reviewer, we will also apply these Amundsen calibrated gamma values to the other regions of the Antarctic ice sheet for the quadratic relation. This will lead to lower melting sensitivities than in the Amundsen region since by its definition the quadratic relation implies that sectors that are melted by warmer waters are more sensitive than colder sectors. The nonlinear relation between melt and temperature change found in observations (Figure 5, Jenkins et al. 2018, Nature Geoscience) also suggests that the quadratic calibration based on the Amundsen region might be applicable to the cold-water sectors.

In addition to the Amundsen calibration, we believe that an Antarctic-wide calibration could still be interesting to apply as a lower bound for the future projections. This Antarctic-wide calibration will provide information on what would happen if the total Antarctic ice sheet would keep the same basal melt sensitivity to ocean warming in the future. Calibrating on the total Antarctic ice sheet also means that all regions will be sensitive to future warming and no single region will have a gamma of zero. This could be considered as a low-end scenario since it includes regions with very small ice discharge during the calibration period,

which are expected to melt in the future as the climate warms. We will apply both the linear and quadratic relation for the Antarctic-wide calibration, so that we can compare our results with Levermann et al. (2020).

Another improvement will be that for both the regional Amundsen and Antarctic-wide calibration we will only select the model pairs that have a calibrated gamma value greater than zero. Models that could not be calibrated (i.e. with a gamma of zero) will not be used for the projections.

Since these would mean major changes to the manuscript and potentially the central findings, in the following I mostly omitted comments on the methodology and results that will be affected by the major comment above:

Minor comments:

the study uses “discharge” in some places, but what it actually means is “changes in discharge”, as the latter should in principle be associated to changes in sub-shelf melt. Please check and correct.

We agree, this will be corrected.

page 1, line 24: specify “long-term”

We mean thousands of years. This will be specified.

page 2, line 53-55: ISMIP6 used more than the quadratic calibration, and those were two options for calibration (not done at the same time)

Thanks for pointing this out, we will correct this.

page 3, line 69, line 85: this is discussed as if the scaling factor in LARMIP only has disadvantages, but it actually has the advantage that also uncertainties in the global mean temperature changes (and not only the CMIP trajectories) were included in the uncertainty estimate

We will include this advantage of using GSAT in the text.

page 5, line 100; Table 2: the sub-surface temperatures used as ocean forcing look too shallow for me. The relevant water masses are those at depth of the continental shelf that drive the melting close to the grounding lines. I would expect these more around 800 to 1000m depth.

We used the same depth values as in Levermann et al. (2020), which are the mean depths of the ice shelf undersides. Ocean circulation characteristics, such as cavity circulation, are not explicitly included in this approximation, although the quadratic parameterisation attempts to account for this. To test how our results depend on the depth, we will repeat the computations with 800-1000m depth ocean temperature data and discuss how this influences our results in the manuscript.

page 6, line 118-120: why do you not correct for individual model biases but by the ensemble mean?

We actually do correct the individual models with the ensemble mean of the *reanalysis* products. We will clarify this in the text.

Fig. 3 caption: change to “Annual... time series of the CMIP6 multi-model mean (green) , model drift and bias-adjusted, and the GREP ensemble mean (orange). Both are smoothed by a five year running average filter.”

This sentence will be rephrased.

page 7, line 121: this sounds weird, the water is not warmed in the cavities, but it's the change in the pressure that lowers the freezing point and increases the thermal driving

What we mean to say is that the water in the cavities becomes warmer over time due to inflow of warmer water, but we understand the confusion. We will rephrase the sentence.

page 6, line 137: that T_f is a constant for each region is a very coarse assumption, could be discussed in the discussion.

We will discuss this assumption in the discussion.

page 9, line 171-174: why this condition for the “unbounded”? Would you get a better fit for the Amundsen Sea if you would remove this?

It should be noted that the best models for the Amundsen Sea are not selected, only the best models for the Antarctic summed response (Figure A2). By applying model selection on the Amundsen region we will get a better fit for the Amundsen Sea. To avoid confusion, we will test if this upper bound has a significant impact, but since it only applies to a few models it will probably not have a large influence on the multi-model mean response.

section 3.2: LARMIP2 did not calibrate the melt-factor with observed changes in ice discharge, but it did compare the obtained mass losses over the historic period, which actually look like a much better fit than your results (Fig 6, Levermann et al. 2020).

It should be noted that our Figure 6 shows the results of all model pairs, including those that could not be calibrated ($\gamma = 0$), resulting in an underestimation over the historical period. We will replace this Figure by one that includes only models that could be calibrated ($\gamma > 0$).

section 3.1: do you have an estimate on the uncertainty of your calibrated melt factors for each combination?

We do not have error estimates yet for each ESM-RF combination, but it is a good idea to include these. We will add an error estimate for the calibration based on the error estimate

in ice discharge observations by calibrating discharge on the observed discharge +/- one standard deviation error in ice discharge for each ESM-RF pair.

figure 5: instead show the sensitivity in m/a/K, your legend is hard to see, maybe increase the intensity of the colors?

Thanks for pointing this out, we will increase the color intensity.

page 13, line 239: but they should, by construction. This indicated the underlying problem with the methodology.

We agree that this shows that calibration is not possible in all regions. Therefore, we will follow the suggestion of the reviewer by applying the regional calibration only on the Amundsen region and include only the models that could be calibrated (about 80% of all model pairs).

discussion: you could add Payne et al. 2021 for a comparison between CMIP5 and CMIP6 effects on AIS sea-level projections; you should discuss the errors that arise through not including surface mass balance changes in your fitting procedure.

We will add a short discussion on AIS sea-level projection based on Payne et al. 2021. However, it should be noted that we do not aim to reproduce total mass loss, but purely ice discharge (without surface mass balance changes). This was also our motivation to choose the Rignot dataset for fitting, since it includes pure ice discharge. Therefore, our fitting procedure does not include an error due to surface mass balance changes.

Comment on tc-2021-348

Anonymous Referee #2

Review of van der Linden et al. "Calibration of basal melt on past ice discharge lowers projections of Antarctica's sea level contribution".

General Comments:

In this paper, the authors use linear response functions of ice sheet models from LARMIP-2 to estimate contributions to sea level in three future emissions scenarios. Forcing is derived from ocean conditions in CMIP6 models and is further bias-adjusted and corrected for model drift. Their methodology resembles that of LARMIP-2, however the key difference is that instead of calibrating basal melt rates on observed melt rates (as in ISMIP6), basal melt rates were calibrated on observation-based estimates of ice discharge (Rignot et al., 2019). The authors use two different forms of melt rate parameterization (linear and quadratic) and then use their derived values for gamma to update sea level rise projections under three SSP projections out to 2100. They conclude that their method results in lower projections of Antarctic sea level contribution than the other main methods used in IPCC AR6 (ISMIP6 and LARMIP-2).

I want to thank the authors for this work, and their new approach in calibrating to a different observational dataset than has been typically used in the community so far. They have recognized and tried to address a need to better-constrain gamma values. That said, I have a few broad-brush concerns about the paper.

First, the calibration of basal melt rate is based on past ice discharge in regions (e.g. Ross, Weddell) where very little ice loss was taking place. As a result, gamma values are often equal to zero, and so future contributions to sea level from these regions is inevitably muted despite the potential for actual increases in thermal forcing. My concern here is that it is hard to derive a meaningful value for gamma in regions where there is minimal mass loss, and therefore hard to deduce the actual sensitivity of the region to changes in ocean temperature. The uncertainty around CMIP6 ocean warming coupled with the lack of historical mass loss in some of these key regions makes finding a valid range of gamma very difficult. This problem is manifested when the authors find that ESM-RF pairs underestimate the magnitude of observed Antarctic sea level response despite being tuned to match it. Furthermore, after picking the top 10% of ESM-RF pairs, some regional hindcasts do not capture the observed sea level response (Figs A1 & A2). As a result, I'm concerned that the conclusions drawn about the future contributions to sea level, and their uncertainty analysis, are rendered less credible.

We agree that it is difficult to derive gamma values from regions with minimal mass loss over the calibration period. This could indeed lead to an underestimation of the contributions of these regions in future projections. We will therefore adjust our method, by applying regional calibration only on the Amundsen region (in addition to the Antarctic-wide calibration), as proposed by reviewer #1, since this region shows significant mass loss that is mainly caused by ocean forcing. For a more elaborate answer we refer to our reply to the first comment of reviewer #1.

Second (and this is a correctable issue) I felt that the rationale/motivation for doing the calibration on ice discharge rather than basal melt rate was not well-articulated. I think the authors could add stronger language for why this method is worthwhile. This should also be put in the context of ISMIP6 and LARMIP-2 methodologies, and what potential issues calibrating to basal melt rate could create.

We will specify this better in the text. We will include these two main points: 1) We want to capture the ice discharge acceleration during the historical period which cannot be derived from basal melt itself due to the delayed response of ice discharge to basal melt. 2) ice discharge measurements capture the entire ice sheet through satellite measurements of ice height and velocity and therefore are better constrained than basal melt estimates which are not measured for the full ice sheet and for the full time period that we use for calibration.

Specific Comments:

L41: UK-ESM has included an evolving ice sheet in a GCM framework. And there are other efforts currently in the works in other groups to include an evolving Antarctic ice sheet in a ESM. So not sure what is meant by 'short-term' here I guess.

We mean that it is not yet the standard CMIP6 setup so ESM setups with coupled ice sheets cannot yet be used in the 'short-term' in multi-model studies. With 'short-term' we mean in the coming years. We will clarify this and shorten the discussion on coupled ice sheet-ESM modelling in this paragraph, since it is not the main topic of the current paper.

L69: Please specify why GSAT is a less desirable metric than subsurface temp. Or alternatively, why is it a more useful metric? Easier to derive?

GSAT is easier to derive, but it does not account for (future changes in) Southern Ocean dynamics. Deriving forcing underneath ice shelves from global mean surface temperatures is quite a rough assumption. Southern Ocean dynamics is probably difficult to capture in just a scaling factor and time lag. It could be expected that a regional metric has a better relation with forcing underneath ice shelves (see also Lambert et al., 2021). We will expand the discussion on this issue in the text.

L72: How is future loss consistent with past loss? Please elaborate on pros/cons of ISMIP6 calibration methodology and your methodology here. What makes your useful? Motivation needed.

We mean that the physics should be consistent, under the assumption that no new processes are taking place. So, if we apply a certain melt relation for future projections, it should also be applicable during the historical period. Otherwise, the melt relation is not consistent. Therefore, we apply the same melt relation to the past and the future. If this does not work, it shows that either there is a deficiency in the melt relation or new processes are taking place. We will specify this better in the text.

Table 2: Is this the mean grounding line depth? I believe so, but if it is, shouldn't Amundsen GL depth be closer to ~500m?

It the mean depth of the ice shelf base, similar as in Levermann et al. (2020).

L108 – L114: Could rearrange this to start with the motivation first (ocean T bias will affect magnitude of basal melt rate in quadratic estimate, therefore these are the steps we take to deal with it). Currently this feels like you don't know why you're dealing with bias correction until after it is explained.

Thanks for pointing this out, we will rephrase this paragraph.

L118: Why remove the mean bias, which may even have the opposite sign to the model-specific bias? Why not just remove the bias for each model?

This sentence is not well-written. We mean that we remove the bias for each model using the mean of the *reanalysis* products. We will clearly state this in the text.

L143: Please note why/how this assumption is flawed.

The assumption that the inflow/ambient temperature remains constant in the cavity is flawed since it is affected by mixing with meltwater. Also, ocean dynamical processes are not included in this assumption (e.g. the cavity circulation). We will add this in the text.

L148-149: Is there wide variability in salinity between far-field and sub-shelf? I.e. Please comment on whether using far-field salinity climatology is a fairly broad assumption, or not? Please discuss. The same goes for T_f too.

Salinity is used to compute the freezing point temperature T_f . We will analyse the sensitivity of T_f to salinity changes and comment on it in the discussion.

L166: consider rephrasing this since but here you say sea level contributions are calibrated on ice discharge, but the rest of the paper states that basal melt rate is calibrated on ice discharge. I understand what you mean here, but this may be confusing for readers.

We understand this confusion and will change the sentence to be consistent with the rest of the paper.

Section 2.3 in general requires more work to increase clarity for the reader. Please describe the iterative process, how it works. And please expand on the rationale for why to do unbounded vs. bounded methods, as well as the discussion on how the 'unbounded' calibration range is determined. The final paragraph of this section also needs some more clarification.

We discuss these items indeed very shortly and will further elaborate on the methodology.

L138: Please define T_0 . It is not defined.

It is defined near Equation 1 in Section 2 (Methodology).

L227: What calibrated gamma values? All of them? The full range? The median? Mean? This is unclear to me.

The values are shown in Figure 5. We use for each ESM-RF pair the best-fitting value (with the lowest RMSE compared to Rignot ice discharge). This will be specified in the caption.

L239-242: What is "reasonable extent"? Also, this seems fairly problematic to me. If the ESM-RF pairs cannot capture the magnitude of the observed Antarctic sea level response, even though they are calibrated to do so, what does this say about the methodology? This is assuming that the past conditions and sensitivities hold into the future as well, which may or may not be true, particularly when feedbacks become triggered (e.g. MICI). This is illustrated by the observation that pre-2010 discharge is overestimated, and post-2010 discharge is underestimated.

Reasonable extent is a subjective term. We believe that if half of the models has a hindcast with a magnitude that is less than half of the observed discharge we should not use them for future projections. Therefore, we apply a model selection in which we only select models

that had the best results in reproducing the past. It could be the case that some models do not resolve the relevant processes such as ocean dynamics. We do indeed assume that past conditions hold in the future since we apply a constant gamma value, although it should be noted that the quadratic relation also depends on the absolute ocean temperature and therefore the basal melt sensitivity will change over time. Our results could lead to the conclusion that this assumption of a constant gamma is not valid, which would also be useful information. However, deficiencies could also be explained by a lack of some (past) feedbacks in the models, such as the freshwater-temperature feedback. This could for example (partly) explain why the discharge acceleration is not resolved well. This has already been discussed in the discussion section, but we will add some arguments in the results section as well.

Figure 6: What gamma is used? Median? Mean? Full range? Is the shaded area intermodal spread, are you using a spread in gamma too or just a singular gamma value?

The best-fitting gamma value (lowest RMSE) for each ESM-RF pair. The shaded area is intermodel spread based on the single best-fitting gamma-values for each ESM-RF combination.

L407: Please spell out the argument that applies to ISMIP6 gamma-values.

We will combine this sentence with the previous one, since similar arguments apply.

In general I enjoyed the discussion, and thought that plenty of this type of language could have been used as motivation/context in the introduction. Just something to consider.

Thank you. We will consider using parts of the discussion in the motivation/context in the introduction and result section.

Minor Comments:

L 7-8: Explicitly state all three SSP scenarios used.

We will write out the full descriptions.

L23-24: include references for model (ice sheet models, I assume?) and geological data being referred to.

We will add a table with the model names (both ice sheet models and CMIP6 earth system models)

L25: Moreover, melt of Antarctic land ice

We will add Antarctic.

L32: Using similar methodologies to what? To each other, I assume...?

Yes. We will add this in the text.

L32-35: Can you mention why uncertainties are increasing?

Yes, we will add this if known.

L36: Future projections are always based on modeling (not often).

We will change often to always.

L38 & L42: What do you mean “balanced by data from ESMs”? Please clarify

We mean that because they are coupled their boundary conditions cannot be prescribed but are specified through their interaction with the ocean and atmosphere of ESMs (L38). And that observations act as boundary conditions (L42). We will clarify this in the text.

L39: Is there a citation for claiming ice sheet-ESMs show undesirable/unrealistic trends ?

We suggest to remove these sentences (“because ice sheet models in the short term.”) as discussing this topic in detail might not be directly relevant for our paper.

L46: Used as onebasis for projections...

We will change ‘a’ to ‘one’.

L56: In that study, the temperature...

We will add the comma.

L75-76: What does it mean to arrive at an estimate of future mass loss that is consistent with observed mass loss over the past four decades?

That the physics is consistent (in this case the computation of basal melt).

L86: Reference Lambert et al (2021) here would be appropriate.

Indeed, we will add this reference to the sentence.

L91: Please state explicitly what Levermann et al calibrates to.

Levermann et al. does not calibrate the basal melt parameterization, but uses melt sensitivities derived from observations of Jenkins (1991) and Payne et al (2007), resulting in a range of 7-16 m/yr/K. The sensitivity is drawn randomly and uniformly from this interval.

L97: What does “delayed ice sheet response” mean here?

The delayed response to basal melt.

Table 1: What level does “subsurface ocean temperature” refer to here? I assume it is the mean depth of the grounding line (table 2), but please be clear.

It is the mean depth of the ice shelf underside since the LARMIP-2 linear response functions are forced with a spatially constant basal melt anomaly. We will specify this in the text.

L103: The ocean temperature time series...?

Yes, we will specify this.

L107: Remove text: “For the quadratic melt parameterization”.

OK. We mentioned this because it is only relevant for the quadratic relation, but it does not matter for the linear one so we will remove it from the sentence.

L123 & 142: Do any CMIP6 models represent cavities?

Not that we know of.

L126: Table 3 is referenced, but perhaps prematurely. I suggest T3 and T4 are swapped in order.

Yes, that makes sense. We will see if we can add the reference to Table 3 at a later moment.

L138: “See Table 4...” already noted earlier.

OK, we will remove this sentence.

Figure 5: Increase font size here please.

OK.

L227: “as specified on top” à “as specified in the titles”

OK, we will change this in the text.

L281-285: This text seems to be repeated again later on from L290-295. Please give this another proofread to make sure there is no longer repetition and the flow works well.

We will correct this repetition.

L326: “The differences with ISMIP6 and LARMIP-2...” I think you mean difference between this study’s results and these two studies, but this phrasing leaves this ambiguous.

Indeed, we will rephrase this sentence.

Fig11&12: Label the x-axis please. And possibly increase fontsize.

We will add a label to the x-axis and increase the fontsize.

L345: "is computed using forcing from state-of-the-art..."

We will change 'with' to 'using forcing from'.

L346: "CMIP6 applied to linear response functions..."

We will change this.

L364: Is there an appropriate reference for the comments made about ESM performance?

We will add a reference here.

L373: By climate-state dependent do you mean time-evolving?

Yes, depending on e.g. the subsurface ocean temperature.

Table 7: Please state what the percentages refer to. Also add units.

Yes, we will add an explanation to the percentile numbers. Units are specified in the caption.

L376: Please elaborate on the processes you refer to here.

We will elaborate on the processes.

L384: The better prediction of future mass loss you do achieve is for less physically defensible reasons, no?

If the gamma value would be a constant, like we assumed for our study, it would indeed be less physically defensible. However, if it turns out that gamma is time-evolving it could be defensible. For example, if the cold ocean sectors become warmer and their temperature closer to the Amundsen region it could be expected that Antarctic-wide or calibration on the Amundsen region works better than regional calibration on those specific sectors. We will elaborate our discussion on this point.

L386: Elaborate on the advantages and disadvantages.

We will elaborate on this.

L395: Therefore, it is interesting...

We will add a comma.

L434-435: Please state what you mean by 'highest' and 'lowest' methods.

The methods that give the highest and lowest projections. We will specify this in the text.

FigA1 & A2: Font sizes of tick labels and axis labels could be increased for better readability.

We will increase the font sizes.

In general, section headings could be more clear. For example, "Magnitude and Rate" and "Best Estimate" could be more specific to help the reader understand what the section is about.

We will make the headings more specific.