The referees' comments are reproduced in black hereafter, and our responses are shown in blue.

Anonymous Referee #2

In the manuscript parameterisations and their calibration to generate basal melt rate forcing for the ISMIP6 experiments are presented. In a first step, a present-day climatology of the ocean is generated from different datasets and extrapolated by horizontal filling underneath ice-shelf and into currently ice-covered regions. The derived, local temperature and salinity then inform the parameterisations. The authors present two different parameterisations, both have a quadratic dependency on thermal forcing, one based on the local thermal forcing and one on a mixture of local and basin-wide averaged thermal forcing. Furthermore, a procedure to tune the parameters including an assessment of their uncertainties is presented. Tuning parameters are a pre-factor, which is constant for the entire continent, and basin-wide temperature corrections δT_b for 16 different basins b. The first tuning approach uses the Antarctic-wide basal mass flux for tuning of the parameters and the second approach observed melt rates close to Pine Island Glacier's grounding line. While present-day melt rates are, by construction, similar for both sets of parameters, melt rate sensitivities are very different and hence the projected melt rates differ by an order of magnitude.
In general, this manuscript is well written and presents a novel and comprehensive approach to systematic tuning of parameterisations including uncertainty ranges for parameters. It clearly indicates problems related to the tuning and potential future developments.

> We thank the referee for this positive review.

Major comments

(1) The aim of this work is to provide a suitable basal-melt rate parameterisation and oceanic input for ISMIP6 projections. Since the two calibration methods you present yield largely different results, it would be useful to identify upper and lower limits for basal melt rate sensitivities and discuss how your parameterisations fit into that range. In particular, do the projected changes in basal melt rates for the (95th percentile of the) PIGL parameterisation represents an upper limit and the (5th percentile of the) AntMean parameterisation a lower limit given current observations and modelling studies? How does the slope-dependent parameterisation fit in there?

> See our response to the 2nd general comment of Referee #1. There are very few observational data to assess the sensitivity of melt rates to changing temperature; the interannual observations at Dotson and Thwaites have been used here to evaluate the γ_0 coefficient (Fig. 8). This suggests that the PIGL method is more realistic than MeanAnt, and that the 95th percentile of PIGL's γ_0 cannot be discarded. However, comparisons to FESOM simulations (Fig. 10) or to the +0.5°C perturbation of Seroussi et al. (2017) suggest that the MeanAnt method may be more realistic, and the 5th percentile of MeanAnt's γ_0 cannot be discarded. So without obtaining more interannual observations or more model simulations (keeping in mind concerns with biases), it is difficult to narrow the range of uncertainty on γ_0.

The slope-dependent parameterization is not part of the standard ISMIP6 protocol, mostly because of calendar constraints. However, we believe that this is a promising way forward with this kind of very simple parameterization, which is why we included it in the Discussion section. As already mentioned in the

Discussion "Introducing the slope dependency also strongly reduces differences between the two calibration methods (Tab. 3), thereby reducing uncertainty in projected melt rates".

(2) More details on the PIGL tuning are required, see specific comments below.

> See our responses below.

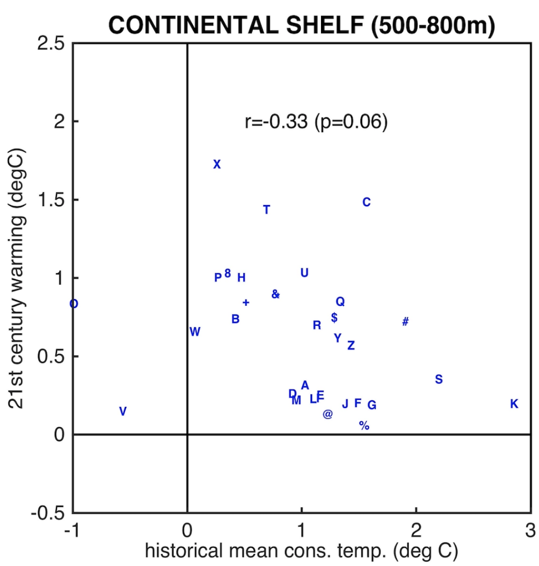
Specific comments

- page 3, line 32: I do not understand this sentence, since you only focus on basal melt rate forcing here?

> Agreed, this has been replaced with "in this paper, we focus on basal melting".

- page 5, line 19: Can such a switch be simulated in CMIP models without representing ice-shelf cavities and the continental shelf?

> Let's have a look at the conservative temperature averaged over 500-800m, averaged in a box representing the Amundsen Sea continental shelf, and for 33 CMIP5 models. Below is a scatter plot of the rcp85 warming projection (future minus present) as a function of the present-day average temperature; each character represents an individual CMIP5 model. We can see that such switches occur (e.g. models "X" and "T"), although this does not explain much of the cross-model variance. Actually, this does not necessary require an ice-shelf, it can also be related to sea-ice: if there is a lot of sea ice at present day, future warming can produce large changes in sea-ice formation and eventually stop deep convection; conversely, if there is already no much sea-ice and associated convection at present day, it will be more difficult to produce strong changes. While this is interesting, we have not added anything about this in the manuscript to keep the focus on the ISMIP6 protocol.



- page 6, line 28: Please explain for readers not familiar with WOA the terms `statistical mean' and `objectively analyzed mean'.

> This has been modified as:

> "We use the "statistical mean" (*average of all available values at each standard depth level in each 1° square*), rather than the "objectively analyzed mean" (*interpolation from irregularly spaced locations to a fixed grid*) values for WOA18p and EN4".

- page 6, line 30: I'm not sure I understand this, you `bin' the WOA18p data onto the same grid?

> Sorry for the confusion here. The WOA18p data are *already* binned on the WOA18p grid, so that's what we have to work with. We regrid them to the standard ISMIP6 grid before combining them with the other data sets because binning, then regridding is a lossy process that we want to try to minimize. We have modified the relevant text as follows:

> "The WOA18p data have already been binned by the creators of the dataset on the native WOA18p grid (0.25° bins in latitude and longitude). We interpolate these data (first, conservatively in the vertical and then bilinearly in the horizontal) to the ISMIP6 standard grid. Since the EN4 and MEOP data are provided at their original locations without binning, we are able to bin-average these datasets directly on the standard grid".

- page 8, line 1: Why do you use different procedures for the datasets? In particular, why do you chose to vertically interpolate the WOA18p data and not the other datasets since this might introduce vertical variations if the other data has vertical gaps?

> As explained in our previous response, the WOA18p data are already binned to a horizontal and vertical grid -- we don't have access to the original point data -- so we interpolate them to the ISMIP6 grid as best we can. Then, we combine with other datasets where we *do* have the point data on the ISMIP6 grid. Again, sorry for the misleading text that gave the impression we were doing the binning of WOA18p, rather than the creators of the dataset.

- page 8, line 22: How are salinities affected?

> The thermal forcing is not a strong function of salinity. Indeed, from 34 to 35 g/kg, the freezing point decreases by only 0.06°C. We nonetheless decided to process salinity in the same way as temperature to provide a clean dataset.

- section 3.2: Do I understand correctly that the only way the compiled observational dataset is vertically extrapolated is by filling the deeper levels with copies of the lowest available data point?

> yes, this is correct.

- page 9, line 10: `ocean model data and the climatology'

> Thank you, it is now “ocean model *and observational* data”.

- page 9, line 12: Is it correct that the open ocean is not separated by the basin boundaries as shown in Fig. 2 for the interpolation (otherwise ocean regions, e.g., in the Weddell Sea, would be very small and might not contain data)? And for the ice-covered regions that the values at the boundaries are used?

> yes, this is correct.

- Figure 2, `shading' should be `colors'. Please add more explanation to the legend, especially make clear what regions are actually used in this study.

> The figure caption has been clarified.

- page 9: Please add figures showing your final datasets for an exemplary depth and along the current ice-shelf draft including basins boundaries and basin averaged values.

> The thermal forcing from the final dataset is already shown along the current ice-shelf drafts in Fig. 5, and Fig. 1c already shows temperatures of the combined dataset before extrapolation. We consider this sufficient, and have not added a figure. We have nonetheless mentioned in page 9 that “The resulting thermal forcing along the current ice-shelf drafts is shown in section 4”.

- page 10, line 5: Please give an example here.

> We have added the example of Pine Island and Thwaites.

- page 13, line 8, page 15, line 5: Do I understand correctly that you fit melt rates in units of average m/a for each region, not in Gt/a? How different are results depending on the choice of average or aggregated melt rates?

> Our description was not clear, and this has been clarified. We actually fit mass loss rates (in Gt/a), not average melt rates (in m/a). This would be an important distinction if the ice shelves in BEDMAP2 (Fretwell et al. 2013; used in our study) had a different area compared to ice shelves in Rignot et al. (2013) and Depoorter et al. (2013). While there may be small differences, all these studies use observation representative of the 2000s, and should be consistent. Rignot et al. (2013) even used BEDMAP2 to map the ice thickness for a majority of ice shelves where Operation Ice Bridge did not make measurements.

- page 13, line 30: More detail is required here. In particular, do I understand correctly that you use the highest 10 melt rates from the spatial pattern? Do you fit such that melt rates in the respective location are similar, or that melt rates in the area close to the grounding line have a similar melt rate?

> This has been clarified: “we estimate gamma0 by randomly sampling one of the 10 grid points with the highest melt rates (with equal probability) and associated thermal forcing (normally distributed error) underneath Pine Island ice shelf. This is repeated 10^5 times to obtain the median, 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles of gamma0”.

- page 15, line 2: Wouldn't it help to better constrain the melt sensitivity in PIGL by using the temporal variation from Figure 8 for calibration?

> Yes, the analysis shown in Fig. 8 came after the ISMIP6 protocol design, but this would have been an option, although it does not guarantee that melt rates are high enough near grounding lines (which was the motivation for the PIGL method).

- Figure 3 and 4: Please add explanation to the legend.

> We have added a brief explanation in these two figure captions.

- page 15, line 6: δT should represent changes of water masses being transported into the cavities as well as uncertainties in observations. Since the first would only act to decrease temperatures at depth, shouldn't a decrease in temperatures be favored over an increase (i.e., not a normal distribution be assumed)?

> As written in our manuscript, δT accounts "for biases in observational products, ocean property changes from the continental shelf to the ice shelf base (not accounted for in the aforementioned extrapolation), tidal effects and other missing physics". The PIGL calibration does create negative δT in most sectors. With the MeanAnt calibration, we first adjust γ_0 to get the correct melt rate for the entire ice sheet, then δT to get the observed melt rate in each sector. So by construction, there must be regions with positive δT and regions with negative δT. This would be one more argument to prefer PIGL over MeanAnt, but as δT accounts for many imperfections of our parameterization, we prefer not to over-interpret this. We have simply added this sentence when we describe the δT distributions:

> "We note that MeanAnt δT values are positive and negative by construction, while PIGL δT values are negative, as expected if this correction represents changes in water mass properties along the ice draft (keeping in mind that it also likely accounts for missing physics)."

- page 18, line 7: If you add $\delta T = 1.07K$, how large are temperatures in the Amundsen Sea then for present day? And how do they compare to observations in that region?

> It is difficult to answer this question as our observational gridded dataset is the best estimate that we had for the climatological annual mean temperature. So in that sense, $\delta T = 1.07$ K is 1.07 K too warm compared to observational estimates. But again, δT represents more than a correction of observed temperatures, it also accounts for missing physics, which is why we did not try to overinterpret the meaning of δT values.

- page 18, line 26: Do I understand correctly, that you retune parameters here for the Amundsen region? Since changing γ or δT affects the basal melt rate sensitivity, the comparison to observations is not very meaningful for the other parameter choices. Also, do you apply the observed T, S profiles as anomalies to your climatology? Such a procedure might be better to assess the parameterisations, since the melt sensitivity of your parameterisations depends also on temperature.

> We do not retune γ, which is the coefficient mostly responsible for the melt sensitivity to ocean warming, and which is actually the coefficient that we want to evaluate here. The δT value determined previously in this paper was calibrated to correct the sector-averaged thermal forcing in the non-local parameterization (so the entire Amundsen sector here). To compare to interannual observational data, we only have either Dotson or Pine Island (on different years), and we cannot calculate Amundsen-averaged thermal forcing. As such, keeping the standard δT value for the Amundsen sector does not make sense, and we choose to re-calibrate δT to match the mean observational melt rate and focus on the interannual variability. This has been clarified in Fig. 8's caption and in the associated text.

- Figure 6: Please add also uncertainty ranges based on different values of γ, similar figures for the local parameterisations as well as for the slope-dependent ones.

> Fig. 6 already contains many lines, and adding percentiles would make it difficult to read. Further, this figure is used as an illustration to explain the behaviour of our two calibration methods, and we do not think that adding percentiles would better illustrate our methodology. We have therefore kept Fig. 6 as it was.

- page 23,24: One explanation for the discrepancy could also be that with your parameterisations all ice shelves have the same melt rate sensitivity (modulated by their respective temperature), however, FRIS might have a lower sensitivity than PIG, not only due to the initial temperature, but also due to its geometric properties (see Holland et al. (2008) testing this for an idealized geometry).

> We agree, and this is the reason why we introduced the slope dependency in the Discussion section.

- page 24, line 26: It might be key to include the basal slope in parameterisation. How does this affect future changes in BMR as shown in Figure 10?

> We have added the equivalent of Fig. 10 but for the slope-dependent version (see Fig. 12 below). As already mentioned and shown in Tab. 3, the difference between PIGL and MeanAnt is reduced when the slope dependency is introduced, and the projected basal mass loss is generally much lower than with the standard ISMIP6 method.

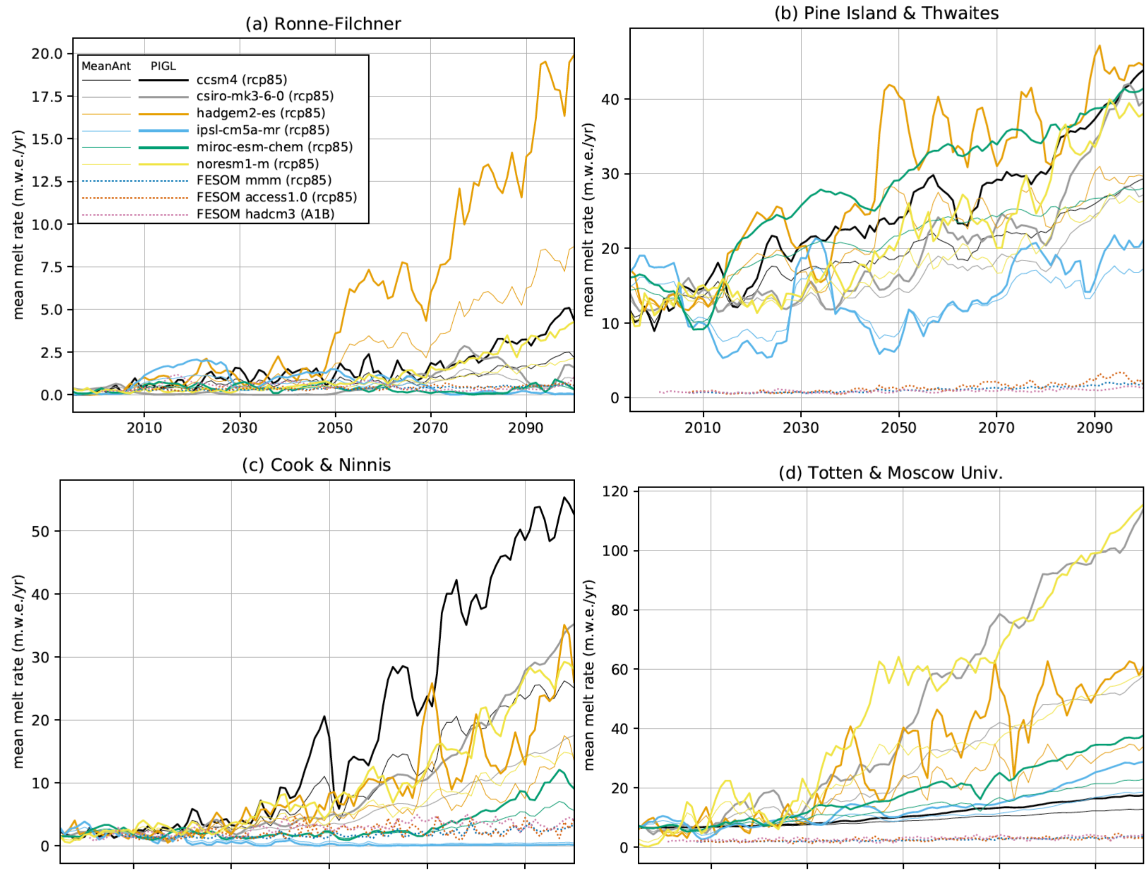


Figure 12. Same as Fig. 10, but using a slope dependency (eq. 1 multiplied by $\sin\theta$).