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

Anonymous Referee #1

The authors describe the protocol that will be used to compute melt rates at the base of ice shelves in ice sheet models driven by output of global climate models in the framework of the Ice Sheet Model Intercomparison Project for CMIP6 (ISMIP6). The global climate models included in CMIP6 have generally a too coarse resolution and do not simulate explicitly the circulation and fluxes in the ice shelves cavities. It is thus important that all the groups participating in ISMIP6 use a similar protocol to derive the melt rates from those global model results so that the origin of the differences in their results can be more easily investigated.
The manuscript is very clear. It describes precisely and justifies well the choices performed in the approach. It also proposes several options to sample the uncertainties associated to the computation of the fluxes. This will be very helpful in the development of the intercomparison project. Consequently, I just have minor suggestions for improvements.

> We thank the referee for this positive review.

I have first two small general points

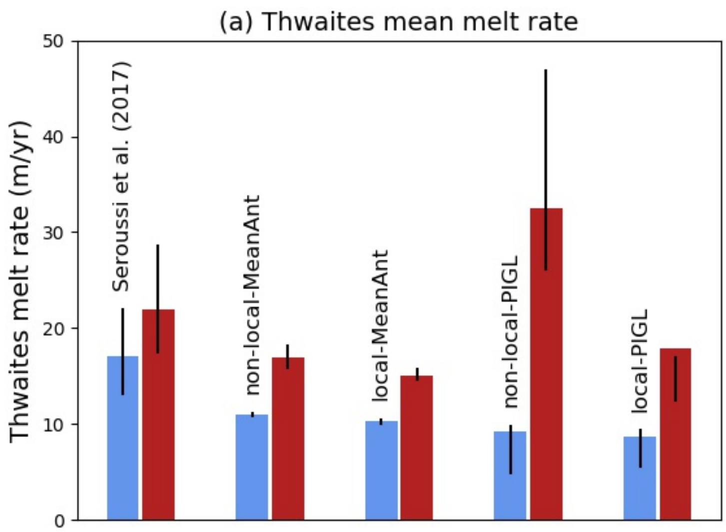
1/ If I understand well, despite a relatively sophisticated approach to obtain the melt rates for present-day conditions, the warming signal simulated on the continental shelves by global climate models for future conditions is transferred without modifications into the cavities. The warming is also homogenous in the cavities, because of the extrapolation applied. If this is the case, maybe it is good to write it explicitly, for instance in the final section, to avoid misinterpretations.

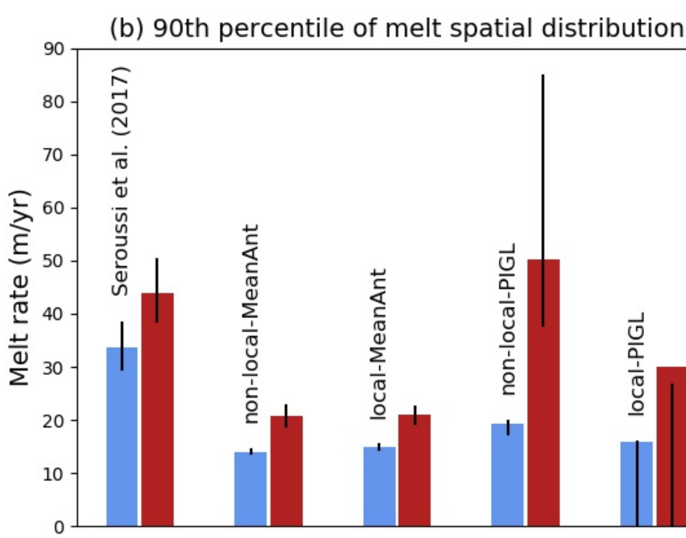
> The warming signal simulated by the CMIP models is indeed transferred without modifications into the cavities, but it is not homogenous, it keeps the vertical warming profile of CMIP models and possibly regional patterns, which could produce east-west warming gradients in the largest cavities. This has been clarified as follows: "This method keeps the same vertical structure inside ice-shelf cavities as in the "ambient ocean" of observations or CMIP5 anomalies, which omits several physical processes".

2/ The computed fluxes can vary by one order of magnitude between the different parametrizations. This is a very large uncertainty and I guess this would have a major impact on ice sheet model results. I know this point is not the topic of this paper but more information on the uncertainties of those fluxes would be very helpful. Results of simulations with FESOM are suggested as a benchmark but I was wondering if other results could be included too to have a broader discussion of this important point.

> First of all, we would like to remind the reviewer that the entire paper is about the uncertainty on parameterized melt rates. This is our motivation to calculate the percentiles and to explore two very different methods to calibrate our parameters. Having said that, our range of uncertainty is indeed huge, and it is legitimate to try to reduce it. As pointed out in the review of Asay-Davis et al. (2017), so far, "few projections have been performed with ocean models including ice-shelf cavities". This statement remains valid until now, so it is difficult to use model projections to evaluate our parameterization and possibly

reduce the range of acceptable parameter values. In addition to FESOM (the only CMIP-based projection so far), we can use the study by Seroussi et al. (2017), focused on Thwaites, and in which the ocean initial and lateral-boundary conditions were uniformly warmed by +0.5°C. Applying a similar perturbation of 0.5°C to our climatology, we obtain the following results (present-day in blue vs +0.5°C in red):





However, it is again not easy to conclude on which parameterization performs better. First of all, the present-day parameterized values are underestimated at Thwaites compared to Seroussi et al. (and to observations). MeanAnt seems in better agreement with Seroussi et al. (2017) in terms of relative increase for both average and high-end values, but “warm” (+0.5°C) non-local-PIGL melt rates are quite close to “warm” melt rates in Seroussi et al. (2017). This figure and its description have been added at the end of section 5.

Specific points

1. Page 4, lines 4-5. I do not understand what is meant by 'coupled ice sheet ocean models are not ready to be used with CMIP boundary conditions'. Is the problem that ice sheet models are not coupled to ocean models for the majority of ISMIP6 models or that those models cannot be used on the spatial-timescales of interest? I guess that, for coupled ice sheet ocean models, a protocol can also be defined to drive them by CMIP boundary conditions (but it is out of the scope of ISMIP6?) – see also page 4, line 20.

> An option for using ocean—ice-sheet coupled models was proposed in the early ISMIP6 protocol, but none of the groups was ready to run such simulations at the scale of Antarctica. So far, papers published on ocean—ice-sheet coupled models are limited to idealized configurations or regional configurations (see references in Asay-Davis et al. 2017 and Favier et al. 2019). Although some groups have started to run such global coupled models, switching to the pan-Antarctic scale and circum-polar ocean – while keeping a realistic present-day state – remains challenging. To make things clearer, we have specified "are not ready to be used with CMIP boundary conditions *at the pan-Antarctic scale*".

2. Page 8, line 17. The authors mention that the errors due to sampling, interpolation/ extrapolation are likely much larger than those due to the temporal bias. However, large interannual variability and trends have been observed in several coastal regions around Antarctica. That would thus be helpful to quantify the bias associated with the choice of the different periods, maybe using some of the data in the regions with the best coverage or using oceanic reanalyses (which have their own biases too).

> Except near a very few well-observed ice shelves (e.g. Pine Island, Dotson), there are not enough data to properly estimate the interannual variability or trends in most coastal areas. This is particularly true over the very narrow continental shelf in East Antarctica, where the inclusion of elephant-seal data decreases the temperature by more than 1°C (Fig. 1c,d). To our knowledge, there is no evidence that ocean temperature could differ by such a large amount between 1995-2017 (WOA+EN4 data) and 2004-2018 (MEOP).

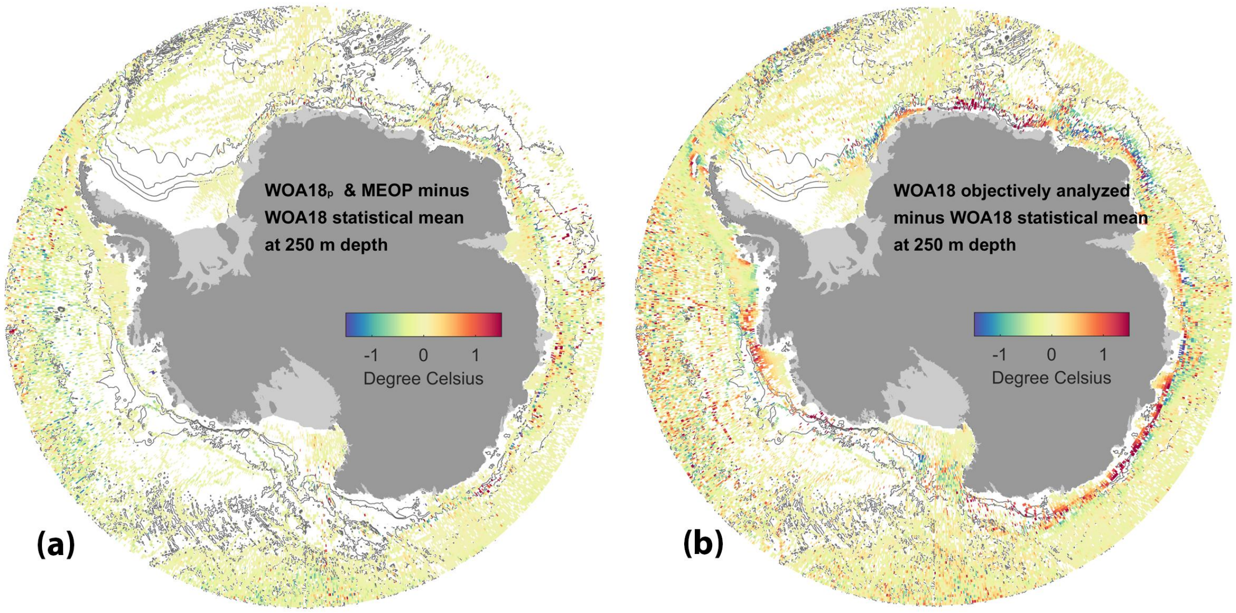
About the inconsistency between the ocean data and melt rates time windows (2003-2008), we are aware about work in progress to provide interannual ice-shelf melt rates, but so far there is no interannual ice-shelf melt rate estimates published for a large majority of Antarctic ice shelves. In the case of Dotson, the average melt rate varies from 41.7 Gt/yr over 2000-2008 to 40.6 Gt/yr over 2009-2016 (the interannual peak is in 2009; Jenkins et al. 2018), so the time inconsistency between the ocean and melting datasets is likely unimportant in this case. To clarify this, we have added:

> "Estimates of interannual variability of ocean properties exist only for a handful of coastal regions around Antarctica. Based on Jenkins et al. (2018), we believe that the uncertainties due to temporal variability between these two time periods are smaller than those due to the spatial interpolation/extrapolation."

About the reviewer's suggestion to use an ocean reanalysis, there is no such reanalysis based on a model that represents ice-shelf cavities. Furthermore, existing ocean reanalyses are mostly constrained by summer observations near the ice-sheet margins so data assimilation is unlikely to compensate the absence of ice shelves. We therefore decided not to rely on such reanalyses.

3. Page 8, line 24. The dataset proposed is different from the latest release of the World Ocean Atlas that use similar observations as input. I understand the reasons for this choice but, as many scientists will likely use this version of the World Ocean Atlas, it would be needed to highlight the main differences, for instance by showing a few maps in the supplementary material.

> Our merged dataset (WOA18p+EN4+MEOP) actually gives very similar temperatures as the latest version of WOA18's statistical mean, as shown in panel (a) below:



However, there are gaps in WOA18's statistical analysis, which make it impractical to constrain ice-shelf melting. It could be tempting to use WOA18's objective analysis instead of our merged dataset, but this objective analysis does not seem able to account for the strong horizontal gradients over the very narrow continental shelf of East Antarctica (see panel b). This suggests that our merged dataset may be more adequate for providing continental-shelf properties for these regions. This is now detailed in section 3 and the figure has been included as supplementary figure S1.

4. Section 4.2. Is 'thermal forcing' defined ?

> yes, it is defined in the Approach section as "the difference between the in-situ far-field ocean temperature (not modified by the buoyant plume) and the in-situ freezing temperature"

5. Page 13, line 18. The authors mention that they take samples in the melt rate and the error in the thermal forcing, using normal distributions. I may miss something but I think they take samples in the distribution of melt rate and thermal forcing (not in the error of thermal forcing). Same for Figure 3.

> The thermal forcing is different at each grid point, but we randomly add a uniform error in each sector. We do sample this error in a normal distribution, not the thermal forcing itself. We have not modified these sentences.

6. Page 13, line 30. Gamma0 is estimated by sampling the 10 highest melt rates. Would using all the melt rates for the Pine Island ice shelf lead to values that are closer to the ones obtained for the MeanAnt method?

> We have not examined this option as our aim is either to be representative of the entire ice sheet, or to get high melt rates near Pine Island's grounding line.

7. Page 15, line 9. If the temperature correction deltaT accounts for 'ocean property changes from the continental shelf to the ice shelf base' (page 12, line 12), I would assume that deltaT should be negative in most regions. Are the positive values obtained for the MeanAnt in many regions a sign that deltaT is rather compensating for a too weak exchange coefficient?

> As written in our manuscript, deltaT 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 deltaT in most sectors. With the MeanAnt calibration, we first adjust gamma0 to get the correct melt rate for the entire ice sheet, then deltaT to get the observed melt rate in each sector. So by construction, there must be regions with positive deltaT and regions with negative deltaT. This would be one more argument to prefer PIGL over MeanAnt, but as deltaT accounts for many imperfections of our parameterization, we prefer not to over-interpret this. We have simply added this sentence when we describe the deltaT distributions:

> "We note that MeanAnt deltaT values are positive and negative by construction, while PIGL deltaT 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)."

8. Page 18, line 13. Is the underestimation of the melting at surface in the PIGL method a consequence of using constant deltaT on the vertical while the correction may be smaller closer to the surface?

> It depends how deltaT is interpreted. If it is seen as accounting for the water mass transformation, it should be higher in the upper layers. If it is seen as accounting for biases in the observational datasets, it should probably be higher in the thermocline. But again, many things are hidden behind deltaT, and we don't want to over-interpret this, so we have not added any comment about this.

9. Figure 6. The last but one and last but two sentences of the caption are repetitions of the second line.

> Thank you, this has been corrected.

10. Page 22, line 9. 'estimated' instead of 'reconstructed'?

> This has been modified as suggested.

11. Page 24, line 14. I would suggest 'selected' instead of 'identified' as the choice is mainly based on past results, not on new analyses performed in the manuscript.

> This has been modified as suggested.

12. Page 24. It is not clear from the discussion if the parametrization with a slope dependency is suggested or not as an option for ISMIP6.

> We have specified “While we encourage testing this parameterization, it is not part of the ISMIP6 standard protocol”.