Response to anonymous referee RC2 on ”Assessment of Sub-Shelf Melting Parameterisations Using the Ocean-Ice Sheet Coupled Model NEMO(v3.6)-Elmer/Ice(v8.3)” by Favier et al.

April 26, 2019

The authors present the first results from a newly developed offline coupling between the ice model Elmer/Ice and ocean model NEMO. They use the standardized MISMIP/MISOMIP intercomparison framework to assess the impact of ocean melting on ice dynamics for a range of (idealized) future ocean conditions, demonstrating the capabilities of their model. They then use these (sophisticated) coupled ice-ocean results as a benchmark to assess the performance of a range of (simplified) ice-shelf melting parameterization that are commonly used in a stand-alone ice models. This is a solid piece of work, which recognises the importance and complexities of simulating ice-ocean interactions in Antarctic ice shelf cavities, and the need for careful model development and validation. The manuscript presents the first comparison between results from a coupled ice-ocean model and a comprehensive set of commonly used meltrate parameterizations, making this a timely and valuable reference for further research. The model development sections and appendices contain sufficient amounts of (technical) detail, the methods are sound and in line with previously published developments, and the experiments are explained in a comprehensive way. The authors demonstrate the capabilities of their new coupled ice-ocean model, and pave the way for further research. I highly recommend this work for publication in GMD, although I would like the suggest a few points for further clarification/improvement.
We are very happy with this positive comment

1) For the calibration of the melt calculations you use the WARM ocean conditions, but adopt different criteria to fix the exchange coefficients in the coupled (¡melt¿=30m/yr below 300m) and parameterized (¡melt¿=8.5mr/yr over the entire ice shelf) melt calculations. This -somewhat ad hoc- choice of calibration subsequently has large effects on the results (comparing Figure 7 and H1) and it therefore seems rather important. Could you please clarify why you do not adopt a universal calibration for all (parameterized and coupled) methods, e.g. ¡melt¿=8.5mr/yr over the entire ice shelf and what would inform such a calibration? For a universal calibration, the initial differences in SLC between different melt calculations are fully attributed to differences in the spatial distribution of melt, at least in the WARM scenario, rather than spurious effects due to the calibration method.
The target of 30 m/a of melting below 300 m depth using the WARM ocean conditions is part of the MISOMIP protocol. We simply used this protocol because the WARM ocean conditions lead to a quick relaxation of the ocean, which allows to find the calibration without cumbersome calculations. However, when forced by the WARM ocean conditions, all the sub-shelf melt parameterisations produce substantial melting above 300 m depth, as opposed to ocean models that produce small melt rates. We thus chose to take the average of melting without depth limitation, but in the end, we have the same melting average of 8.5 m/a (+/- 1) for all parameterised and coupled simulations. All this was explained, and has now been clarified in response to reviewer 1, in Section 3.2.

2) On a similar note, I wasn't expecting the spread in total melting between different melt parameterizations (Figure 5) to remain more-or-less constant through time. I was expecting a divergence, with most parameterizations doing progressively 'worse' over time, compared to coupled simulations. Do you have any insights in to why that is?
We are not sure about what the reviewer means here. In terms of sea-level contribution, it is quite clear from revised Fig.7-8, that the distance between the parameterizations and the coupled model and the distance between different parameterizations increase in time. In terms of melt rates it is different because melt rates remain highly constrained by the temperature scenarios.

I understand that the initial spread is defined by how sensitive the melt parameterizations are to the forcing for a given geometry, with the constraint that total melting is the same for all parameterization in the WARM (calibration) scenario. To disentangle the geometrical feedbacks and the initial sensitivity to ocean forcing, I

think it would be instructive to see a plot similar to the panels in Figure 5, but for the WARM scenario. By nature of your calibration, all simulations should agree on the total melting at time 0, and (perhaps) divert for t¿0 due to ice-ocean feedbacks. By imposing a common starting point, you will be able to unambiguously identify which parameterizations are 'close to' or 'far away' from the coupled simulations for that particular forcing.
This is an interesting suggestion. Actually, there are two reasons why melting will divert at t¿0: (1) the ice/ocean feedback, i.e. how the change in ice draft depth modifies melt rates, and (2) the various melt parameterizations all have different sensitivities to warming, and as restoring temperature evolves, melt rates will diverge even without ice/ocean feedback. So investigating this would also probably require simulations with fixed cavities. This is interesting but our approach is to evaluate melting through the ice sheet response because we don't want to decide a priori where melt pattern has to be accurate. Overall, this suggestion seems to be quite disconnected to our approach, and to keep the paper relatively short and clear, we have decided not to follow the reviewer's suggestion.

3) About the RMSD criterion (Figure 7) that you use to assess the performance of parameterized melt rates compared to coupled ice-ocean simulations, I wonder if this criterion might be too simple and perhaps even misleading. As you explain, the spatial distribution of melt rates for a given geometry will, to a large degree, control the dynamic response of the glacier. However, this is not clearly captured by the RMSD criterion. In particular, you could identify parameterizations with a low RMSD as 'performing well', but the spatial distribution of melt could be totally wrong, and therfore the RMSD is low for the wrong reasons. As a result, this parameterization might not be suitable for other (more complex) geometries. Perhaps a simple RMSD criterion should be supplemented by a measure of how well the spatial melt distribution compares to the coupled scenario, making the assessment more robust. To achieve this, it would be instructive to see Figure 4 but after 50 years of simulations.
This is clear that the distribution of melting from parameterised simulations and coupled simulations are quite different, which is actually discussed in the paper ("All parameterisations yield too large melt rates in thin ice areas and too small melt rates ..."). However, we do not agree that the RMSD is misleading even though it is not ideal. Assessing the spatial distribution of melt would also be not ideal because it is difficult to identify which part of the melting pattern is important for the ice dynamics (e.g. see "tele-butressing" in Reese et al. 2018b). For this reason, we have decided to evaluate the parameterizations through an ice-sheet simulation rather than through the melting characteristics.

Finally, here is a list of some smaller comments, typos etc.
p1, l15 shelt → shelf
Changed

p4, l19 'floating nodes only': please clarify if you impose melt for nodes in partially grounded elements
Even though the melting is 0 at the last grounded point and 1 at the first floating point, the finite element method will give an averaged solution for the element containing these two (or three here) points. Thus the last grounded node may be affected by melting if the thickness is sufficiently modified to make it float. Otherwise, the last grounded node will stay grounded. As the treatment of melting is the same by Elmer/Ice in both parameterised and coupled simulation, we have not added those technical points but now simply say "Melting is applied to floating nodes but not to grounded nodes, meaning that the first floating element (partially or not) may be affected by melting."

p5, l1 different ocean models will use different ways to parameterize a 'boundary layer' and calculate u_{TBL}. Perhaps you could be more specific here about u_{TBL}, unless this methodology has been published elsewhere?
As described by Mathiot et al. (2017), u_{TBL} is averaged over a constant thickness, assumed to represent the thickness of the TBL. We added the reference in the text.

p5, l17 do you always average over the entire coupling period, or do you use the final week/month/... of that period?
Melt rates are always averaged over the entire coupling period in order to conserve mass as much as possible, which we added in the paper. By contrast, the final ice draft of Elmer/Ice's coupling period is always sent to NEMO.

p5, l22 'too thin to be captured by NEMO': could you be more specific please? Do you impose a minimum water column thickness? Do you adjust your geometry to allow for this etc?
We impose a minimum water column thickness of 20m, which allows NEMO to have a minimum of two vertical cells under the partial-cells condition, which we added to the paper. If the new water column opened by Elmer/Ice is thinner than 20m, it is not simulated by NEMO, i.e. zero melt rate is sent back to Elmer/Ice.

p6, l10-12 You say you have shown convergence with coupling timestep, but in Figure A1 all results fall on top of each other. To fully show convergence, you need to present results from eg 48 and 24 months, and show that they 'converge' to the solution for 12, . . . months. On a similar note, you present results for 1 particular scenario. In the caption of Figure A1 you say this is the COM-Ocean1 experiment, but I'm not sure if you mean Ocean1r or Ocean1ra? As the total melt goes down over time, I'm assuming this is a 1ra scenario with cold forcing? This could be important, as convergence might be harder to achieve in a warm ocean scenario?
This is the retreat IceOcean1r scenario (warm), which has been specified. The initial strong melt rate is related to the initialization method in MISOMIP, but after 5years, melt increases over 95 years. We do not see the point of showing results for 24-month and 48-month coupling periods because we never use such coupling period. Of course, if we make the coupling period very long, it will make a difference, but the point is more that taking any coupling period between 1-month and 12-month does not affect the result. Also, nowhere in the text we wrote that we have convergence of the solution when decreasing the coupling period. We simply mentioned that there is no sensitivity to the coupling period in the 1-12 months range.

p6, l17 As the calibration procedure is so important, perhaps you could provide a 1sentence summary of the protocol from (Asay-Davis et al., 2016)?
We have added the sentence "more details of the protocol relevant to our study are given in Sec. 2.2.3, and the protocol is fully described in Asay-Davis, 2016, Sec. 3.2.1" at the end of Section 2.2.3, and in Section 3.2.1 we have added the sentence "The remaining steps of our calibration, described here below, differ from the ISOMIP+ protocol and are specific to our study" so it is clear what is part of the ISOMIP+ protocol and what is specific to our calibration. The relevant details of the protocol are still described at the previous sentence.

p6, Table1 Of course this list of experiments is not/cannot be exhaustive, but it would be nice to have some more motivation for the choice of these specific experiments. Do they somehow represent end-members that allow you to put generous 'error bars' on the coupled results? For example, why are you interested in constant tidal velocities, knowing that Pine Island is not subjected to strong tidal currents? What defines T_{CPL} and why is it different for the different scenarios, given that you claim 'convergence' below 12 months. Why do you not consider changes in the drag coefficient, etc?
The coupling frequency is different but has no effect. We do not change the drag coefficient C_d because this is somewhat equivalent to changing Γ_T which is our tuning coefficient (see Jourdain et al. 2017). The rest is based on our experience of what is likely to affect melt rates: vertical mixing, vertical resolution, and horizontal resolution. Tidal velocities directly affect heat exchange and were therefore thought to be important. We agree that tides are unimportant for Pine Island (e.g. Jourdain et al. 2018), but this idealized cavity represents a range of small cavities in warm and cold environments, not only Pine Island, although it was initially chosen to mimic Pine Island in the MISOMIP protocol.

p8, l10 'inspired BY Jourdain et al. (2017) WHO ...
Changed

p10, l9 'to the local pressure' → 'ON the local pressure'
Changed

p12, l25 do you mean figure 2?
Yes, changed

p14 Perhaps you can point out that you do not impose any 'density compensation' by changing the salinity for each temperature profile, so the density profile will be different in each scenario.
We have added the following sentence where the scenrions are described: "Note that none of the temperature profiles account for a salinity compensation (as opposed to the MISOMIP protocol), so the density profile is different in each scenario"

Figure 3. The labels for the different sub figures got mixed up in the caption. I guess D should be C, E should be D and F should be E?
Yes, corrected

Figure 4. In the top row, one panel shows contours. Are they ice draft?
Yes, it is now clearly specified

p17, l9 conbined → combined
We have changed this part as a response to the other reviewer

p17, l20-25: How do you explain this difference in timescale between the initial melting pulse in the parameterized vs coupled case?
We think that when a melt pulse in the coupled model, a lot of fresh water is added to the system that will further decrease melting. Such feedback is either not or poorly accounted for in the parameterizations. We have added this explanation to the paper.

p17, l25: Do you mean 'lasts longer for the former, about 20a, than for the latter, about 5a'?
Yes, this typo was also pointed out by the other reviewer. We have corrected it

p17, l26: 'a melting minimum': This is not entirely obvious to me, it is certainly not 'universal', e.g. PME1 and $BM5_{500}$ seem to be monotonically decreasing?
This is true, we have rephrased the paragraph

p17, l30: surface → sea surface
Changed

Figure 5. Caption: 'Same as figure 6': please describe here, and refer in the caption of Figure 6 to Figure 5. The solid light grey curves are very hard to see, so I would consider using a darker shade or a different colour. Also, I would like to see these curves included in the figure legend. I'm also finding it impossible to distinguish between Mlin, PME1 and PME4 as they all print in pink. The same happens for Mquad and PME2. Perhaps you could consider expanding your range of colours, at the risk of making this figure even harder to interpret?
This was also a point from the other reviewer. For each of these figures, we have now two new figures for simple and more complex parameterisations. It should be easier to interpret and understand now.

p20, l5-7 I'm a bit lost here, could you please reformulate this statement?
We have removed this part

p21, l5-10: do results converge with a larger number of boxes, or is there an optimal choice for the number of boxes?
In the SI of Reese et al., 2018, a convergence is showed above 5 boxes. In our study we don't find a convergence, but it is unclear whether this is related to different resolutions or to the specificities of the MISOMIP geometry.

p21, l22 and Figure 7. It's unclear to me why you chose 50 years as a time for your performance indicator. As you have 100 years of simulations, why not use 100 years?
As we say in the first paragraph of the discussion, a significant part of the ice shelf is melted out by the parameterisations after 50 years of simulations, which makes comparisons difficult after this period since this is less the case for coupled simulations

Also, is there any way you can account for the difference in initial pulse, which lasts 20 years in the parameterized cases compared to 5 years in the coupled setup? This difference might severely bias your performance indicator, altough it is probably a constant offset.
The duration of this initial pulse is not equal between the different experiments, especially for parameterised simulations, thus we would not know what period to use. We nonetheless agree that assessing the ability of a parameterisation to cope with such a pulse would be interesting.

p26, 27: Appendix B is used twice, please check the numbering
Checked and corrected

p31, l3: remove 'the'
Done

p31, l4: anto → anti and remove 'until'
Done

p33, Appendix F: I'm not sure where this appendix has been refered to in the main text
The figure E1 was cited instead, we have modified it and now cite Ap E