Response to Reviewers

We thank both reviewers for their helpful and insightful comments. Our responses are compiled below. For clarity, we have reproduced the original reviewers *comments in italics*, whilst our responses are in normal text.

In addition to the specific reviewers comments, many of the model descriptions have been updated to include additional comments following the initial review by the editor, Other small typographical errors have been corrected.

Reviewer Comments 1

This is a valuable community paper and an important first step for systematic evaluation of calving in numerical ice-sheet models. The manuscript is well motivated, the experimental design is appropriate for a phase-1 algorithm-focused MIP, and the results will be useful for model developers and users. I recommend publication after following revision.

1. *Line 245, "floating cells adjacent to ....", does this mean CISM only capture floating shelf calving? Can it capture grounded glacier calving activity?*

When the CalvingMIP experiments were designed, not all participants were confident of their model's ability to calve grounded ice. As such, the Thule experiments were framed in such a way as to explicitly not allow grounded calving, despite most models being able to represent this. The fractional area method used by CISM can also allow advance/retreat of grounded cells in a similar fashion, although it is not doing so during CalvingMIP.

2. *Table3, there is a clear resolution mismatch for IMAU UFEMISM (20km) coarser than all the other models. Can results be replaced with 5km resolution simulation instead? In fact, shouldn't IMAU be consistent with NCAR CISM's results with the same resolution (at least for circular domain where calving rate direction doesn't make difference)?*

IMAU UFEMISM does indeed have a larger grid resolution than other models. Due to a lack of time, compute resources and personnel they were unable to perform higher resolution simulations. As CalvingMIP relies on volunteer efforts, we still wanted to include the simulations they were able to perform in the final results. The following text has been added for clarity.

"Due to a lack of time and available resources, higher resolution simulations were not able to be performed." (Line 359)

There might be some confusion on the reviewers part between IMAU-ICE (a regular, square grid-based model) and IMAU-UFEMISM (the unstructured-grid based model used during CalvingMIP. It would not be a trivial task to replace the IMAU-UFEMISM runs with 5 km simulations on an unstructured mesh.

In a separate response to Reviewer 2's comments, we have decided to remove the IMAU UFEMISM results from Fig.5 and Fig.11, as they decided they were not able to be completed satisfactorily with the available resources. Text has been added to the model description and figure discussion to this effect.

"Note that there are no results from IMAU UFEMISM." (Line 419)

"Note that, as in the circular domain, there are no results from IMAU UFEMISM." (Line 481)

“However, due to a lack of time and resources satisfactory results were not able to be obtained for ice thickness at the calving result, and as such no results are shown for IMAU UFEMISM in Fig. 5 and Fig. 11.” (Line 363)

3. *Have all these models conducted convergence analysis on grid size? Might be a good idea to put it in SI and make sure the paper presents converged result from each model.*

We agree that convergence with resolution is always a factor to be aware of with numerical modelling studies. However, as CalvingMIP relies on volunteer participants with no funding or resources provided by CalvingMIP there are practical limitations on how much effort we can expect participants to devote to convergence studies if the projects’ goal of maximising the number of participating model groups is to be achieved. As such, participating models were instructed that “There are no requirements on model resolution or time step beyond the fact that participants should be convinced that their results have converged with both.” but were not expected to submit additional formal, high resolution results demonstrating this.

In the early, initial design stages of CalvingMIP a 10 km results grid was used for the circular domain. Results were very similar to the 5 km results we present here, with differences believed to be more due to other areas of the calving algorithm being actively under development than down to grid resolution. It was decided to increase the final results grid resolution to 5 km to have a resolution more comparable to that typically used in large scale predictive models of Antarctica.

4. *In conclusion text, the author claims calving algorithm does not make a noticeable difference for the result. To justify that, the author should first justify in table3, other variations do not skew simulation results. (for instance, is it safe to claim in current configuration, HO and SSA has similar results, any scaling argument to justify this? The author has already justified for the circular domain, calving rate direction shouldn't make any difference on the results)*

We agree with this comment and have added the following text to the discussion.

“We also find minimal difference in results between the three different DART ISSM configurations, each using a different stress balance (SSA, Higher order, and MOno Layer higher Order), with the exception of slightly more asymmetric results in both the circular and Thule domain for DART ISSMho and the Thule domain for DART ISSMmolho. This implies that calving algorithms in other models should also be largely independent of stress balance.” (Line 550)

5. *Line 385, when interpreting Figure3, I can see the outliers are IMAU & AWI YELMO, is it a typo that the authors emphasize PIK PISM instead?*

Since our initial submission, AWI YELLMO have updated their results and are now more in line with the other models. IMAU UNFEMISM remains an outlier. In this figure, and only this figure, PIK PISM was incorrectly plotted and labelled as NCAR CISM in our initial submission. This has now been corrected. We bring up PIK PISM due to its behaviour in the first 250 years of the simulation along the diagonal profile (Fig 3:b). In line with the above changes, text now reads:

“The evolution of the calving front position in time for the Circular domain is shown in Fig. 3} along with a target envelope of 5 km (one grid cell on the common results grid) deviation from the expected position. Out of 13 models, 12 are within the target envelope for the full duration of the run for Profile A (Orthogonal to results grid orientation) whilst 11 are within the envelope for Profile B (Diagonal to results grid orientation). IMAU UFEMISM falls short of the expected amount of total retreat, approximately 10 km outside the target envelope after 500 years and approaching 100 km of expected total retreat. IMAU UFEMISM is symmetric with this offset in both the retreat and advance stage, finishing the simulation with a position very similar to their initial position. PIK PISM shows a reduced rate of retreat for Profile B falling outside the envelope for the first 250 years of simulation whilst for the remaining 750 years of the simulation the front position falls within the envelope (Fig. 3b). In the case of PIK-PISM, this is most likely a result from interpolating the results onto the common results grid, or from the used analysis method of estimating the grounding line position along the diagonal by using mean fractional area coverage. There is some evidence for a drift in calving front position when comparing initial and final positions, with most models ending position being offset from their starting position by a non-zero amount. However, this is a rather small amount (in all cases less than the target envelope) when compared to the temporal and spatial scales of the simulation." (Line 394)

6. *Fig.3 and Fig.9 are not consistent. The outlier changes from IMAU & AWI YELMO to IMAU & PIK PISM, why?*

As previously noted, there was an error in the labelling of PIK PISM in Fig. 3. In combination with AWI YELLMO updating their results, there is now consistency between Fig3. and Fig9. with IMAU UFEMISM and PIK PISM being outliers in both plots.

7. *For each figure, after stating which models deviate from theory/other models, the authors can add more explanation of why that happens, and mitigation strategies to fix that (if possible).*

We agree with this comment and have strived, to the best of our ability, to do so. As the calving algorithms in question are still very much in active development, there is not always a known and justified explanation. The following text has been added/updated in the relevant sections for each figure:

Fig3:

"A possible explanation for this discrepancy with the expected results in the comparatively low resolution of 20 km used in the IMAU UFEMISM simulations combined with estimating the front position via the fractional area of calving front elements." (Line 400)

"In the case of PIK PISM, this is most likely a result from interpolating the results onto the common results grid, or from the used analysis method of estimating the calving front position along the diagonal by using mean fractional area coverage." (Line 404)

Fig 5:

"This is most likely results from how the sub-grid scale properties at the calving front are being calculated, as judging from their other results these oscillations in ice properties are contained to the calving front itself. " (Line 422)

Fig7:

"In line with the reduced rate of retreat shown by IMAU UFEMISM (Fig. 7it shows less change in floating ice area throughout the simulation. " (Line 432)

Fig 9:

"IMAU UFEMISM stands out from this grouping, with noticeably reduced rates of both retreat and advance, similar to its results in the Circular domain. PIK PISM shows a sudden acceleration in retreat when compared to other models after 350 years, a behaviour which is not seen in the other models. A possible explanation for this is the common result profile interacting poorly with the method used for converting a fractional area on a regular grid into a subgrid scale front position." (Line 454)

Fig 10:

"It should be noted that, unlike in the circular domain, the results profile is not necessarily orientated in the same direct as ice front velocity, potentially allowing for more variation in results due to how each model implements sub grid scale interpolation." (Line 469)

Fig 11:

"This is most likely results from how the sub-grid scale properties at the calving front are being calculated, as judging from their other results these oscillations in ice properties are contained to the calving front itself. "(Line 483)

Fig 12:

"The one exception to this is ULB KORI, which undergoes a spike in model asymmetry after 600 years, with values increasing up to a 20\% difference in quadrant ice area (note the different scale for ULB-KORI in Fig. 12k). This persists for approximately 100 years before model symmetry returns to values that are similar to other models. As this is not reflected in other results from ULB-KORI it is assumed to be a post processing error." (Line 488)

Discussion:

"The calving algorithms demonstrated here are all still in active development, with the results here showcasing a snapshot in time of their capabilities." (Line 523)

8. The project is very valuable as stated by the author "whilst undertaking the work for this project most models significantly improved their results for those initially submitted, both in terms of accuracy as well as spotting previously unnoticed errors in their code". Can the author highlight what bugs in the code for different models were fixed during this project (maybe summarize in a table and attach as SI)? I think the list of "common mistakes to make" is very valuable for students who just start learning ice sheet models. The table can also highlight the mitigation strategy to be implemented in the future, for those models that exhibit limitations during this project.

Whilst we do not have an easy way of fitting these into a table, we have added the following text to discuss some of the problems and challenges faced by participating models.

"Some, but by no means all, examples of this include UNN UA implementing a shock capture scheme to minimise oscillations in ice thickness at the calving front, PIK PISM updating their calving algorithm to incorporate a calving rate as a vector and IMAU UFEMISM had some initial problems determining ice thickness, H, with a moving ice front in a fractional area approach due to the multiple realisations of H(t+dt) needed to determine an implicit solution. Fixing this problem is ongoing work for the IMAU UFEMISM team." (Line 515)

Reviewer Comments 2

This paper presents an inter-comparison of several ice sheet models on a pair of test cases for calving algorithms. The goal is to assess the ability of numerical models to implement calving algorithms with a fixed law for how the terminus evolves as a step towards studying the physics of calving. This is an important first step. If numerical methods for simulating terminus evolution are not reliable, then that problem has to be solved before making claims about physics based on simulation results.

The paper succeeds in getting a comparison among many different ice sheet models and in improving calving algorithms during the process. I think this is a valuable contribution to the collective knowledge and should be published. I have a few reservations. I think the authors make a few claims that are not justified by the results. I also think there was a missed opportunity to probe further at the characteristics of the different numerical methods that were used.

The focus on continuous terminus evolution makes sense given the simulation techniques we have today (typically level-set methods). But the paper makes statements about several places about its applicability to Antarctic calving. Continuous evolution of the terminus is a good approximation for Greenland but it is well established that much of the calving mass loss in Antarctica is due to massive calving events that occur every few decades. That timescale is longer than a typical timestep used in a numerical model. The paper claims that its results lend confidence in models' ability to simulate calving. I would walk this claim back to saying that its results lend confidence in models' ability to simulate one type of calving. This is not to impugn this (very important) work; arguably it isn't reasonable to expect models to be able to do large infrequent tabular calving events with today's numerical methods.

We agree with the general points of this comment, although we believe a more in-depth discussion on this topic will be better placed in planned future experiments focussing on calving laws rather than in this current work on calving algorithms. If a models calving algorithm is able to accurately represent a continuous, time varying rate of retreat then we can assume that it would be able to remove the correct amount of ice from a discontinuous calving event. If feasible, future CalvingMIP calving law experiments intend to include both continuous and discontinuous approaches in their intercomparison.

We reworded the following section of the introduction to more clearly justify our focus on continuous calving:

"We do not see continuous calving as the optimal way of representing calving, but given current model limitations and understanding of the physical processes behind calving it is likely to be the dominant method used in large scale predictions of future ice sheet mass balance for the foreseeable future. As such, increased awareness of its capabilities and limitations are desirable, both in its own right and as a stepping stone to more sophisticated calving algorithms. If a model is able to accurately represent a continuous, time varying rate of retreat then it should also be able to represent a single discontinuous calving event. As such, we focus here on the harder technical challenge of a continuously evolving calving front." (Line 64)

The following section of the discussion has been modified to better describe our future plans:

"To that end, the next phase of CalvingMIP will focus on comparing the impact of different calving laws on ice sheet dynamics. We also intend to, if feasible, compare the choice of a continuous or discontinuous approach to calving. The recent developments in stochastic calving (Ambelorun et al.,2025,Benn et al., 2026), where calving rate is represented by a probability density function of individual calving events, would be another interesting point of comparison." (Line 568)

My other criticism is that the authors make a few statements to the effect that there is broad agreement among the models. The paper is still valuable even if the models don't broadly agree: a negative result is still a result. I think the results, which the authors communicate and illustrate very well, don't justify these statements. For example, figures 9 and 10 show (to my eye) substantial behavior differences between the models. Again, this is a significant result in itself -- it means that numerical modelers have more work to do! I think it would be better to say for example that figure 9b shows that the models relax to the same end state but that some of their trajectories differ rather than claiming that there is broad agreement.

On re-reading, we agree that the emphasis in our comparison was too focused on the initial agreement in the early stages of the simulations, and less on the growing differences, particularly when compared to the circular domain cases. As such, we have reworded the following sections:

Fig 9:

"Whilst there is no a priori solution for the position of the calving front (unlike with the circular domain) there is an initial agreement amongst most of the models, although significantly less so as the simulation progresses and when compared to the circular domain." (Line 452)

"Note that a deliberate choice was made in the experimental design to not specify the rate of advance during the second 500 years of simulation so as not to cause an unphysical situation where the calving front advances faster than the ice velocities at the calving front. As such, the rate of advance is highly dependent on the velocity of ice at the calving front and not necessarily due to differences in the calving algorithm." (Line 462)

Fig 10:

"There is a greater spread and variation of results than was seen for the evolution of the front position (and significantly more than during the circular domain), with the majority of models front-speed increasing as it retreats before decreasing as it advances in the second half of the simulation." (Line 467)

Fig 11:

"Front speed at Halbrane Ice Shelf show a qualitatively similar amount of variation to Caprona ice shelf, with significantly more than the circular domain. The general trend shows ice front speed fall as the ice front retreats and then mostly level off after 250 years. PIK-PISM goes against this trend, with an increase in ice-front speed observed after 250 years. AWI YELMO and IMAU UFEMISM show a relatively sudden drop in calving front speed during the advance phase that is not seen in other models." (Line 475)

Discussion

"Results with more complex bathymetry in the Thule domain show an increasing amount of variation within the simulations. This has the potential to increase further with real world bathymetry, and trying to understand the impact of this variation on predictive models is a key area for future research. However, the overall trend of results show calving algorithms that are ready to accurately and reliably implement calving rates that are generated by separate calving laws, albeit with a few caveats." (Line 533)

Conclusion:

“In a simplified, circular domain different models perform with a high degree of qualitative similarity, however a greater spread in results is observed in the experiments utilising more complex synthetic ice sheet bedrock bathymetry” (Line 576)

Many of the figures (4, 9, 10) show substantial differences between how smooth the model trajectories are. I would have liked to see some more in-depth analysis as to why some show clear stair-step patterns and others do not. For example, if the results show that level-set methods or anisotropic meshes can give smoother trajectories than other schemes, then that is valuable information guiding how future models should be constructed. In the discussion section, the authors state that there were not systematic differences attributable to the scheme but rather argue that the more extreme results are due to how they were interpolated to a regular 5km grid. If that's the case then it's not clear whether the project is an intercomparison of the accuracy of calving algorithms or of interpolation methods.

We agree that this is a fair comment. For the results presented here the calving algorithm used to determine front position should be independent of ice properties at the calving front, as the rate of position change is imposed and has no input from ice thickness or speed and how they are determined via interpolation. We present them in our results to highlight what may end up being a significant issue when we progress to future work involving calving laws. The following paragraph has been added to the discussion to clarify this

“Some models (AWI YELMO, PIK PISM, IMAU UFEMISM) have attributed some of the variation in their results to the interpolation method used to determine front properties. As the rate of retreat imposed in our experiments is independent of modelled properties at the calving this should have had minimal impact on how the front position evolves in a simplified domain (Fig. 3}, even if it is possibly responsible for oscillations in ice speed and thickness at the calving front. This is likely to have a significant impact on results once models generate their own calving rates via modelled ice front properties and a calving law.” (Line 538)

Specific comments:

22-25, "widely varying spatial and temporal scales": I think it's more fair to say that the reason for this simplification is because free boundary problems are demonic and most people don't want to deal with them if they can avoid it.

We strongly agree with the reviewer’s assessment of the demonic nature of these problems and have slightly reworded our statement to diplomatically convey this sentiment.

“One main reason for this simplification is that implementing ice-shelf calving in numerical models is a non-trivial problem due to processes operating at widely varying spatial and temporal scales, with a step change moving boundary between ice and ocean combined with observed individual calving events occurring in a semi-stochastic fashion. Correctly implementing them in large scale models requires a significant amount of time and additional computational resources.” (Line 22)

25-27: I agree if you make this statement about Antarctica but it's harder to make the claim for Greenland.

Sentence changed to:

"Efforts are further hampered for Antarctic Ice shelves by the lack of long term observational data sets to calibrate and validate numerical models of calving, with calving cycles tending to a period of slow advance over potential decades followed by a large calving event inducing rapid retreat (e.g., Cook Glacier; Miles et al., 2018}." (Line 27)

38-39: This is more of a philosophical debate but if you believe that stochasticity is important then it's a good question to ask whether a hypothetical calving model should exactly reproduce an individual observed trajectory or whether we need different ways to evaluate hypothetical models.

We agree that there is definitely scope for a philosophical debate on this topic! Our personal view on this is it depends on what the aims of the overall comparison are. Arguably the most pressing issue for glaciologists is how much does calving impact future sea level rise contributions. If the stochasticity of calving is important for this (which we are of the belief that it is) then we definitely need to take it into account when determining the best method of validating calving models against past observations. Section changed to acknowledge this point without going into excessive detail:

"Some existing calving laws have been evaluated across glaciers in Greenland (Choi et al, 2018) and Antarctica (Wilner et al., 2023) in a single model, and no particular law was seen to consistently reproduce observed calving fronts. These approaches remain marginal, and have never been clearly evaluated in a multi-model environment with an investigation of their impact on sea level rise contribution. As calving is a stochastic process, there remains debate about how best to evaluate model performance when compared to the observational record." (Line 39)

77-78: Is it possible that some desirable properties of the model cannot be achieve with a rectilinear grid? For example, if you try and compute the perimeter of a circle but you discretize the interior/exterior on a regular grid then you never get the right answer.

This is a fair point, but there are valid reasons why models with rectilinear grids exist and continue to be used. CalvingMIP is intentionally designed to allow as many models as possible to participate. We have adjusted the text as follows to make it clear that our goals are not hard and fast requirements, but desirably qualities in a calving algorithm with pre-existing model limitations.

"To this end, when comparing results between participating models, we have four key principles that we believe models should strive to adhere to if they seek to accurately apply a calving rate derived from a calving law to their particular calving algorithm" (Line 72)

362: Should this be a melt rate equal to the local ice thickness divided by the model timestep? A melt rate has to have units of thickness / time.

Sentence changed for clarity:

"Ice points reaching the ocean domain (positive LSF values) are removed using a melt rate equal to the local ice thickness, removing all ice over one model time step." (Line 374)

401-410: What's going on with UFEMISM here? This looks like a sampling error which could unfairly make that model look worse than it is.

This is correct. Unfortunately, this is an issue that can not be remedied without the IMAU UFEMISM group redoing their simulations and they do not have the people and computational resources available to do so. As a compromise, we have removed the potential distracting IMAU UFEMISM thickness results from Figure 5 and 11 and explained their absence in the text.

Figure 8: These look near identical in the eyeball norm, it would be better to plot the differences against some baseline (e.g. the average thickness among all models) with a diverging color map.

Both figure 8 and 14 have been updated to show model thickness relative to the model mean thickness. As a results of this, some trends in ice thickness that was previously hidden by our choice of colour scale have been identified and discussed in more detail.

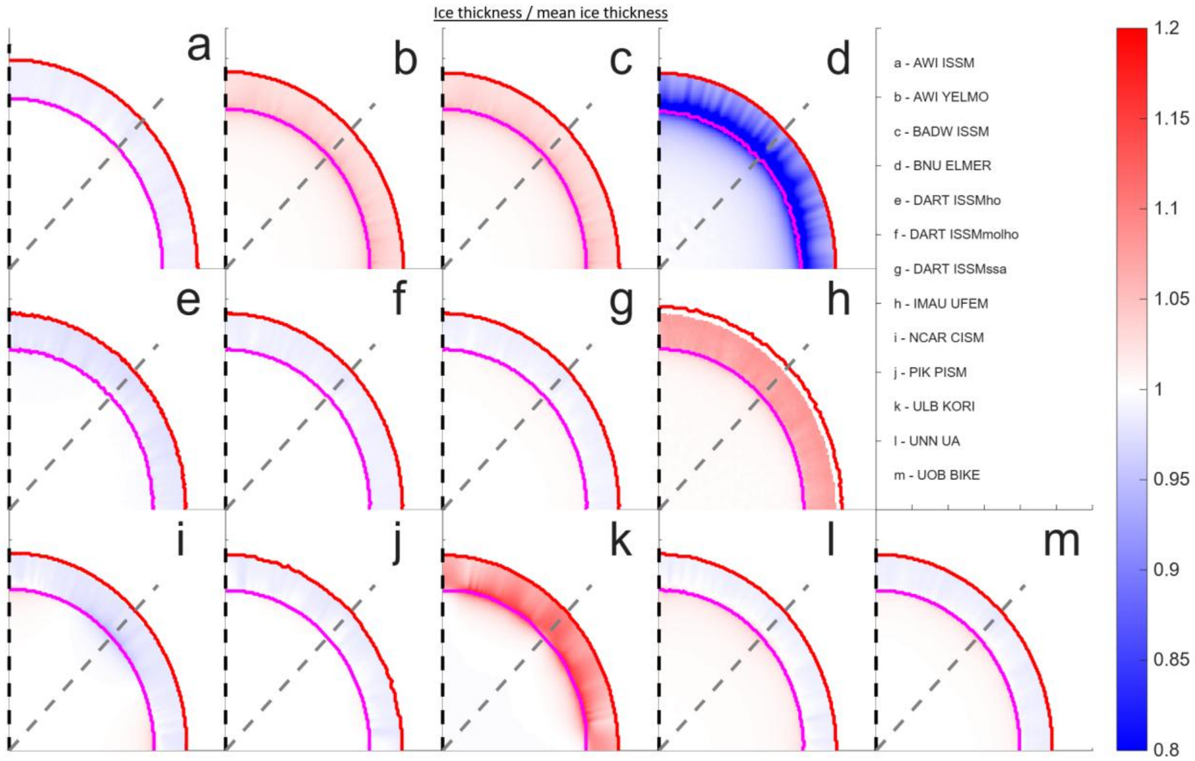

Figure 1 New version of Fig 8.

"Fig. 8 shows modelled ice area and thickness relative to the mean thickness for all models at the minimum calving front extent after 500 years of simulated time. Qualitatively, there is a very high degree of similarity for grounded ice, with more variation apparent in the floating ice. Floating ice thickness typically varies between \pm 10%. BNU ELMER is noticeably thinner, with floating up to 20% thinner than the model mean, however at the calving front itself the ice thickness becomes much closer to the model mean (Fig. 5}. ULB KORI has floating ice up to 20 \% above the model mean, although with noticeable spatial variability. This is most likely responsible for the relative higher asymmetry for ULB KORI shown in Fig. 6." (Line 444)

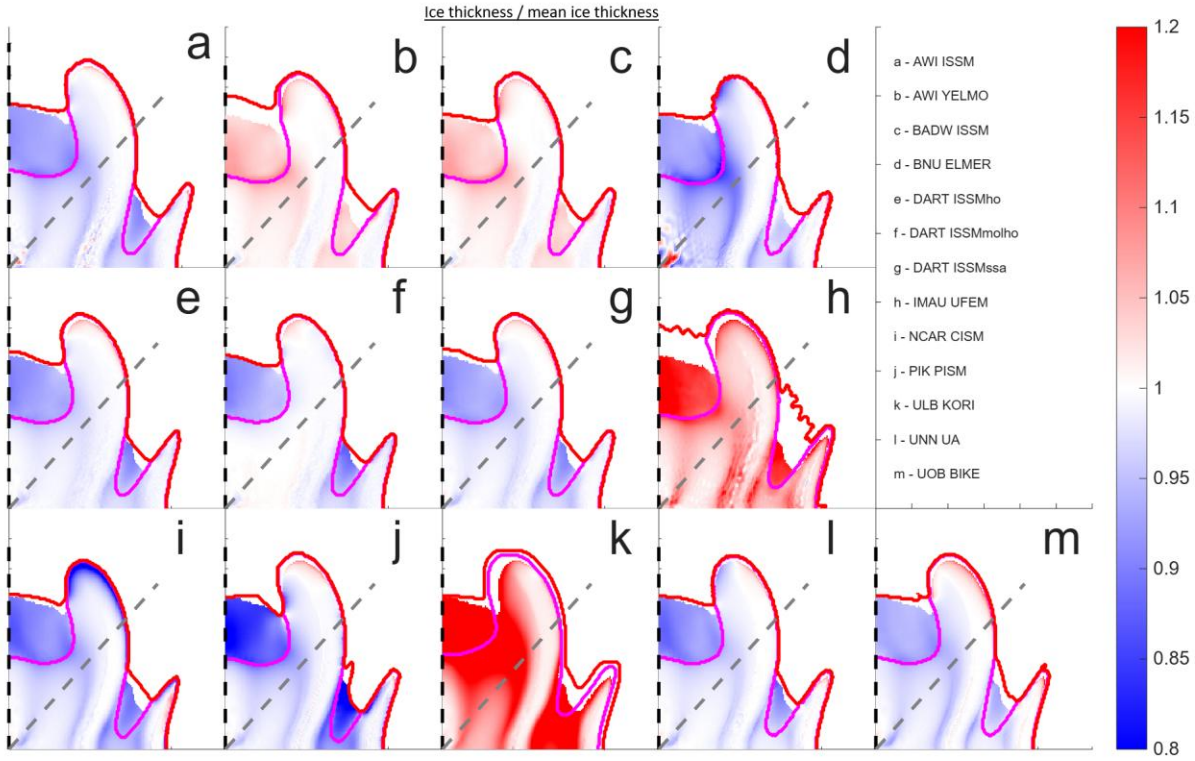

Figure 2 New version of Fig 14.

"Fig. 14 shows modelled ice area and thickness relative to the mean of all models at the minimum calving front extent after 500 years of simulated time. Similar to the Circular domain there is a greater variation in ice thickness for floating ice when compared to grounded ice, although unlike the circular domain there is now noticeable variation in the grounded ice thickness. This is most likely a result of the change in buttressing strength provided by varying floating ice thickness impacting grounded ice. As before, thickness varies by up to /pm 20% with ULB KORI and IMAU UFEM being noticeably thicker than the average and PIK PISM relatively thinner." (Line 501)

"Whilst we can not attribute the variation in ice thickness across the whole domain of our models to ice shelf calving alone, it does highlight the importance of ice shelf buttressing on long term ice sheet behaviour." (Line 562)

435, "good general agreement": these look noticeably different, I'd say the final states agree but the trajectories differ

In line with your previous larger comment, this section has been updated.

451: The absolute magnitudes are close enough but there are noticeable differences among the models. Some are oscillating, others are smooth. If we could rule out the possibility that this was a sampling or interpolation artifact then those oscillations looks like a numerical artifacts. That's understandable when simulating such a challenging system but it's hard to draw a conclusion one way or the other if, as the authors say, this might just be because of interpolation.

In line with your previous comments, we have added some discussion on interpolation to the discussion section.