Final Response Reviewer 1

We thank the Anonymous reviewer #1 for their insightful and very detailed review. Below we explain in detail how we intend to address the reviewer's comments during the revision process. Our replies are highlighted in italics.

Beckmann et al. analyze ice-sheet model outputs from ISMIP6. They are concerned with disentangling different sources of uncertainty. I think the main finding of the paper is that the initially most significant, melt-related uncertainty declines past 2100, after which calving-related uncertainty becomes the most significant uncertainty source.

We will highlight more the increase of calving related uncertainty after 2100 as this is a point also Reviewer 2 pointed out.

General comments:

1) TF time series: The authors face a major challenge in terms of inferring the thermal forcing present through the simulations, that was however absent from the outputs. They choose to simply interpolate temperature to the modeled ice base geometries (which are present in the output), independent of what parameterization (producing the resulting TF) was actually used. It remains unclear whether that quantity has anything to do with the actual parameterized TF. I think it needs to be shown (or referenced if that has already been done elsewhere) how well these two quantities are actually related. This could be done, for example in some separate ice sheet model run where the TF inferred and TF calculated are evaluated for all parameterizations used in the analyzed ensemble.

We agree with the reviewer that our approach represents a simplification, and we acknowledge that the lack of explicitly outputting thermal forcing (TF) fields poses a limitation. TF is not supplied as an output in ISMIP6, so we had to make a methodological choice on how to approximate it. Most ice-sheet models in the ensemble employ melt parameterisations that depend on thermal forcing beneath the ice shelf (often in a quadratic formulation). Based on this, we approximated TF by interpolating ocean temperature to the modelled ice-shelf base geometry. However, we agree that this inferred TF is not identical to the TF used internally by each model. The effective TF entering the parameterisations depends not only on geometry, but also on model-specific choices such as temperature corrections, reference fields, and implementation details. We also note that some parameterisations (e.g. PICO , PICOP and linear melt schemes) rely on far-field temperatures rather than local under-shelf TF. This implies that there is no uniquely defined TF across models, and a direct one-to-one comparison is not possible within the protocol and outputs of ISMIP6. A full evaluation of inferred versus internally used TF across all parameterisations and models would require additional dedicated simulations and diagnostics, which is beyond the scope of the present study. We will clarify this limitation more explicitly in the revised manuscript. Nevertheless, we expect that our definition of TF is appropriate for our aims. Specifically, our TF input data are derived from predominantly coarse-resolution CMIP ocean fields, extrapolated beneath the ice shelves using a simple flooding algorithm (e.g. Jourdain et al., 2020). Under these assumptions, we do not expect large systematic differences between temperatures at the ice-shelf front and beneath the shelf, except in regions with strong bathymetric gradients. To further address the reviewer's *concern, we will include an additional analysis for one representative case (e.g. PIK-PISM), where we examine the evolution of TF and quantify the potential differences arising from our approximation. This will be discussed explicitly in the revised manuscript.*

2) MS factor normalization: The normalization by area in the calculation of the MS factor seems to be problematic for comparing different ice shelves, whose melt rate distributions have different shapes. This normalization choice, rather than a more fundamental physical reason, is probably why large ice shelves have lower, and small ice shelves higher MS factors. This choice does not affect the variability of the derived MS time series, and therefore doesn't affect conclusions about the uncertainty attribution. However, this choice does affect the discussion about geographic variability of the MS factor across regions.

We thank the reviewer for this important point and agree that the choice of normalisation by area may influence the comparison between ice shelves of different sizes. This choice was made to enable a consistent comparison across models and ice shelves, particularly given their differing geometries and temporal evolution. We did consider alternative approaches; however, each introduced its own limitations. For example, we could have chosen to evaluate melt rates close to the grounding line (C1) or melt at depth (C2), but both are highly dependent on the resolution of the ice model. On the other hand, using integrated melt alone can bias the comparison towards larger shelves. Therefore, while area normalisation is not without limitations, it provides a pragmatic and consistent framework for inter-model comparison. We agree with the reviewer that this choice could influence the apparent differences between small and large ice shelves, and we will explicitly acknowledge and discuss this effect in the revised manuscript, particularly in the context of geographic variability of the MS factor.

3) MS (and DS) factor derivation from data: There were a few fitting choices listed in the paper, and there is a fairly complicated and sometimes confusing explanation of how different choices and methods were used for different models and regions, but the reasoning isn't always very convincing. Perhaps it is just a problem with presentation, but I was not left with the impression that the derivation of the factors is particularly robust. I think this should be addressed somehow, but I don't have a good recommendation. Perhaps starting with a goal of what timescales/regimes these MS/DS factors should be useful for (rather than the applicability range being an outcome) could be a good start, but I am not entirely sure.

We thank the reviewer for this thoughtful comment. We agree that the current presentation of the MS and DS factor derivation may give the impression of being overly complex and insufficiently motivated. We will revise this section to clarify both the methodological choices and their underlying rationale.

Our objective in deriving the MS/DS factors is to obtain values that allow the most accurate reconstruction of the simulated basal melt response and that are most fair to the models. Specifically, we determine MS such that the predicted basal melt (B_predict), based on a linear *sensitivity to thermal forcing, closely matches the simulated basal melt (B) over the full time series.*

The use of different fitting approaches reflects the need to ensure a robust representation of this relationship across models with differing behaviour, rather than arbitrary methodological choices. While the method used to estimate MS may vary depending on model characteristics, the criterion for evaluating its performance, i.e.,how well B_predict reproduces B, is applied consistently across all cases.

We will clarify this objective and streamline the explanation in the revised manuscript to make the derivation and its robustness more transparent.

We believe these revisions will make it clearer that, while methodological choices are required due to the diversity of the ensemble, the resulting MS/DS factors are robust within the intended scope of the analysis.

> 4) There are multiple instances of interpretation of results that do not seem to be reflected in the figures or based on the formulas provided, and rather seem based on observational understanding of the system, however, without showing that the modeled results are consistent with these observations (e.g. comment 458 below). In general, more frequent figure referencing would help, and perhaps even completely solve this concern.

We thank the reviewer for this important observation. We agree that, in several instances, the connection between the interpretation in the text and the supporting figures or quantitative analysis is not sufficiently explicit. Our intention was to interpret the model results in the context of established physical understanding; however, we acknowledge that this can give the impression that some statements are not directly supported by the presented results. We will revise the manuscript to ensure that all key interpretations are clearly grounded in the model output. In particular, we will strengthen figure referencing throughout the text so that each interpretation is explicitly linked to the relevant figure or quantitative result. Where support is not directly shown, we will revise or qualify statements to avoid overinterpretation. Where needed, we will include additional quantitative evidence to better support key claims, and we will clearly distinguish between interpretations based on our simulations and those informed by established theory or observations. We believe that these revisions will improve the transparency and traceability of the analysis and address the reviewer's concern.

In-line comments:

18: 'will' feels a bit too definitive

We will adjust.

22: Do you also want to specify scenario for the lower bound of sea-level rise projections as you do later for the upper bound?

Yes we can add the specific scenario in the manuscript.

27: Sentence starting 'The largest sources...' too long to be clear, please rephrase or break up.

We will rephrase.

34: "show a much larger spread, with no clear correspondence to the climate forcing (Fig. 1b)." - this seems to be the case only after some time can you clarify if that is the case? Also can't really see that clearly because the plotting order seems to be by color - perhaps it would be better to interlace the colors.

Correct, after 2100 we see that spread. We will add that to the text and adjust the Figure with interlacing colours.

35: and also as function of the CMIP model output.

Absolutely correct, we will add this to the manuscript.

44: Coriolis is not a property like salinity or temperature. The whole parenthesis isn't really clear actually - just remove or actually list processes that are simplified or missed.

Agreed, we will delete the parenthesis.

47: Probably worth adding a citation to and ESM that actually has coupled ice sheet - e.g., Siahaan 2022.

Agreed, we will add the citation.

102: Please provide reasoning why you restrict your analysis only to high-emissions scenarios.

We restrict the analysis to high-emission scenarios because they provide a larger range of thermal forcing changes, allowing for a more robust estimation of melt sensitivity. In lower-emission scenarios, the signal is weaker and harder to distinguish from internal variability. We will add this explanation to the manuscript.

127: rather then "acting" something like "that would have acted" to indicate that you are trying to infer something close but not exact to the field that was used in the simulations but was not saved.

Agreed, we will adjust the text.

128: what is the climate dataset? do you mean the ocean thermal forcing dataset mentioned earlier? If yes, please keep terminology consistent.
We meant the ocean data set provided by ISMIP6 for the ice sheet modellers . We will clarify in the text.

125-130: I am not following here. Why do you linearly interpolate properties, rather than using the Jourdain extrapolation algorithm that was actually used in the simulations? If these results were close, why would the more complicated Jourdain extrapolation actually ever be used?

The ISMIP6 thermal forcing fields are based on CMIP5/6 ocean model output, which does not resolve ice-shelf cavities. Therefore, Jourdain et al. (2020) extrapolated ocean temperatures throughout the entire Antarctic Ice Sheet (i.e. assuming no grounded ice) to provide continuous forcing fields (x, y, z, t) that can be applied as ice shelves evolve. In our study, we aim to reconstruct the effective thermal forcing experienced by each ice sheet over time. Because ice-shelf geometry evolves (e.g. retreat into shallower regions), the thermal forcing entering the melt parameterisation also changes spatially. To account for this, we interpolate the 4D forcing fields onto the time-evolving ice-shelf geometry of each model simulation. This approach therefore does not replace the Jourdain extrapolation, but builds on it by mapping the provided forcing consistently onto the evolving geometry, allowing us to diagnose the thermal forcing actually experienced at each time step, in the absence of this field being provided as a model output in the ISMIP6 protocol. We will clarify this more clearly in the revised manuscript.

135: I don't agree with this statement about consistency. It is not important to have one single method for all parameterizations, the important thing is to recover the thermal forcing at the ice shelf base equally well (with respect to what was actually present at the time of simulation) for all parameterizations.

We thank the reviewer for this clarification and agree that the key objective is to recover the thermal forcing at the ice-shelf base as accurately as possible for each model.

However, because thermal forcing is not explicitly archived and differs in its implementation across parameterisations (e.g. under-shelf vs. far-field, temperature corrections), a direct reconstruction of the "true" forcing is not possible without requesting significantly more work from the ISMIP6 model groups. Our approach therefore applies a consistent diagnostic method by interpolating the provided 4D forcing fields onto the evolving ice-shelf geometry. We acknowledge that this may not capture all parameterisation-specific differences. To assess the impact, we will include a quantitative comparison for a representative case (e.g. PIK-PISM with PICO) and discuss the implications in the revised manuscript

131: 'may differ slightly' seems like an understatement.

We will quantify the difference and formulate a more precise statement.

155: I am not sure about the purpose of the note. Perhaps add references that do either of the two approaches so that there is some context?

We agree that the note is not sufficiently clear in its current form. Our intention was simply to define the dynamically driven contribution to sea-level rise within the simulations. To avoid confusion, we will remove this note and instead clarify the definition of the dynamic contribution more directly in the revised manuscript.

Equation 1: What is the meaning of c? Shouldn't there be at time $t = 0$, 0 change in TF and BMB, therefore making $c = 0$? In that case the fit should be forced to pass through 0 - indeed from fig S3 there is a strange offset in the black, relative to blue lines which seems to be caused by nonzero c.

Also, TF has also already been shelf averaged, correct? in that case it seems that the shelf area just cancels on both sides, no?

We thank the reviewer for spotting this. The formulation of Eq. 1 was indeed misleading and will be corrected. In the actual MS derivation, we do not fit anomalies forced through zero, but regress the simulated basal melt against thermal forcing as:

$\overline{BMB} = MS \cdot TF + c\overline{BMB} = MS \cdot TF + c$

Where $\overline{BMB} = BMB/Shelfarea$

Here, c represents the intercept or background melt contribution in the ice-sheet model, i.e. melt not explained by the shelf-averaged TF predictor alone. This is needed because the predictor is an area-averaged TF over the shelf; local regions may still experience positive TF and melt even when the shelf-mean TF is small. In addition, model-specific choices such as regional temperature corrections or melt-parameterisation details may allow non-zero melt for low values of the averaged TF.

We will correct Eq. 1 and clarify that TF is already shelf-averaged.

165: How is TF defined? In situ - freezing temperature at each ice shelf base location, correct? Can you make that explicit somewhere, if that is what you do?

Yes, the difference between the in situ ocean and the in situ freezing temperature is the thermal forcing which was provided as the 4d climate data forcing set (Jourdain et al). We then average TF over the shelf area. We will refer to this explicitly in the new manuscript.

170-175: Could you please include these plots with fits from which you derive MS (and analogously for DS) in the supplement so that we have a sense of how reasonable it is to fit a line to there? Ok, later on I see there is an example in S3, so perhaps cite it earlier? Also, can you provide examples from each end of the spectrum of Fig S2? Since the goodness of fit and melt factor value follow the same curve, I wonder what that actually says about the utility of the melt factor across all models, so some figures would help there.

We will add figure and reference them in the text to support our statement here.

175: Is it the abruptness of the change, or the fact that the geometry changed enough? And if it is simply the latter, does that mean that your method is good for estimating melt rate in static ice shelf cavities, but no so good for evolving ice shelf cavities? And how much shape evolution is too much for it?

It is the abruptness of the change due to large calving events. indicated in Fig S2 the area (grey line) does already change before BMB max and BMB can be reconstructed with the MS factor but this abrupt change leads to a discontinues timeseries which is harder to fit to .We will reference to this Figure in the manuscript to underline our statement.

Fig S2: x labels on the bottom of the right column are incorrect. Also, what is the grey dashed line and why are there two legends - please clarify in the caption.

We will correct. The grey dashed line is the ice shelf area (A) which we need to add to the figure caption.

Fig S3: the legend for the bottom 3 panels is at the top panels and the legend for the top three panels is missing.

We will rearrange the legend position and add the missing legend.

Also, can you visually mark the time period to which the line is fitted? (for example by different color of dots)

Great idea, we will mark the dots used for fitting.

181: What is 2K in TF<2K?

2 Kelvin of thermal forcing, so the difference of the mean ocean temperature and the mean freezing temperature (depth and salinity dependent).We will clarify that in the manuscript.

185: I don't understand why it is reasonable to use a different fitting method for different model simulations, since what is compared at the end of the day, are the coefficients across simulations. Because of that, it seems to me more important to have the same method for each simulation used for a given sector, then the same methods used for all sectors within a simulation. Still though, it feels that the same method should be used for all of these, if all are to be compared. Ultimately, I think it is important to show and communicate whether these choices influence the results or not, and I don't think I got that out of reading the paper.

Our objective in deriving the MS factor is to obtain values that allow an accurate reconstruction of the simulated basal melt response. Specifically, we determine MS such that the predicted basal melt (B_predict), based on a linear sensitivity to thermal forcing, closely matches the simulated basal melt (B) over the full time series.

The use of different fitting approaches for MS reflects the need to ensure a robust representation of this relationship across models with differing behaviour, rather than arbitrary methodological choices. While the method used to estimate MS may vary depending on model characteristics, the criterion for evaluating its performance—how well B_predict reproduces B—is applied consistently across all cases.

We will clarify this objective and streamline the explanation in the revised manuscript to make the derivation and its robustness more transparent.

Equation 3: same as eq 1 - what is the physical meaning of d, can you provide analog of figs S2 and S3 for DS?

The constant d represents the intercept or background dynamic ice loss represented by the model that is not explained by the cumulative BMB alone. We will add the corresponding figures to the supplementary.

3.1.2: I am not clear on if you are classifying models based on calving laws implemented, or if you simply noticed a bimodal distribution in the results and you are pointing out that observation.

We find a bimodal distribution in shelf area changes and cumulative BMB. The distribution can be physically interpreted in terms of the different calving approaches used by the models.

To strengthen this point, and following the suggestion of Reviewer 2, we will include a formal statistical test to assess the significance of the differences between the identified groups. This will provide a more robust basis for the grouping and its interpretation in the revised manuscript.

291: can you call it G1 etc same as in Fig 5?

Yes, this will clarify the text. We will add this to the new manuscript.

290-310: can you add figure references to statements? Also elsewhere in the paper.

Yes we will add figure references here and elsewhere in the manuscript.

320: Is it delta BMB not BMB here and further in the section?

Yes we will carefully review the whole manuscript and change BMB to delta BMB where appropriate and define exactly.

329: Why don't you do this for all sectors?

We did this for all sectors but did not find a significant relationship as per line 336-337, which we will move earlier in the paragraph to be clearer.

338: What do you mean by "basal mass balance evolution is dominated mainly by the calving group"? Are you saying that the same bimodal distribution that you observe in ice shelf are change also appears in delta BMB? Or do you mean something else?

Yes exactly, we will state this more clearly in the revised manuscript.

Similarly, what do you mean by "basal mass balance evolution is explained solely by the MS factor"

We will clarify this and add figure reference.

Fig 9: Why don't the IMAU models have different regions shown? Do all regions have the same factor value for this model? - ok I see later this is mentioned in the text, please add to figure also.

Yes will add the figure reference to the text.

399: Figure reference to statement please.

Yes will add the figure reference to the text.

409: Figure ref please.

Yes will add the figure reference to the text.

450: I think the term 'predict' is incorrect here and elsewhere, I think 'diagnose' would be more accurate.

Agreed, but maybe "reconstruct" ice sheet model behaviour is even better.

458: I don't think it is the control value of TF but the change in TF that would give rise to high MS factor (see Eq 1). Further, is the highest change in TF actually taking place in the Amundsen Sea - some figure that would show the continental shelf TF anomaly time series for the different regions would be useful.

The reviewer is correct that the MS reflects the response to changes in TF rather than the magnitude of TF itself. Our original statement was intended to relate this sensitivity to regions such as the Amundsen Sea, where both high TF and strong melt are observed, which could point to a high MS. However, we agree that this needs to be clarified: high background TF does not necessarily imply a high MS factor.

460-480: Isn't the reason for the difference between low MS factor at Filchner-Ronne and high MS factor at Amundsen just the fact that you normalize by ice shelf area? And if that is the case, what does that really tell us about the usefulness or meaning of this MS factor? I understand the desire to normalize by area, but I am not convinced in this case it is very meaningful. The reason for that is that fundamentally, melting (and change in melting) is highest at deeper portions of the ice shelves, and smaller, shorter ice shelves (Amundsen) have higher fraction of high-melting portions than big ice shelves (Filchner-Ronne) - so I think your MS factor differences just reflect this fact. Conversely, if you do not normalize, you would probably get much higher MS volume flux factors for bigger ice shelves than smaller ice shelves, which is also not very informative. But perhaps there is some better metric that takes into account the distribution of melt rates over ice shelves and compares their distributions in a more meaningful way than a simple average, which seems more appropriate for gaussian distributions (I don't think melt rates on an ice shelf are necessarily gaussian).

As the reviewer correctly points out, normalising by ice-shelf area can influence the interpretation of MS factors, and part of the signal may indeed reflect differences in the distribution of melt rates (e.g. the relative contribution of deeper, high-melt regions). We will clarify this limitation more explicitly in the revised manuscript. But- as the reviewer also mentioned correctly- not normalizing the BMB would also lead to biases.

We have explored alternative approaches to better capture these effects. In particular, we considered deriving MS factors based on melt rates near the grounding line and at different depth levels. However, due to the relatively coarse resolution of some ice-sheet models (i.e. 16-32 km), the definition of "near the grounding line" becomes ambiguous. Nevertheless, for some regions, an MS factor based on near-grounding line melt can still provide useful insight into sensitivities (see Appendix C1), and we will expand this comparison in the revised manuscript.

We also investigated whether focusing on deeper parts of the ice shelf—where higher melt rates are often expected—would provide a more robust metric. However, this assumption is not universally supported, as the distribution of melt rates depends not only on geometry but also on the chosen melt parameterisation and the evolving ice-shelf configuration. In Appendix C2, where we analyse melt contributions across different depth levels, we do not find a consistent statistical relationship supporting this assumption. To further address the reviewer's concern, we will include an additional figure comparing depth-dependent melt profiles across representative ice shelves, providing a more quantitative illustration of how melt distributions differ and how this relates to the MS factor.

505: Is there even any physical reason you would expect MS and DS factors to correlate? If yes, perhaps start with that as a motivation. Reading the results often feels like you are just correlating random time series and waiting if something pops out of it (so the danger there is spurious correlation).

There is no physical reason to expect that the MS and DS factors correlate, but we wanted to investigate whether there was a relationship between the two. We will revise this paragraph to indicate that, and provide our physical motivation to investigate the relationship, before talking about the correlation between the two sensitivity factors.

531: Can you elaborate on how these factors would be useful? Do you mean using them to train emulators? They don't seem too robust, especially across a range of different regimes, so it seems far better to just strain the emulators on the data that was used to produce the MS and DS factors.

Yes, we will clarify our intention. The main purpose is twofold:
(1) to derive meaningful MS values (e.g. from observations) that can support improved calibration of ice-sheet models. We discuss their utility in more detail in discussion section 4.3 points 1-3. And
(2) to provide a quantitative measure of the sensitivity of ice-sheet models to changes in thermal forcing or basal melt, enabling a clearer interpretation of sea-level rise projections, analogous to the concept of climate sensitivity in GCMs.

536-537: So those calving groups G1 and G2 actually have different sets of calving parameterizations? If that is the case that should have been stated clearly in results already (maybe I missed that, but was definitely confused when reading results).

Yes, they do, or on some cases they use higher parameter values (e.g. higher minimum thickness thresholds in minimum-thickness-based calving schemes). We will make sure to state this clearly in the revised manuscript.

551: Resolution of which model? Of the ice sheet model? Or of the melt rate parameterization (if such a thing applies)?

Yes, ice sheet model. We will clarify this.

567-568: So what else then probabilistic projections is appropriate? Or are no projections appropriate at all at this point for these sectors?

Our point here is that taking into account all the different models with such a high range in MS, most likely not close to observation, will lead to misleading probabilistic SLR projections. Some should be appropriate depending on whether the MS matches the observational values. We will improve this sentence for clarity in the revised manuscript.

569-576: You seem to suggest that the high sensitivity in these regions comes from modeling choices, but what if it is simply 'intrinsic variability' - that is high sensitivity as a result of particular group of geometries and locations?

If the reviewer is referring to the possibility that strong bedrock gradients may influence the initial ice-shelf configuration and thereby the thermal forcing available beneath the shelf, this is indeed an important point. We will include this consideration in the discussion.