We thank the reviewer for their time and effort in providing feedback on our manuscript. Our responses to their comments are given below, in red. In the case where no direct response is given, we will implement the reviewer's suggestions directly into the revised manuscript without further change.

General comments

This paper presents a thorough and rigorous method for producing calibrated projections of sea level rise from the Amundsen Sea sector of West Antarctica, using a multistep process to consider and constrain uncertainty from a wide range of sources, by combining both a numerical ice flow model and a statistical surrogate model. The resulting sea level contributions by 2100 under RCP8.5 and Paris2C are at the lower end of the range of estimates from previous studies and similar between the two scenarios. They also extended some simulations to 2250 and found the two scenarios diverged after 2100 due to differences in snow accumulation.

In general, the study is very well presented (save a few consistency errors – see technical corrections below) and structured clearly, with a lot of methodological detail. However, I think certain later sections could benefit from further discussion and interpretation, e.g. Section 4, in particular, is very brief and offers no real interpretation, which is a bit of a shame – see specific comments for more detail.

All reviewers agreed that further discussion and interpretation of the results would be beneficial to the paper and we will expand on numerous sections based on this feedback.

Specific comments

There are several places where statements are made without supporting citations:

- L83-84: give examples of studies that use a mixed sliding law
- L114-115: reference for 50% figure?
- L148: reference for melt due to GHF?
- L537: reference for observed rate

We will add supporting citations for these statements

L106: what is the threshold for the limit on superimposed ice, and where does this come from? How sensitive is runoff to this threshold – couldn't this also feed into the Bayesian inference?

Following the description of Janssens and Huybrechts (2000), we use the capillary retention model that takes into account the refreezing process and the capillary suction effect of the snowpack. The potential-retention fraction is given by:

$$p_r = min\left[\frac{c}{L}T.\frac{C}{P} + \left(\frac{C-M}{P}\right).\left(\frac{\rho_e}{\rho_o} - 1\right); 1\right]$$

Where c and L are the specific heat capacity and latent heat of fusion of ice, P is the mean annual total precipitation, T is the mean annual temperature, M is the total annual snowmelt, ρ_o and ρ_e are the densities of dry snow and water-saturated wet snow, respectively. We will give more details in the revised manuscript.

Regarding the sensitivity of runoff to this threshold; certainly, there will be an affect and therefore a sensitivity that could feed into the Bayesian inference. We selected two parameters of the PDD model to explore two mechanisms: the increase in precipitation and increase in temperature extremes as CESM1 temperatures change during the course of our simulations. The parameter σ_M already essentially changes the amount of runoff in our model, and so sensitivity to this parameter and the retention fraction would presumably be similar. Since the difficulty of our Bayesian inference and the number of simulations required scales nonlinearly with the number of parameters we explore, adding another parameter would only make sense if that allows to include an entirely different mechanism. The degree day factors, for example, could also be included but we did not for the same reason.

L113-114: This sentence implies that there are no other contributors to basal melt of grounded ice – perhaps worth clarifying that this is referring to your model rather than reality. For example, geothermal heat flux is mentioned later on as being of less importance, but this sentence implies it doesn't contribute to melt at all.

We will clarify this and include a comment on geothermal heat flux in the revised manuscript.

Figure 2: caption could include more detail, e.g. define the colormaps. In the main text, are you able to offer any physical interpretation of what these PCs are showing?

We will include a colormap in the revised manuscript. Regarding a physical interpretation of the PCs; broadly speaking PC1 represents an overall thinning and retreat signal of the entire ASE region, which is unsurprisingly the main mode of change in the region, and subsequent PCs represent variation in this response between the Pine Island, Thwaites and Dotson/Crosson catchments. We are cautious about ascribing these smaller PCs to any one particular physical mechanism, however, they are more like ingredients that can be combined in different amounts to yield a spatial pattern that most closely matches the result of each simulation, and the mechanism behind that spatial pattern could be very different in each case.

L369: is the assumption that observational errors are spatially uncorrelated robust? E.g. velocity errors (in magnitude and direction) are dependent on flow speed.

This is a simplifying assumption that makes the probabilistic inference easier, but one that is commonly used in similar papers, e.g. Wernecke (2020)

L446 / Section 4: the interpretation and discussion in this section could be expanded. For example, why do think that the weight given to surface elevation change in the inversion (along with other inversion parameters) is the biggest contributor to uncertainty? Does this indicate that the model is most sensitive to C and A, or are the inversion parameter priors simply considerably wider than those of other parameters? Given the sensitivity, what is the implication therefore of keeping the C and A fields constant throughout century-scale forward simulations (I realise this is the way it's done, but interested if your results give further insight into whether this can really be justified)? How do the various inversion parameters translate to variation in C and A, in magnitude and spatially? Do these fields differ a lot between the 1996 and 2021 initialisation?

We will expand on this section in the revised manuscript, but we address these questions directly. There are a number of things to disentangle here, but essentially what is clear is that the parameters related to our model initialisation contribute the most to uncertainty and among those, the weight given to the surface elevation change in the inversion through parameter $\dot{h}_E$ is the most important.

The model response is indeed sensitive to spatial variability in the A and C fields, as would be expected since for uniform A and C fields the model would completely fail to match observed velocities. However, we don't expect the parameter priors play an important role, since in the Bayesian framework the importance of the prior diminishes rapidly as you increase the amount of 'data', or in our case model simulations. Comparing the A and C fields between simulations is challenging, since every inversion simultaneously changes every parameter including *m* and *n*, which therefore change the units of A and C. This also precludes a comparison between 1996 and 2021 initialisations, since they are separate experiments sampling different parts of the parameter space so again no two simulations will have the same combination of m and n. Regarding the assumption that A and C fields do not change during the course of our simulations, clearly this is a limitation of many ice sheet modelling studies. We would argue that for forecasts over relatively short timescales (80 years for most of our simulations) this assumption is not necessarily a bad one, however we cannot quantify that with these results and are merely speculating. Again, we will expand on a discussion of these results and their implications for modelling studies that use this type of initialisation.

For some parameters there is a big difference in the Sobol indices between RCP8.5 and Paris2C (e.g. basal mass balance parameters) – could you comment on this?

When comparing all Sobol indices between the two scenarios there is a general tendency for RCP8.5 to show more sensitivity to parameters related to external forcing (p, Gamma_{TS}, E_{ID}) and in contrast for Paris2C simulations to show more sensitivity to parameters related to internal ice dynamics (m and inversion parameters). This makes sense since, as we show in the paper, the changes in mass balance are the main cause of differences between these two simulations in terms of cumulative sea level contribution. We will add a comment emphasising this point in the revised manuscript.

L478: perhaps worth noting that this rate is very similar to the present day observed rate for the sector.

We will add a note to this effect.

L485-486: "Secondly, using adaptive mesh refinement…" – many of the studies cited above use BISICLES, which also uses AMR, with a resolution of 250 m at the grounding line, so this sentence is a bit misleading. In general, this paragraph could be developed further – e.g. the inclusion of a plume model rather than a very simple melt rate forcing used by some of the others could be discussed.

We agree this should be framed better, since no single aspect of our study is entirely new and the novelty is in bringing all of these components together, so we will reword this in the revised manuscript.

L561-564: It seems strange to introduce the concept of testing the reversibility in the final paragraph of the conclusions. Perhaps it would be better suited to the discussion, along with the relevant citations.

We will add some brief discussion related to this earlier in the revised manuscript.

L580-581: I think limiting ocean driven melt to strictly ungrounded nodes is a sensible and widely used approach, but here or elsewhere, perhaps it is worth commenting on recent modelling and observational work that indicates that ocean water intrusions far upstream of the grounding "line" could be contributing significantly to melt? E.g. Bradley and Hewitt (2024, 10.1038/s41561-024-01465-7), Rignot et al. (2024, 10.1073/pnas.2404766121).

We will add a remark along these lines in the revised manuscript.

Appendix B: perhaps I missed this, but how do you sample the parameter space for the Ua-obs ensemble used to train the surrogate?

Parameters are sampled using latin hypercube sampling and we will clarify this in the revised manuscript.

Technical corrections

Please ensure that all abbreviations/acronyms are defined at point of first use (e.g. SLR, line 22; RNN, figure 1).

Table 1: check inversion parameter symbols vs names

Check formatting of citations, e.g. years not in brackets (e.g. L151), use of et al. vs and others.

L348: thinnning -> thinning

L390: remove extra "model"

Fig. 4: E_dhdt -> hdot_e?

Fig. 7: hdot_E -> hdot_e (probably worth checking consistency of symbols throughout)

L440: colormap for GLs is brown (as indicated by the figure 6 caption) not red.

L585: “is draft” -> “ice draft”?

All technical corrections will be implemented in the revised manuscript.

References:

Janssens I, Huybrechts P. (2000) The treatment of meltwater retention in mass-balance parameterizations of the Greenland ice sheet. *Annals of Glaciology* **31**:133-140. doi:10.3189/172756400781819941

Wernecke, A., Edwards, T. L., Nias, I. J., Holden, P. B., and Edwards, N. R. (2020) Spatial probabilistic calibration of a high-resolution Amundsen Sea Embayment ice sheet model with satellite altimeter data, The Cryosphere, **14**, 1459–1474, https://doi.org/10.5194/tc-14-1459-2020.