<u>Response to anonymous Referee #1.</u>

We thank the referee for the time spent on our manuscript and for these comments. We have realised that several aspects of our methodology were not so clearly explained and we believe that these comments will greatly improve the clarity of our manuscript. We believe that most of the points raised by the referee can be addressed by clarifying our method and we hereafter provide more details on how we have addressed these comments.

We have also made some changes that were not directly suggested by the referees but that will hopefully improve the article:

- The use of "emulation" has been generalised for the three methods instead of a mix of "extension", "reconstruction" and "emulation".
- We use the same weights for the 16-model ensemble (until 2100) and the 8-model ensemble (after 2100), which actually gives a more realistic representation of the ECS likely range than with equal weights after 2100 as implemented in the initial version, and gives more continuous results around 2100.
- We have removed the "approach" subsection that was initially presented in the results, to have all the methods described in section 2 and only the results of the ensemble of projections in section 3.

1) The distinction between "meltwater production" and "runoff" is unclear throughout the paper. This is especially the case as the authors investigate both runoff as a contribution to sea level changes and the "emergence of runoff conditions necessary for hydrofracturing". I find these two outcomes of 'runoff' a bit contradictory because if meltwater runs off into the ocean and contributes to sea level rise, then it can't induce hydrofracture events, which result from the pressure of ponded meltwater (i.e. Bell et al 2018). I think the authors should be careful to clarify this terminology throughout the paper. For example, in L27 the authors state: "The exact warming level needed to trigger important production of runoff on a given ice shelf depends on the amount of snowfall and on the snow/firn temperature and density (Donat-Magnin et al 2021, van Wessem et al., 2023)" However, these studies specifically look at meltwater production, which is the important component for hydrofracture on ice shelves, not runoff.

⇒ We acknowledge that our initial manuscript did not use a clear terminology and that some assumptions were not clearly stated. We have significantly modified the Introduction and the Methods section to make things clearer:

"Runoff is a negative contribution to the surface mass balance. It is produced if surface melt and/or rain rates are high enough to (i) percolate and bring the temperature of underlying snow and firn layers to the freezing point, (ii) saturate the pore space in the snow and firn layers, which is sometimes referred to as firn air depletion (Pfeffer et al., 1991; Kuipers Munneke et al., 2014; Alley et al., 2018; Donat-Magnin et al., 2021), and (iii) flow into the ocean. The liquid water beyond firn saturation, hereafter referred to as "liquid water in excess", does not necessarily flow into the ocean. Liquid water in excess can indeed alternatively form ponds or be transported horizontally within the firn or at the ice surface (Kingslake et al., 2017; Bell et al., 2018).

In some circumstances, the presence of liquid water in excess can trigger ice shelf break-up through hydrofracturing: in favorable conditions of ice shelf stress, the weight of liquid water can destabilize a fracture and lead to its unstoppable propagation as long as liquid water keeps filling the fracture (Weertman 1973; Lai et al., 2020). Stress variations associated with surface meltwater ponding and drainage, causing flexure and fracture, can amplify this mechanism and propagate its effects spatially (Banwell et al., 2013, 2019). An entire ice shelf break-up nonetheless likely requires large amounts of meltwater production all over its surface, as observed before the break-up of

Larsen A in 1995 and Larsen B in 2002 (Skvarca et al., 2004; Scambos et al., 2003; van den Broeke et al., 2005; Sergienko et al., 2005; Robel1 et al., 2019; Wille et al., 2022).

[…]

In the next section, we first use our methodology to estimate the SMB contribution to changes in sea level. As we will see in the next section, very warm conditions may lead to runoff production over the grounded ice sheet, as high as ~1000 m high, and a large part of the melt water not retained in the firn is expected to drain into ice shelves located downstream (Kingslake et al. 2017). We therefore assume that all liquid water in excess over the grounded ice sheet flows to the ice shelves downstream or directly into the ocean. Therefore, as far as the grounded ice sheet is concerned, the production of liquid water beyond firn saturation is a negative term in our SMB calculations.

We then use our methodology to estimate when the surface conditions have the potential to trigger ice shelf hydrofracturing. The relatively flat ice shelves are treated in a different way than the grounded ice. Indeed, in addition to the liquid water produced locally, the ice shelves receive the liquid water that was produced beyond firn saturation over the grounded ice sheet. To account for this, we assume that an ice shelf receives a fraction of the liquid water produced over the grounded ice of its drainage basin. The fraction is taken as the fraction of the basin coastline occupied by the ice shelf.

Another specificity of ice shelves is that they are relatively flat and can bend, so that it is impossible to estimate the amount of liquid water forming ponds or flowing into the ocean without a dedicated hydrology–firn–ice-shelf model. This is why we introduce an empirical threshold on the production rate of liquid water in excess to assess the potential for hydrofracturing. The idea is to have a rate that is sufficiently high to form ponds and fill crevasses even if a part flows into the ocean".

We have also moved the part about the ice shelf SMB to Appendix C, which helps clarify our two main objectives: (i) SMB (including runoff) over the grounded ice sheet and sea level, and (ii) production of liquid water beyond firn saturation and related potential for ice shelf hydrofracturing.

2) Methodology

There are many parts of the methodology that are not well explained and remain unclear. This makes it challenging to assess the robustness of the methodology and therefore the results. I suggest a graphical figure detailing the methodology.

⇒ This is a very good suggestion. We have added two figures, one with a schematic for each of the three emulation methods (Fig. 1 in the revised manuscript), and one for the actual methodology used to build the ensemble of simulations from 1850 to 2100 or 2200 (Fig. 6 in the revised manuscript).

Below are some specific sentences or sections I did not understand or believe require additional detail:

a) L109: "To extend surface variables to a given local warming or cooling level, we always start from 20 different years (i.e. different values of Tref), then we average the 20 extended values." What does this mean? Does this essentially create a smoothed reconstruction?

⇒ We hope that the new schematics (Fig. 1) and a slightly reworded text in section 2.2.1 will make this part clearer:

"In section 3, we emulate surface conditions for periods (Fig. 1a), scenarios (Fig. 1b), or CMIP models (Fig. 1c) that are not covered by existing MAR simulations. […] To emulate a

surface variable at a given warming or cooling level, we always calculate the emulation from 20 different years (i.e., different values of T_{ref} and ΔT), then we average the 20 emulated values (Fig. 1). This is done to better sample natural variability and to generate an emulated variability that is mostly related to the CMIP model temperatures. It also makes the emulation more robust from a statistical point of view".

b) L114: "The a and b parameters are obtained through a least-mean-square-fitting of an exponential curve for SMB minus runoff on the one hand and the surface melt rate of the other hand." A supplementary figure off this exponential curve would be very helpful.

⇒ We have added this figure in Appendix B as suggested. We have also moved a part of the previous text about this fit to the Appendix, which hopefully makes the Method section easier to read.

c) L115: "The fit is done on the original model grid as regridding does not preserve exponential relationships." I find this statement to be concerning. My understanding is that the exponential relationship between the two variables may weaken in the regridding but should remain? Further, if the exponential relationship parameters are fit on the original grid, and this relationship is not preserved in the regridding, is it then appropriate to apply this fit on the regridded data?

⇒ Yes, in principle, regridding weakens rather than removes the exponential relationship, but we have removed this statement because the simulations were interpolated conservatively from a 35 km grid to a 4 km grid, so the values on the interpolated grid are almost the same as on the original grid.

d) I am unsure how to interpret the r parameter (Eq. 3, 4, L121-124). Is this the percent of excess meltwater production that is converted to runoff (as opposed to that which ponds or refreezes)? If so, I expect that this value might be different on the grounded ice sheet vs ice shelves due to due to higher slopes on the grounded ice sheet.

⇒ This has been clarified using the new formulations (see 1st comment). The *r* parameter is a threshold over which liquid water is produced beyond firn saturation, which occurs when the melt rate exceeds what can be stored and refrozen in the ongoing snow/firn accumulation. Here we do not attempt to know whether the excess of liquid water forms ponds or flows directly into the ocean (runoff), which is why the same *r* value is used on the grounded ice sheet and on ice shelves.

e) Section 2.2.4: It is unclear how you reconstruct a cooler scenario from a warmer one. Do you use SSP5 to reconstruct SSP1? Or use warmer years as a reference time and reconstruct back in time? Also, what is the purpose of this and how will this be useful?

⇒ We have added schematic (Figs. 1,6) and have reorganised the Methods section to (*i*) better show the purpose of individual emulation methods, and (*ii*) better explain how these methods are combined in the final ensemble.

To answer this specific question: if we have SSP5 over 2101-2200, we mostly use it to emulate 2101-2200 under SSP1 because this is more accurate than emulating the entire period from 2081-2100. And it is just for 2101-2120 that we combine both methods in a ramping transition.

f) L157: "Similarly as in the previous subsection, each reconstructed year is the average of 20 reconstructions from a reference ranging from 10 years before to 9 years after the reconstructed year."

I don't fully understand this sentence and have had to read it several times. It is perhaps related to the point mentioned in a) above?

⇒ Yes, see previous response. This has been clarified through the schematics and the text.

g) Section 2.2.5: A graphic or schematic outlining the workflow here would be extremely helpful because I don't understand how the emulation is done. Additionally, Figure 3 is not very intuitive for me and should be better explained as I am unsure how to interpret it.

⇒ Yes, the workflow is now outlined in Fig. 6.

We have rewritten the caption of Fig. 3 (now Fig. 4) to make it easier to understand. This way of presenting the results is also found in other papers (e.g., Figs. 4 & 7 in Barthel et al., 2020).

h) Figure 4: In panels e and f, the reconstructed runoff anomaly is too low, despite the melt anomaly (panel c and d) being fairly accurate for ice shelves and too high over the grounded ice sheet. Does this suggest some mis-parameterization in your method? Perhaps the wrong r value?

⇒ The reconstructed runoff anomalies are both slightly too high in Fig. 4 (now Fig. 5), as indicated by the positive bias values indicated in brackets for panels c, d, e, f. If the referee was thinking about Fig. 2 (now Fig. 3) and not Fig. 4, there is indeed a positive bias in the melt anomaly and a negative bias in the runoff (i.e., excess of liquid water beyond firn saturation). Decreasing the *r* value would clearly reduce the runoff bias in this figure, but not in others. This indeed shows that the parameterisation is not perfect. We have added a comment on this.

i) L205-208: For these seven simulations… apply a ramping transition between the two methods from 2101 to 2120." I don't follow what is being done here and again, I think a figure or something would help the reader understand the methodology.

⇒ Yes, this has been better explained thanks to the addition of a schematic (Fig. 6), and the corresponding text has been rewritten.

j) L280: The choice of a 100kg/m2/yr runoff threshold for triggering hydrofracture seems extreme and is not well-defended in the text. This is 50-67% less than the average meltwater production estimated prior to the collapse of Larsen B (200-300 according to the text). How was this threshold "empirically" chosen? Was it just based on Larsen A/B?

⇒ We have improved this part of our manuscript by using a threshold distribution to better account for its uncertainty:

"Another specificity of ice shelves is that they are relatively flat and can bend, so that it is impossible to estimate the amount of liquid water forming ponds or flowing into the ocean without a dedicated hydrology–firn–ice-shelf model. This is why we introduce an empirical threshold on the production rate of liquid water in excess to assess the potential for hydrofracturing. The idea is to have a rate that is sufficiently high to form ponds and fill crevasses even if a part flows into the ocean. The average production of liquid water beyond saturation over Larsen B prior to its collapse was estimated between 200 and 300 kg m^{-2} yr^{-1} (Holland et al. 2011; van Wessem et al., 2016; Costi et al., 2018), so our threshold has to be lower than that. There is nonetheless a large uncertainty on the threshold and we sample it in a normal distribution of 150 and 61 kg m^{-2} yr^{-1} of mean and standard deviation, respectively. This is chosen to obtain 90% of the threshold values between 50 and 250 kg m^{-2} yr^{-1}. The lower

end of this range is chosen empirically so that not too many ice shelves are above the threshold in present-day conditions. The uncertainty on the threshold is hence included in the calculation of the probability of a given ice shelf to be over the threshold".

k) L291: How do you define "likely" or "very likely"?

⇒ This was indeed only defined in some table and figure captions. We have added the definition in the main text: "*we present our results as confidence intervals, which we define as in the IPCC reports: 17–83th percentiles for the likely range (66% probability) and 5–95th percentiles for the very-likely range (90% probability)*".

3) Figure 1

L148: It seems a bit of a stretch to say that the extension of the RCM simulation is suitable for 25 years over the grounded ice sheet…. Really, there is just one year anomalously high SMB year at ~2125. Otherwise, the original MAR simulation and the reconstruction have opposite trends, even during this first 25 years.

In general, Figure 1 is concerning for me. It seems that this reconstruction method cannot be applied in a warming climate. However, the authors do apply some sort of reconstruction to obtain the results in Figures 6-12. How were these reconstructions obtained when Figure 1 demonstrates issues for applying this method in a warming climate? How can we trust the results presented here in light of Figure 1?

⇒ First of all, to build our ensemble of projections, the emulation from another period is only used for the extension to 1850, i.e., to a colder climate, and in a 20-year ramp-down transition from 2101 to 2020 (towards a warmer climate). We believe that this is now much clearer in the revised manuscript, in particular thanks to the new schematics (Fig. 6).

Second, we have modified Fig. 1 (now Fig. 2) to evaluate both a backward and a forward emulation, and we have indicated the biases over 20 years. This confirms that the emulation backward (towards a colder climate) is less biased than the emulation forward (towards a warmer climate). The emulation forward nonetheless produces biases that remain small over the first 10 years, and reasonable over the first 20 years for $r = 0.5$ and $r = 0.6$ (~10% bias for the SMB over the grounded ice sheet and the production of liquid water in excess over the ice shelves, Fig. 2a,f). We also want to point out that our intention is not to assess the detail of the emulated interannual variability, but to describe the overall bias that could affect an ice-sheet model in the long term.

4) Relation to previous studies

In general, this manuscript is lacking some references to and context within recent literature. For example, how do the results in section 3.3 add to or fit within the context of previous ice-shelf potential instability studies (i.e. van Wessem et al., 2023; Dunmire et al., 2024; Alley et al., 2018; Lai et al., 2020)).

Additionally, some reverences to previous AIS SMB studies are missing (e.g., Gorte et al., 2020, Noel et al 2023).

⇒ We have added more comparisons to similar types of predictions. Some studies were published just before (Noel et al., 2023) or just after (Dunmire et al., 2024) our initial submission, but we can now include them. We have also included the recent projections by Veldhuijsen et al. (2024). All these references are now cited in our new section 3.2

(previously 3.3) together with references that were already cited (Kuipers Munneke et al., 2014; Donat-Magnin et al., 2021; van Wessem et al., 2023).

Regarding the other articles mentioned by the reviewer, Lai et al. (2020) was already mentioned as describing the mechanical conditions that can lead to hydrofracturing in the presence of meltwater beyond firn saturation. We have cited Gorte et al. (2020) for their reduction of uncertainty by weighting CMIP model based on their misfit with observations and Noel et al. (2023) for their novel downscaling method, but we do not compare our AIS SMB estimates to theirs because Gorte et al. do not account for runoff and because the projection of Noel et al. is based on a single CMIP model.

We have also added a paragraph in the Conclusion to explain what our projections bring compared to the aforementioned studies:

"*We believe that most original aspects of our projections compared to recent studies (van Wessem et al., 2023; Dunmire et al., 2024; Veldhuijsen et al., 2024) are (i) the time coverage back to 1850 and until 2200 while other studies cover ~1980-2100, (ii) the use of a large weighted ensemble of CMIP models to account for the models uncertainty while keeping a plausible equilibrium climate sensitivity, (iii) the relatively large number of MAR simulations used to assess and calibrate our simple emulator*".

5) Finally, the motivation for this work, and specifically how this method could be used in the context of ISMIP7, should be elaborated.

⇒ We have reworded a paragraph of our conclusion:

"Our method is useful to populate ensembles of surface mass balance and production rates of liquid water which are needed to constrain ice sheet model ensembles and to estimate the likely range of future sea level rise. This approach has been used to complete the set of RCM simulations used to drive ice sheet simulations until 2150 in the PROTECT European project (Durand et al., 2022; Mosbeux et al., 2024). This could also be useful for the upcoming ISMIP exercise as it is relatively simple to use, it can be applied before obtaining all the CMIP 6-hourly data and before running RCMs, and it can provide the information needed to trigger the hydrofracturing mechanism in ice sheet models".

Minor comments

L7: "After correcting the distribution of equilibrium climate sensitivity of 16 climate models…" From just the abstract, it is unclear what this means.

⇒ We have reformulated as *"After weighting 16 climate models to obtain a realistic distribution of the equilibrium climate sensitivity, we find…"*. We believe that the concept of "equilibrium climate sensitivity" is sufficiently known in the climate community to appear in the abstract, but we have provided a definition in section 2.2: *"the global mean surface air temperature increase that follows a doubling of atmospheric carbon dioxide"*.

L7: "… we find a likely contribution of surface mass balance to sea level rise of 0.4 to 2.2 cm from 1900 to 2010…". It does not make sense that the contribution of SMB to SLR would be positive for this period so I'm assuming this is with respect to a reference period? Same for the SLR contribution ranges in the following lines?

⇒ Embarrassingly, the entire sentence was wrong and has been replaced with:

"we find a likely contribution of surface mass balance to sea level rise of ***-2.2 to -0.4*** *cm from 1900 to 2010, and -3.4 to -0.1 cm from* ***2000*** *to 2099 under the SSP1-2.6 scenario, versus -4.4 to -1.4 cm under SSP2-4.5 and -7.8 to -4.0 cm under SSP5-8.5".*

L25: "Hydrofracturing may strongly enhance the contribution of upstream glaciers to sea level rise." This is a bit misleading. The papers cited (among other work) indicate that the removal/collapse of ice-shelves (perhaps due to hydrofracture events) causes a speed-up of upstream glaciers, not just the hydrofracture event itself.

⇒ We have reworded as:

"When occurring on ice shelf parts that buttress the upstream flow, hydrofracturing and the resulting ice shelf collapse may strongly enhance the contribution of upstream glaciers to sea level rise".

L41: "Because of these difficulties, only… which is generally insufficient to sample the CMIP model diversity." The "- when produced –" in this sentence threw me off a bit and I had to read it a few times to understand what was being said.

⇒ Agreed, we have removed this part of the sentence.

L43: "… correct unrealistic Equilibrium Climate Sensitivity…" A brief explanation for this concept would be helpful here.

⇒ We have removed the mention to the concept of Equilibrium Climate Sensitivity from the Introduction, and we now define it in section 2.2.

L45: "Over the years, Antarctic Ice Sheet modellers have often scaled their best estimates of present-day accumulation to temperature anomalies from the CMIP models…". This sentence fragment is unclear to me.

⇒ We have replaced with:

"Over the years, Antarctic Ice Sheet modellers have often scaled their best estimates of present-day accumulation to temperature anomalies from the CMIP models (e.g., based on the Clausius-Clapeyron relationship as in Gregory and Huybrechts, 2006)".

L67: "The surface mass balance and melting… Donat-Magnin et al (2020) and Kittel et al (2021)." I think a brief explanation of the results of these papers would be helpful here. I am left wondering: And how does MAR do in comparison to observational products?

⇒ In the new Appendix A, we have summarised the main results of the melt and SMB evaluations presented in previous articles and we have added a plot to further evaluate MAR's runoff in comparison to a satellite estimate of melt pond volume.

L101: Assuming that all precipitation is entirely made of snow is a big assumption to make, especially for projections that extend to 2200 in high-emission scenarios. The impact of this assumption should at least be discussed somewhere in the paper.

⇒ We already had this sentence:

"The presence of rainfall makes Eq. (3) more complex (see Appendix B of Donat-magnin et al. 2021), but rainfall is generally negligible compared to snowfall in present-day conditions and to surface melt rates in much warmer conditions (Donat-Magnin et al., 2021)".

We have added:

"*Furthermore, the only RCM simulation until 2200 (MAR–IPSL-CM6A-LR, SSP5-8.5) does simulate the actual rainfall/snowfall distribution and still has a positive surface mass balance over ice shelves in 2200 (see section 3), which indicates that snowfall is still dominant over rainfall".*

L132-134: Should this really be interpreted as a 'mass loss rate' if Figure 1 shows anomaly values with respect to a reference period? The SMB for the reference period is positive (although the specific reference period value from MAR should be mentioned somewhere in the paper). Even though the line in Figure 1b decreases throughout the timeseries, mass loss doesn't occur until it reaches the negative magnitude of the reference period. For example, if the reference period SMB for ice-shelves ~500 Gt/yr, then surface mass loss doesn't really occur until approximately 2120 (when the time series reaches -500 Gt/yr).

⇒ Yes, this has been reformulated as:

"the SMB in the original MAR simulation remains steady until ~2090 then drops due to increased melting with a SMB in 2200 that is 2000 Gt yr^{-1} lower than in 1995–2014".

L138: "… although this is still an improvement compared to the original IPSL-CM6A-LR outputs". I think it would be very interesting to have these original ESM timeseries plotted in Figure 1 as well.

⇒ We have added the SMB time series directly calculated from IPSL-CM6A-LR. See the black dashed lines in new Fig. 2a,b.

L123: "… covering the aforementioned range, i.e. 0.5 to 0.9." This range is different from that mentioned before (0.6-0.85, L98).

⇒ We have reformulated as:

"we will assess values in the 0.50–0.90 range, which includes the values used in previous work".

Section 2.2.3: Somewhere in this section it should be clarified that Equation 4 is used to do this reconstruction.

⇒ A reference to Eq. 4 has been added.

L179-181: "The realistic SMB reconstructions derived from MAR-ACCESS1.3 are mostly compensations between overestimated melt and overestimated accumulation". Why do you say this?

⇒ Because the melt rates emulated from MAR-ACCESS1.3 are largely overestimated, as seen by the large green pentagon in Fig. 3c,d, which means that the realistic SMB is due to underestimated accumulation. This is to argue that emulations from MAR-ACCESS1.3 are outliers despite realistic emulated SMB over the grounded ice sheet. A reference to Fig. 3c,d has been added to make this sentence clearer.

L210: "… with a 20-year transition." What does this mean?

⇒ This has hopefully been clarified, see previous responses.

L238: Figure 8 is mentioned before Figure 7

⇒ This has been corrected.

L253: What does "weaker SMB" mean?

⇒ This has been replaced with "lower SMB".

L265: "Spatially, a net surface mass loss arises for several ice shelves…" I find this to be a bit misleading since Figure 7 shows SMB anomalies from a reference period, not absolute SMB. Is there actually a net surface mass loss or is it just lower than the reference period?

⇒ Agreed, this has been reformulated.

L294 – It should be mentioned that George VI ice shelf has compressive stresses which do not promote hydrofracture occurrence (Labarbera et al., 2011)

⇒ We have added this statement and the suggested reference.

L 311: What is the A1B scenario?

⇒ This is a scenario that was used in CMIP3 and IPCC-AR4. This entire paragraph has been rewritten with no mention of the A1B scenario.

Technical corrections:

⇒ These have been corrected.

L33: progresses à progress
L62: Citation needed for the pore close-off density.
L180: "ooverestimated"
L185: "emulation" à emulations
L234: "scenario" à "scenarios"
L 254: End parenthesis after "(Fig. 9."
L262: "to the exception" à "with the exception"
Figure 8 caption: "same as Fig. ?? is also shown"

Response to anonymous Referee #2.

We thank the referee for the time spent on our manuscript and for these comments. We have realised that several aspects of our methodology were not so clearly explained and we believe that these comments will greatly improve the clarity of our manuscript. We believe that most of the points raised by the referee can be addressed by clarifying our method and we hereafter provide more details on how we have addressed these comments.

We have also made some changes that were not directly suggested by the referees but that will hopefully improve the article:

- The use of "emulation" has been generalised for the three methods instead of a mix of "extension", "reconstruction" and "emulation".
- We use the same weights for the 16-model ensemble (until 2100) and the 8-model ensemble (after 2100), which actually gives a more realistic representation of the ECS likely range than with equal weights after 2100 as implemented in the initial version, and gives more continuous results around 2100.

We have removed the "approach" subsection that was initially presented in the results, to have all the methods described in section 2 and only the results of the ensemble of projections in section 3.

General
This is an interesting paper that addresses an important problem: when can we expect runoff from Antarctic ice shelves, leading to mass loss and/or meltwater ponding which is commonly regarded as a precursor for ice shelf hydrofracture. The authors use multiple MAR simulations of future Antarctic climates, with which they calibrate a simple statistical model to emulate melt, runoff so as to approximate SMB and (instantaneous) firn saturation. Not considered are sublimation, rain and the time it takes to saturate the firn. When compared directly to MAR output, the emulator gives mixed results, yet provides valuable insights in the uncertainties that arise from the driving global models. The writing can be clarified in places (see below), the figures are generally clear.

Major comments

The abstract is at places confusing. It goes in one step from describing runoff emulation to AIS SLR contribution. Although it is stated that the mass loss numbers are based on SMB alone, the link that is made to hydrofracturing in the title and the presented numbers in terms of sea level rise could trick the reader into thinking that dynamical effects are also considered. It should be made clear from the outset is that this study treats runoff in two ways: as a source of mass loss and as a threshold to induce ponding and hydrofracturing (after which the dynamical effect on the ice sheet is not considered).

⇒ We have reworded the abstract to hopefully make our approach clearer. The introduction and methods have also been reworded and completed along these lines.

Also in the abstract: "Based on a more limited and uncorrected ensemble, we find a considerable uncertainty in the contribution to sea level from 2000 to 2200". It is unclear whether this enhanced uncertainty derives from the fact that the ensemble is smaller and uncorrected, or from the method, or all of these. Please only include major results in the abstract, the details can be elaborated upon in the text.

⇒ The sentence has been rewritten as:

"The contribution from 2000 to 2200 is highly uncertain: between -10 and -1~cm in SSP1-2.6 and between -33 and +6~cm in SSP5-8.5 depending on the model".

l.59: The statistical model is based on MAR data, which is a fair choice given that this model has been used for multiple future runs. But for the statistical model to be useful, MAR must provide reliable runoff. Here it is stated that the atmosphere in MAR is coupled to a 30 layer snow model "...representing the first 20 m of snow/firn with refined resolution at the surface". But in many places the firn layer in Antarctica is considerably deeper than 20 m? What does this imply for runoff simulated by MAR? See also my comments on treatment of percolation and runoff evaluation below.

⇒ Only using MAR is certainly a caveat but this was the only RCM for which we had an ensemble of projections sufficiently large to be robustly emulated. This caveat was already acknowledged in the Discussion section. We nonetheless agree that simulating 100 m of firn would be better than 20 m, although the potential presence of ice slabs makes it difficult to anticipate the exact effect. The firn depth is expected to change the timing of firn saturation/air depletion, but not the threshold for firn saturation which is more related to the melting and snowfall rates. We have added the sentence in red to the relevant paragraph in the Discussion section:

"*Our approach has consisted of emulating MAR simulations. Other RCMs, possibly combined to elaborated firn models, have similar skills in representing typical Antarctic conditions (Mottram et al., 2021), but there is likely a considerable spread in their response to surface warming. For example, the depth and vertical resolution of firn models probably make important differences in the timing of runoff production. The 20-m firn layer simulated in MAR thus likely reaches liquid water saturation earlier than models with a thicker firn layer. One of the next priorities will therefore be to emulate the diversity of RCM sensitivities, which would make the uncertainty ranges much more comprehensive.*"

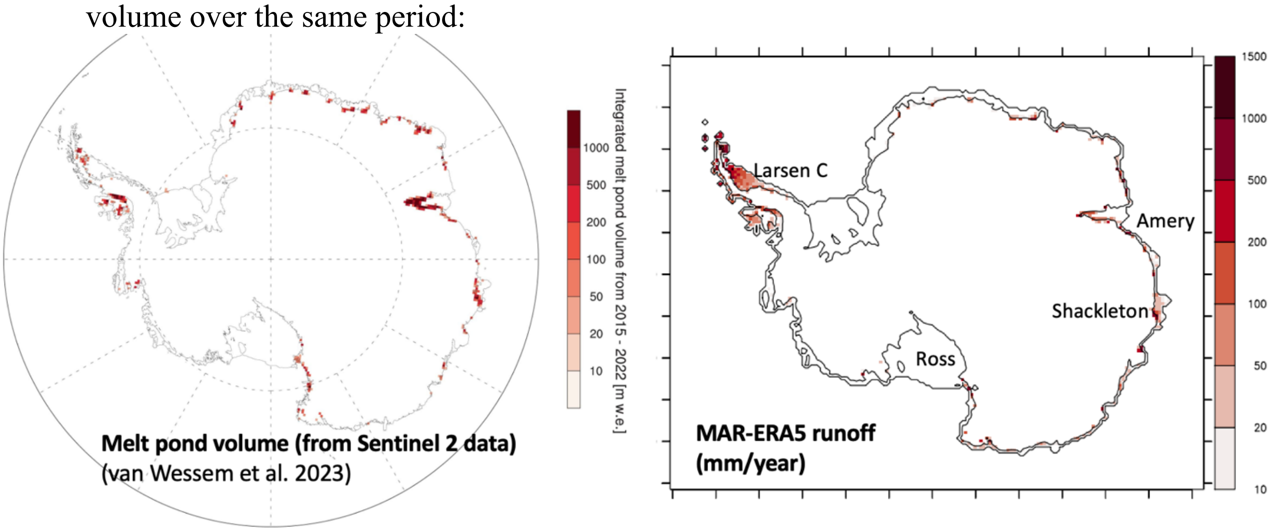
l.65: " If liquid water is not able to percolate further down, then it fills the entire porosity space of surface layers, and the excess is considered as runoff and removed from the snow/firn model (there is no representation of ponds or horizontal routing)". From this description it is not clear to me how water percolation interacts with ice lenses in the firn layer and how the process of filling up works before runoff starts.

⇒ This sentence has been clarified so that the entire paragraph reads:

"*In the presence of surface melting or rainfall, liquid water percolates downward into the next firn layers with a water retention of 5% of the porosity in each successive layer. The firn layers are fully permeable until they reach a close-off density of 830 kg m^{-3}. To account for possible cracks in ice lenses and moulins, the part of available water that is transmitted downward to the next layer decreases as a linear function of firn density, from 100% transmitted at the close-off density to zero at 900 kg m^{-3} and beyond. If liquid water is not able to percolate further down, it remains where it is. When the entire porosity space in the surface grid cell is filled with liquid water or if the surface grid cell is denser than 900 kg m^{-3}, any additional surface melt is considered as runoff and removed from the snow/firn model. There is no representation of ponds or horizontal routing.*"

l.67: "The surface mass balance and melting conditions produced by MAR have been evaluated in comparison to observational." Has there also been an attempt to evaluate the modelled meltwater ponding/runoff? I seem to remember that Van Wessem et al. (2023) compared their predicted ponding potential to satellite observations of melt ponds.

⇒ First of all, we have included more details on the comparisons to observations in this paragraph following a comment by Reviewer #1. Van Wessem et al. (2023) indeed presented a comparison of their RACMO simulations to an observational estimate of melt pond volume (derived from Sentinel 2). This comparison is qualitative as neither MAR nor RACMO simulate ponding (they remove the excess of liquid water and do not simulate horizontal transport of liquid water). We nonetheless show here the runoff produced by MAR forced by ERA5 over 2015-2022 in comparison the melt pond volume over the same period:



We find that the areas of high runoff in MAR generally correspond to the areas where high melt pond volumes are estimated from Sentinel 2 data, even if the area of high runoff over Larsen C is larger than the area of large melt pond volume in the satellite product. This comparison has been presented together with other results from published studies in Appendix A.

Figure 1 shows results for the integrated grounded ice/ice shelves only. How do the spatial patterns look?

⇒ We have added three figures in Appendiix C, one for each type of application of our evaluation method. The patterns are well reproduced.

Minor comments

l. 12: "Based on a runoff criteria.." Please note that 'criteria' is plural, so either use "Based on runoff criteria.." or "Based on a runoff criterion.."

⇒ This has been corrected.

l. 12: "we identify the emergence of surface conditions prone to hydrofracturing" Suggest: "we identify the timing of surface conditions that make ice shelves prone to hydrofracturing" or something similar.

⇒ This has been changed.

l. 13: " A majority of ice shelves could remain safe" Suggest to reformulate: "Our results suggest that the majority..."

⇒ This has been modified as suggested.

l.16: precipitation -> snowfall

⇒ Precipitation consists of snowfall + rainfall, and rainfall is a positive term in the surface mass balance that effectively contributes to mass gain until the firn is saturated with liquid water or until the surface consists of bare ice. But rainfall has a minor effect and this first sentence may be clearer with "snowfall", so we have corrected as suggested.

l. 18: However, if air temperatures exceed -> However, model simulations suggest that, if atmospheric warming exceeds

⇒ This has been modified as suggested.

l. 22: can be conducive of -> can be conducive to

⇒ This sentence has disappeared from the revised manuscript.

l. 31: sea level -> sea level change

⇒ This has been modified as suggested.

l. 36: in particular with regard to firn saturation by meltwater and subsequent runoff -> in particular with regard to firn saturation by meltwater and subsequent ponding

⇒ This has been modified as suggested.

l. 87: Eq. 1: At what level is the warming specified? For this expression to be accurate, the warming should be weighted over the atmospheric column in which the water vapour resides. This is different for Eq. 2 (l. 92), where near-surface warming can be used.

⇒ We have added this paragraph to clarify this point:

"Hereafter, we assume that ΔT is a variation in near surface air temperature in both Eq. 1 and Eq. 2, which is a reasonable approximation given that the troposphere warms relatively uniformly from the surface to ~300 hPa (Donat-Magnin et al., 2021, their Fig. 1)".

l. 103: surface -> near-surface (I assume)

⇒ This has been modified as suggested.

l. 108: " The m parameter is introduced to avoid unrealistically high melt rates" Can you provide a value, and how is it determined? Ah, it is provided in l. 120, but is given in kg m-1 s-1. Can you provide a more intuitive value, i.e. what this means per year?

⇒ We have added "1.80×10^{-4} kg m^{-2} s^{-1}, i.e., 15.5 mm/day" (this is also 5.7 m/year).

l. 109-112: This is unclear, please clarify.

We have modified the text and added schematics in two figures (now Fig. 1 and Fig. 6) to better explain our method.

l. 180: typo "ooverestimated"

⇒ This has been corrected.

l.231: " Our projections over the grounded ice until 2100 agree quite well with previous estimates of sea level contribution reported by the IPCC for the three scenarios". Can these numbers be compared, i.e. did these IPCC assessments also report estimates based on SMB-only?

⇒ Yes, the IPCC emulated SMB values from the AR5 to the SSP scenarios (IPCC-AR6-WG1 Tab. 9.3). The values were already reported in our initial Tab. 4 (now Tab. 3) which was referred to at the end of the quoted sentence.

Section 3.2: It would be informative to provide the modelled atmospheric warming of the various models in addition to the SMB and runoff. Main motivation is that some models project negative SMB over the grounded ice sheet in the 22nd century, which must require a very strong warming, as well as significant rainfall. For these extremes, how well does the initial assumption hold that rainfall remains small?

⇒ We have added maps of near surface warming in 2181–2200 for the two groups of models shown in what is now Fig. 11.

⇒ The assumption that rainfall still has a negligible role in much warmer climate is probably good given that:
- Melt rates increase much more than rainfall rates in warmer climate (Tab. 2 of Donat-Magnin et al., 2021 and Fig. 2 of Kittel et al., 2021).
- Melting is much more efficient than rainfall at saturating the firn with liquid water because melting removes snow/firn and converts it to liquid water while rainfall just adds liquid water without removing firn (Appendix B of Donat-Magnin et al., 2021).
- Furthermore, the only RCM simulation until 2200 (MAR–IPSL-CM6A-LR, SSP5-8.5) does simulate the actual rainfall/snowfall distribution and still has a positive surface mass balance over the ice shelves in 2200 (see section 3), which indicates that snowfall is still dominant over rainfall.

All this is mentioned in the revised manuscript.

l. 293: lead -> led

⇒ This has been corrected.

l. 294: Georges -> George

⇒ This has been corrected.

l. 323: "...while it can take more than a decade to saturate it if melt rates are just above the threshold..." Theoretically, it can take an almost infinite amount of time if the threshold is just passed.

⇒ Yes, this is exactly what we meant. We have replaced "more than a decade" with "an infinite amount of time".

l. 363: criteria -> criterion

⇒ This has been corrected.