Response to Reviewer 2

First of all, we would like to thank the reviewer for their time for reviewing this study. We appreciate the insightful feedback we received, which helps us to improve the readability and structure of the text, and to refine the presentation and interpretation of our results. Responses to the comments of the reviewers are written in **red** and citations of the manuscript are written in **blue**.

Kind regards, Sanne Veldhuijsen

This paper presents important new advances, both in terms of the firn modeling methods and the study of ice slab impacts on ice shelf hydrology. It particularly provides a valuable assessment of how firn meltwater storage will evolve in the 21^{st} century that comes closer than past studies to being grounded in our current understanding refreezing processes in firn. Overall, the work seems to be technically correct and well-reasoned. However, the paper can be a fit hard to follow at times due to the huge amount of work and information that went into the results. The authors could consider streamlining the main text and moving some of the model development work to the supplementary text.

Major Comments:

[1] I found the organization of Section 2-3 somewhat hard to follow, I think in part because the classic methods-results-discussion format falls short when trying to present what is the equivalent of almost two papers worth of work. You might consider first presenting a streamlined discussion of the model updates, calibration, and performance as one section. The goal of this section would be that the reader finishes it convinced that the FDM v1.2A-C run is a reasonable estimate of the future firn evolution. Right now, this point gets somewhat lost because the text is constantly bouncing back and forth between different experiments. Once this is established, then you can introduce how you will use FDM v1.2AD-C to study FAC across Antarctica. From there, it flows more naturally into the results that really only focus on FDM v1.2AD-C simulations. My suggestion for organization would be something like:

Section 2 – IMAU FDM Model Updates

2.1 – Densification expression

2.2 – Calibration Experiments and Atmospheric Forcing (combined since the experiments are just based on the different forcings)

2.3 – Calibration and Validation Data (aka firn cores)

2.4 – Calibration and Performance of FDM1.2AD-E (demonstrate that dynamic densification is reasonable using historical period)

2.5 – Performance of FDM1.2AD-C (demonstrate that future projections are reasonable)

Section 3 – Experimental setup for the future firn evolution and calculation of accessible firn air content

Section 4 – as it is now

I think that much of the current Sections 2&3 can also be streamlined to focus just on the key information needed to explain to the reader how the dynamical densification is achieved, and that the new model produces believable results for the rest of the 21st century. Maybe some of the detailed discussion of calibration results and extra figures for intermediate steps like FDM1.2AD-E could be moved to the supplement.

Thank you for these comments and suggestions. We agree that the paper contains a lot of information, and we can imagine that it can be difficult to follow. To improve upon this, we followed your suggestion by presenting all information of the model development together in Section 2 and the accessible FAC calculation description in Section 3, following the suggested structure. We focus in Section 2 on the development and evaluation of FDM v1.2AD-C and thereby moved some information to the supplement (see Sections 2 and 3 and the supplement of the revised MS).

Minor Comments:

Title: Perhaps should mention ice shelves, since the paper seems to almost entirely focused on discussing FAC change on ice shelves? That is indeed a good idea. Our new title is: "Firn air content evolution on Antarctic ice shelves under three future warming scenarios."

Line 11: Choose to describe the future climate scenarios either in terms of mitigation or in terms of emissions, but not both. It is quite confusing to keep track of whether "strong" means "strong emissions" under SSP8.5 or "strong mitigation" under SSP2.6. We agree and have decided to only use: "low, intermediate and high emission scenarios" to avoid confusion.

Line 159: Not clear why this is done? Maybe this would be better discussed in the experimental setup section? This is discussed in the new Section 2.2: Experimental setup and atmospheric forcing: "FDM v1.2AD-C is used to simulate future firn evolution over the Antarctic continent, and the remaining ones are used for evaluation (FDM v1.2A-E, FDM v1.2AD-E, Section 2.4) and assessing the impact on future firn evolution (FDM v1.2A-C, Section 4.5, tested for SSP5-8.5) of FDM v1.2AD."

Line 173 – 183: This paragraph did not seem particularly enlightening. It just reads like a long list of facts with not much explanation of why they are important, and certainly by the time it becomes relevant in the result, the readers will have forgotten all of this. If any of these statistics are important to the results, I would suggest either bringing them up at that point, or moving to a table and offering a super brief summary of the key points (e.g. accumulation, surface melt, and temperature all increase in the future over all parts of the AIS, with ice shelves seeing the largest increases). We agree, and followed your suggestions by leaving out these two paragraphs, since they are not directly used in the main results section. In addition, we have moved Figure 1 to the supplement.

Figure 1: Something to consider that might help readers visual the different experiments would be label on this plot which experiments map to which time periods with background shading or something. The multiple vertical axes are also confusing. Perhaps just break this up into two plots – one with temperature and one with melt and accumulation. To improve the readability of this figure, we will break it in subplots, and add a different background shading for the periods.

Line 198: Machguth et al. (2016) would be another appropriate citation here. We have included this relevant citation.

Section 2.5: A few additional suggestions that can help to bound the range of ice slab permeabilities that you use:

[1] Ashmore, D. W., Mair, D. W. F., & Burgess, D. O. (2019). Meltwater percolation, impermeable layer formation and runoff buffering on Devon Ice Cap, Canada. *Journal of Glaciology*, *66*(255), Article 255. https://doi.org/10.1017/jog.2019.80

Modeling studied which showed that an impermeability threshold of 1 m for ice slabs led to the best fit between modeled and measured SMB on the Devon Ice Cap.

[2] Charalampidis, C., Van As, D., Colgan, W. T., Fausto, R. S., Macferrin, M., & Machguth, H. (2016). Thermal tracing of retained meltwater in the lower accumulation area of the Southwestern Greenland ice sheet. *Annals of Glaciology*, *57*(72), Article 72. https://doi.org/10.1017/aog.2016.2

Used thermistor measurements to show that no percolation occurred through a 5.5 m thick ice slab at KAN-U even during the summer of 2012.

Note that using the numbers from Culberg et al. (2021) when looking at individual ice slabs is a bit complicated, because what they really show is that a package of many ice lenses that is 1-2 m thick can inhibit percolation to some degree. I think it's okay to use since the numbers are consistent with other papers and a 1-2m layer thickness or impermeability is a conservative take-away, but just good to note that their numbers are not totally comparable to some of these other studies on ice slab thickness. Thank you for these suggestions. We included Reference [2] in the figure as an additional observation and add the following: "In addition, firn temperature measurements in Greenland show that no percolation occurred through a 5.5 m thick ice slab even during an extreme melt year (Charalampidis et a. 2016)." Since Reference [1] is not an observation of actual ice slab thickness and depends on e.g. reanalysis forcing data/model configuration we decided not to include that one in the Figure, to not overcomplicate it. However, an impermeability threshold of 1 m is in line with the other findings. The comment about Culberg et al. 2021 is also how to interpret that study (in combination with Gascon et al. 2013), which we addressed in the prepint with: "While these large-scale radar observations do not give an exact relation between thickness and permeability they do give an indication that ice layers thicker than 0.5 m are at least partly impermeable on a larger scale." and "Henceforth, we refer to (a set of) ice layers that have a substantial impact on the accessible FAC as ice slabs."

Line 232: It's not clear where the > 900 kgm^{-3} threshold is coming from. Machguth et al. (2016) gives an ice slab density of 873 25 kgm^{-3} in their supplement and Rennermalm et al. (2021) gives a density of 862 kgm^{-3} (Machguth et al., 2016; Rennermalm et al., 2021). 900 kg/m3 comes from our own simulations, where we find that the near-surface refreezing layers generally have this value. If we use a lower threshold this impacts the accessible FAC calculations due to changes in high-density non-refreezing layers in the deep firn, which is not what we are interested in. However, we do acknowledge that layers with a density of >830 kg/m3 are usually defined as being impermeable. We rephrase this as follows: "Impermeable ice layers are usually defined as having a density > 830 kg/m3 (the pore close-off density). Here, we use a threshold of > 900 kg/m3 which corresponds to the density of near-surface refreezing ice layers within the model. This choice limits the impact on the accessible FAC of changes in high-density non-refreezing layers in the deep firn."

Line 273: I am confused about this statement about "higher FAC". It seems like the preceding sentence says that FAC decreases on average? To clarify this, we have rephrased this as: "This implies that, the effect of enhanced precipitation is overruled by the effect of projected warming."

Figure 6: I found it very hard to pick out the differences between the top and bottom row. Perhaps a third row with difference plots between rows 1 and 2 would be valuable. We have added a third row to highlight the differences between rows 1 and 2.

Lines 302 – 305: how do these thresholds compare to what we know about the climatic conditions for firn aquifers and ice slabs from either Greenland (MacFerrin et al., 2019; Munneke et al., 2014) or the Antarctica Peninsula (Van Wessem et al., 2021)? We compare this to ice slab conditions in Greenland in the discussion: "This is in line with MacFerrin et al. (2019), who found that ice slabs appear to be absent in regions of high accumulation (> 572 mm yr^{-1}) on the Greenland ice sheet." To also compare this to firn aquifer conditions, we add: "In contrast, high accumulation (>1000 mm yr-1) and relatively warm (>-19) conditions may lead to the formation of aquifers, which aligns with the absence of extensive refreezing (Munneke et al. 2014; Van Wessem et al., 2021)".

Figure 7c: Is the "difference between accessible firn air content and total firn air content" just a subtraction or is it a ratio? It is confusing to me why this would be positive for high melt, moderate accumulation ice shelves where my understanding is that accessible FAC would be lower that total FAC. We clarify in the figure caption how the difference is calculated, and we change the sign.

Lines 328 – 337: How is runoff defined and calculated in IMAU-FDM? I think this needs to be clarified for the reader. I am not fully convinced that the runoff extent or values are particularly meaningful given the caveat that ice slabs remain permeable in the model. The spatial extent of low accessible FAC seems like a far more valuable metric. How consistent is the spatial extent of low accessible FAC with the spatial extent of runoff as calculated by the model? Thank you for this suggestion, we add an explanation in the methods of how runoff is calculated: "Once meltwater has saturated the lowermost firn layer, we assume that it will leave the firn column as runoff instantaneously." We agree with your statement that runoff extent is not particularly meaningful given the caveat that ice slabs remain permeable in the model, which we also repeat in L329: "Since ice layers are fully permeable for meltwater percolation in IMAU-FDM, the runoff time series are closely related to the total FAC time series". However, it is difficult to come up with a threshold of what is low accessible FAC is (at what accessible FAC the firn gets saturated), as this depends on e.g. the annual melt of a specific region. For example, for the dry and cold Amery ice shelf, an accessible FAC of 3 m may hold all melt water, whereas this might not be the case on a wetter ice shelf such as Abbot. In addition, a certain amount of snowfall accumulates each winter, which also complicates this. To overcome this, we want to incorporate the ice slab thickness permeability relation in the meltwater percolation scheme of the firn model and asses the generated runoff in a forthcoming study, which is mentioned in the discussion: **"**Including the impermeability of ice layers interactively within IMAU-FDM will be tested in future work.**"**

Line 338: This seems surprising for an area that is currently a firn aquifer and is projected to see increased accumulation. Is this runoff coming from FAC depletion or from the bottom of an aquifer? It is indeed coming from the bottom of an aquifer, which we also address in Section 4.3: "In addition, runoff here also occurs year-round from firn aquifers, which are perennial subsurface bodies of liquid water, that become more ubiquitous in a warmer Antarctica (Bell et al., 2018)." We rephrase this in Line 338 as: "For Wilkins ice shelf we see a quick increase from 0 to > 90 % runoff extent for both scenarios, which indicates a limited spatial variation in firn state for grid points across the ice shelf."

Line 353: Please explain how you define an "extreme melt season". Consider marking these seasons in some way on Figure 9. We define an extreme melt season if it exceeds 2 times the SD of the detrended time series. When they cause a persistent reduction in accessible FAC in Figure 9 we mark these seasons by a shaded area.

Line 414: Seems appropriate to at least mention something about firn aquifers, though I understand that this is not the focus of this study. We agree and have added a small discussion about firn aquifers at the end of this paragraph: "~~These results highlight the different response of ice shelves with low- and high accumulation rates to atmospheric warming.~~ In contrast, high accumulation (>1,000 mm yr^{-1}) and warm (>-19C) conditions may lead to the formation of aquifers (Kuipers Munneke et al. 2014, Van Wessem et al. 2021), which aligns with the absence of extensive refreezing. Although firn aquifers do not have a depleted (accessible) FAC, the runoff from aquifers have the potential to cause hydrofracturing. Therefore, we will explore the future expansion of aquifers in a forthcoming study."