**Response to Reviewer 3:**
First of all, we would like to thank the reviewer for his time for reviewing and editing this study. We greatly appreciate the insightful comments raised by C. Max Stevens and respond to each of them below. Based on your comments, we have improved the clarity of the text, especially Sections 2.6 and 6, we have made section 5.2 more quantitative and included a discussion on implications.

Responses to the comments of the reviewers are written in red and citations of the manuscript are written in blue.

Kind regards,

Sanne Veldhuijsen

**General comments:**
1. The sections (2.6, 6) describing the sensitivity tests need to be clearer – I am still not sure exactly what you did for them. For example, what exactly are you adding and subtracting for those runs? Are you, e.g., adding 8% to the accumulation over the entire forcing, and that is one of the sensitivity runs? It would help to clarify if you wrote exactly what you did to perform these tests, e.g. "We added 8% to the accumulation forcing, and then re-ran our model calibration procedure to get the optimal MO fits (or whatever) with that forcing. We then ran the model with these MO fits and compared the outputs to our baseline model run". (Or something along those lines). I think what I just wrote is the gist of what you are doing, but this is what I need clarified. The distance between sections 2.6 and 6 made this analysis more difficult to understand: I was looking for a table describing the sensitivity tests when I read section 2.6, but did not find it until the end (Table 4). I would also appreciate if table 4 was more specific: e.g., I can figure out what Accumulation+ is from looking at section 6, but it would be nice to have all the information in the table. Are you only doing these sensitivity runs for the observational sites, or did you run the model for the entire ice sheet in any of the sensitivity runs? When you say "10 additional locations" (line 198), what are those in addition to? Are those 10 sites special, or were they not used in the original calibration (if so, why not)? Finally, when doing these sensitivity tests, if you are adjusting the MO fits (section 6), with an empirical firn model, shouldn't you expect the model outputs to match the data reasonably well? This is indeed what you found, but my point is that if you are tuning a firn model using e.g. biased accumulation fields, the firn model will effectively act as a filter for that bias (or, the correction for it will be built into the model), so that the model output will match the data well. I think the upshot of my question is that I would like to see a bit more discussion contextualizing the meaning of the model sensitivity. Is the new model an improved representation of the physics compared to the previous IMAU FDM, or is it just a numerical response based on using different boundary conditions (RACMO forcing and surface density)?

Thank you for these comments. We agree that the description of the sensitivity analysis should be improved. We suggest replacing Table 4 with Table 2 (in Section 2.6) and Table 5 (in Section 6) to improve the link between these sections (See tables shown below).

**Table 2.** Overview of the sensitivity experiments. The inputs and parameters are adjusted by substracting and adding the uncertainties either during the entire run or only during the spin up. The values used are based on the references provided or on analysis within this work.

| Experiment name | Variable | Variation | When | Reference |
|---|---|---|---|---|
| MO fits | MO fits | + and -95 % confidence interval | entire run | Figs. 3a,b |
| $\rho_s$ | $\rho_s$ | + and -30 kg m$^{-3}$ | entire run | RMSE evaluation Section 3.1 |
| $T_s$ | $T_s$ | + and - spatial variable K | entire run | $\sigma$ from Carter et al. (2022) |
| b | b | + and - spatial variable % | entire run | $\sigma$ from Carter et al. (2022) |
| spin up $T_s$ | $T_s$ | +0/+1 K and -1/-0.5 K[a] | spin up | Stenni et al. (2017) |
| spin up b | b | +0/+10% and -10/-5 %[a] | spin up | Thomas et al. (2017) |
| spin up $T_s$, $b$ | b and $T_s$ | +0/+5 %, +0/+0.5 K and -10/-5 %, -1/-0.5 K[a] | spin up | Stenni et al. (2017); Thomas et al. (2017) |

[a]The values on the left side of the slash indicate the variation for the Peninsula and Ellsworth Land, and the values on the right side of the slash for the remainder of the AIS. Temperature and accumulation for the Peninsula and Ellsworth Land are the second last loop of spin up varied with 6.66 % and 0.66 K and in the last loop with 3.33 % and 0.33 K.

**Table 5.** Overview of the change in annual average firn air content and surface elevation change of the extrapolated experiments and sample locations experiments. The sample location experiments also include the experiments without adjusted MO fits. An overview of the experiments including the prescribed uncertainties are listed in Table 2.

| Experiment name | Extrapolated to entire ice sheet | | Sample locations (without adjusted MO fits)[a] | |
|---|---|---|---|---|
| | FAC (%) | dH/dt (mm $yr^{-1}$) | FAC (%) | dH/dt (mm $yr^{-1}$) |
| MO fits | 5.1 | 0.48 | 5.2 | 0.75 |
| $\rho_s$ | 1.2 | 1.46 | 0.7 (5.8) | 2.05 (2.44) |
| $T_s$ | 1.8 | 0.66 | 2.3 (4.9) | 0.65 (1.50) |
| b | 4.0 | 5.63 | 3.7 (6.8) | 7.79 (7.99) |
| spin up $T_s$ | 0.5 | 4.40 | 0.5 (1.5) | 5.16 (5.22) |
| spin up b | 0.7 | 18.96 | 0.7 (1.1) | 22.45 (22.40) |
| spin up $T_s, b$ | 0.7 | 14.93 | 0.6 (0.8) | 17.45 (17.38) |

[a]The values between brackets indicate the uncertainties of the experiments without updated MO fits.

We propose to clarify the experiment description of Section 2.6 as follows:

"To study the model sensitivity to input and parameter uncertainty we performed a sensitivity analysis for the 106 locations shown in Figure S1a. These locations include 95 locations from firn core sites (the firn cores used are located at 95 unique grid points, excluding 4 sites that were later added for evaluation). To improve the representativeness of the climatic conditions over the AIS, we used 11 additional locations in this analysis that are located in high accumulation or low temperature regions, as these areas are underrepresented in the 95 firn core sites (Figs. S1b,c)

All the performed experiments are listed in Table 2. Verjans et al. (2021) and Lundin et al. (2017) showed that differences in accumulation, temperature, firn densification formulation and fresh snow density can lead to a substantial spread in modelled firn thickness and air content. For each of these components, we separately performed a model sensitivity test, which includes the spin-up as well as the main run. The fresh snow density was varied with the RMSE from the evaluation with in situ measurements (30 kg $m^{-3}$; Section 3.1). To test the sensitivity to uncertainty of the dry snow densification rate, we perform a simulation where we use the 95 % confidence interval boundaries of the MO fits (Eqs. 4 and 5; Fig. 3). To test the sensitivity to uncertainty in accumulation and temperature forcing, we use the spread of snowfall and temperature among regional climate models and re-analysis products (Carter et al., 2022). This spread varies spatially and we vary the accumulation, with the ratio of the ensemble standard deviation to the ensemble mean of each location, and vary the temperature with the ensemble standard deviation.

Another important source of forcing uncertainty is the climate forcing during the spin-up period. As explained in Section 2.2, the spin-up forcing is obtained by looping over the 1979-2020 forcing data. However, firn core (Thomas et al., 2017) and isotopes (Stenni et al., 2017) studies show that in the Antarctic Peninsula and Ellsworth Land, the accumulation and temperature were typically about 10 % and 1 K lower during the last centuries. In addition, over the remainder of the AIS, the accumulation and temperature were typically about 5 % and 0.5 K higher or lower during the last centuries. To investigate the typical effect this has on the IMAU-FDM results, we include schematic sensitivity tests in which we adjusted the accumulation and temperature forcing with these values only during the spin-up period. To mimic a gradual increase in precipitation over the Antarctic Peninsula and Ellsworth Land, we create a spin-up experiment in which the precipitation up until the third-to-last loop (the 42-year reference period) is decreased by 10 %, in the second-to-last by 6.66 %, and in the last one by 3.33 % with respect to the mean precipitation of 1979-2020. To mimic a gradual increase in temperature over the Antarctic Peninsula and Ellsworth Land, we create a spin-up experiment in which the temperature up until the third-to-last loop is decreased by 1 K, in the second-to-last by 0.66 K, and in the last one by 0.33 K. Over the remainder of the AIS accumulation and temperature were both decreased and increased with 5 % and 0.5 K during the spin-up period. As the uncertainties between temperature and accumulation during the spin-up period are dependent, we also performed a test in which we simultaneously adjusted the temperature and precipitation. For all our experiments listed in Table 1, except for the MO fits, we re-ran our model calibration procedure to get the optimal MO fits. We then ran the model with these MO fits and compared the outputs to our reference FDM v1.2A run."

In addition to this, we will contextualize the model sensitivity, notably the role of MO fitting in reducing the sensitivity compared to without updated MO fit. Instead of using uniformly changed precipitation (-/+ 8%) and temperature (-/+ 1.5K) in our revised manuscript we use spatially variable perturbations in the sensitivity experiments. We find a larger sensitivity of the simulated FAC (2- 4%) than with the uniformly changed precipitation and temperature (<0.5%), but still smaller than with the original MO fits (4-7% FAC).

2. In Figure 9 and 5.2: It seems to me that you have taken a rather qualitative approach here in looking at the elevation changes (e.g., "shows comparable patterns", "agreement seems"). I am curious how you have chosen to plot these time series: what is your zero/reference point? To me it would make most sense to plot them all zeroed at the start of the time series (1992) to see a direct comparison of how the data and model differ. The way it is plotted presently is misleading in a few cases: e.g. for site 6 (West Antarctica), the curves appear to be lined up reasonably well, but upon closer examination the data show a clear positive elevation trend while the model predicts a net decrease in elevation. This would enable a more quantitative analysis, such as fitting a trend line to each time series to get the long-term trends. Then, you could de-trend and do a time-series correlation for each site to quantify how well the models predict the shorter time-scale variability.
Thank you for this recommendation. We improved the clarity of the text, thereby highlighting the quantitative part of 5.2 (See our response to comment 42). We followed your suggestions and started each time series at zero in 1992. We acknowledge that the plot is indeed rather qualitative, but it serves as an addition to Figure 8 in order to show the temporal detail. In addition,we include results of the FDM v1.1p2 run, which was not performed for the entire ice sheet due to computational demands, and could therefore not be included in Figure 8. Including this run is important as it tells us that the difference in surface elevation change between FDM v1.2A and FDM v1.1p1 is predominantly due to the updated forcing. We have not made this figure more quantitative as these are only 11 out of ~18,000 locations, and the presented analysis is applicable to the entire ice sheet. In line with this, the trends of the time series are already covered in Figure 8. In addition, we do not focus on the short term-variability in Figure 9, which is done in Section 5.1 for the entire ice sheet, and therefore we also use a 6-month moving average window.

We agree that a more detailed temporal variability analysis (although for the entire ice sheet and not the locations in Figure 9) by separating seasonal, long-term trends, shorter time (yearly/decadal) variability would be very interesting, but this is not done in this study, as this is beyond the scope/to limit the size of the paper.

In response to another reviewer, we have added results from the sensitivity analysis in this plot.

3. Discussion about implications: My general feeling in the paper what that it was light on discussing the implications of the research in the broader glaciological community. Given how much IMAU-FDM has been used for altimetry studies, I think adding a paragraph or a few sentences describing how the model changes affect our understanding of how the AIS is changing. Do the conclusions drawn by the users of IMAU-FDMv1.1p1 need to be updated?
Thank you for this recommendation. We propose to change Section 7: Remaining limitations and outlook, into: Implications, remaining limitations and outlook. We propose to add:

"IMAU-FDM has been used to correct altimetry observations in mass balance studies (e.g., Adusumilli et al., 2018; Willen et al., 2021). The difference in firn density change over time of FDM v1.2A compared to FDM v1.1p1 has an impact on mass change estimates in altimetry studies. Over the period 2003-2019, Smith et al. (2021) found that the total mass change of East Antarctica amounts to 195 Gt $yr^{-1}$, and of West Antarctica to -245 Gt $yr^{-1}$. When FDM v1.2A is used instead of FDM v1.1p1, the mass change over the period 2003-2015 for East Antarctica is 7.3 Gt $yr^{-1}$ lower, and for West Antarctica this is 1.2 Gt $yr^{-1}$ higher."

In addition to this, we will also add a discussion about the different impact of the updated forcing (almost completely responsible for the surface elevation change) and model parameterizations (mainly an impact on average firn density profiles).

**Line by line comments:**
4. line 44: Can give a sentence with a broad overview of what the 2 types are for readers not familiar with the difference?
We suggest to add:
"Physics-based models describe densification using a constitutive relationship between stress and strain for snow, while semi-empirical models use physics based densification equations in combination with parameters that are tuned according to observational density profiles."

5. line 47: "less": change to "fewer"

Based on feedback from reviewer 1, we suggest to change:"semi-empirical models require less poorly known parameters" into: "semi-empirical models require a smaller number of poorly constrained parameters "

6. Table1/section 2.1.1 leaves me wondering what surface density was used for run FDM v1.2A. Perhaps change Table 1 to include a 'Fresh Snow Density' column?
We agree that that information was missing. We suggest to update the table as follows:

**Table 1.** Abbreviations and characteristics of IMAU-FDM versions used in this study.

| Abbreviation | IMAU-FDM version | Forcing | Fresh snow density | $MO_{830}$ fit |
|---|---|---|---|---|
| FDM v1.2A | FDM v1.2 Antarctica | RACMO2.3p2, ERA-5 | Eq. (2); this study | Power (Eq. 5) |
| FDM FS-K | FDM v1.2 Antarctica | RACMO2.3p2, ERA-5 | Eq. (1); Kaspers et al. (2004) | Power (Eq. 5) |
| FDM FS-L | FDM v1.2 Antarctica | RACMO2.3p2, ERA-5 | Eq. (2); Lenaerts et al. (2012) | Power (Eq. 5) |
| FDM v1.2A-log | FDM v1.2 Antarctica | RACMO2.3p2, ERA-5 | Eq. (2); this study | Log (Eq. 4) |
| FDM v1.1p1 | FDM v1.1 | RACMO2.3p1, ERA-Interim | Eq. (1); Kaspers et al. (2004) | Log (Eq. 4) |
| FDM v1.1p2 | FDM v1.1 | RACMO2.3p2, ERA-5 | Eq. (1); Kaspers et al. (2004) | Log (Eq. 4) |

7. Section 2.1.2: Do you use the instant accumulation rate or 'mean accumulation over the lifetime of the firn layer'? If the former, please provide detail as to why you choose this and deal with the fact that densification will be zero if there is a timestep with zero accumulation.
We have used annual average accumulation rate, hence the dot above the b, which we define in line 84. For clarity, we have repeated the definition of $\dot{b}$ in this section, as explained in the comment below.

8. line 106/Equation (3): the Arrhenius factor is missing the e. Also (and this is a bit pedantic), $\dot{b}$ in Arthern et al. (2010) has units of kg $m^{-2}$ $a^{-1}$ , which then gives a densification rate dp/dt with units of kg $m^{-3}$ $a^{-1}$. (Arthern defines the units for the factor D in their Appendix B). Your accumulation units (mm w.e $a^{-1}$) are numerically equivalent, but with them the densification rate in your Equation 3 does not end up with units of density per time.
Thank you for noticing this. We have added the exponent term. In addition, we suggest to add: "$\dot{b}$ is the annual average accumulation rate (kg $m^{-2}$ $a^{-1}$)".

9. lines 114-122: It is not entirely clear what you mean with $MO_{550}$ and $MO_{830}$: is $MO_{550}$ the value that should be used for $\rho < 550$, and $MO_{830}$ for $550 < \rho < 830$? Also, perhaps specify that MO added as a multiplier to equation (3).
This can indeed be improved. We suggest replacing this: "By comparing simulated and observed depths (m) of the 550 and 830 kg $m^{-3}$ density levels ($z_{550}$ and $z_{830}$, respectively) Ligtenberg et al. (2011) found that Eq. (3) requires a correction function, the so-called MO fits (ratio of Modelled to Observed depths), which turn out to depend on the accumulation rate. These MO fits are defined using the ratio between modelled and observed values of $z_{550}$ and 115 $z_{830*}$, where $z_{830*} = z_{830} - z_{550}$, using a simulation in which Eq. (3) is used without correction fits. These MO fits are defined using the ratio between modelled and observed values of $z_{550}$ and 115 $z_{830*}$, where $z_{830*} = z_{830} - z_{550}$, using a simulation in which Eq. (3) is used without correction fits. Ligtenberg et al. (2011) and Brils et al. (2021) used logarithmic correction functions, thus:"

By: "By comparing simulated and observed depths (m) of the 550 and 830 kg $m^{-3}$ density levels ($z_{550}$ and $z_{830}$, respectively) Ligtenberg et al. (2011) found that Eq. (3) requires correction terms $MO_{550}$ for $p < 550$ and $MO_{830*}$ for $550 < \rho < 830$, which are defined as the ratio of modelled and observed values of $z_{550}$ and $z_{830*}$, where $z_{830*} = z_{830} - z_{550}$. The correction terms $MO_{550}$ and $MO_{830*}$ are added as a multiplier to Eq. (3); MO values below one reduce the densification rate, values above one enhance the densification rate. $MO_{550}$ and $MO_{830*}$ are chosen as functions of the long term mean accumulation rate. Ligtenberg et al. (2011) and Brils et al. (2022) used logarithmic correction functions, thus:"

10. line 122: Add a sentence at the end of this section explaining that the parameters are tuned, and this tuning is described in Section 3.
We suggest to add: "The fitting parameters in Eq. (3) and (4) are tuned, which is described in Section 3.2".

This comment is also relevant for section 2.1.2 (fresh snow density), therefore we suggest to add:
L102: This retuning is described in Section 3.1".

11. Section 2.2: Do you set the bottom of your model domain to be the depth of the 830 horizon, or does it go deeper? In some locations in Antarctica, there is enough of a temperature gradient at the bottom of the firn to affect the temperature – if you are not modeling through the depth of the ice sheet, do you account for this heat flux?
In IMAU-FDM we do not set the bottom of the firn layer at 830 kg/m3. The lower boundary conditions for the thermal conduction are described in Section 2.3.1 of the revised manuscript.

12. line 142: not clear – why is the air content increasing?
We explain this as follows: "In the actual simulation after the spin-up, a minor trend (of <0.6 mm/yr averaged over the AIS) in total firn air content remains, because at the bottom of the column, ice with a density between 830 - 917 kg/m3, slowly replaces dense ice (ρ = 917 kg/m3), which originates from the initialization of the firn column prior to the spin-up."

13. line 160: change semicolon between references to "and"
We changed this, thank you for noticing.

14. line 181: This is confusing – at the beginning of the paragraph you state that you use 125+8 density profiles, but then here you say you used 122. I think that a bit of language clarification in this paragraph will help – something along the lines of, 'We gathered data from 125 density profiles from firn cores and 8 density profiles from neutron…'. Then, '104 of our 133 cores fit the dry snow criteria needed for our MO-fitting routine.' Or something along those lines adding a bit more specificity.
This confusion comes from the fact that some density profiles are only used for the surface snow density tuning/evaluation.

Based on feedback of all reviewers, we propose to rewrite this part as follows:
"To tune and evaluate the firn model, we collected published firn density profiles from widely varying locations across the AIS (Fig. 1). We used 115 density profiles from firn cores and 8 density profiles from neutron density probe measurements, combining multiple datasets (van den Broeke, 2008; Schwanck et al., 2016; Bréant et al., 2017; Fernandoy et al., 2010; Montgomery et al., 2018; Fourteau et al., 2019; Olmi et al., 2021; Winstrup et al., 2019). The firn cores and neutron density probe measurements are mostly obtained in summer months between 1980 and 2020. For tuning the fresh snow density parameterization, observations of the density of surface snow, defined as the top 0.5 m, from 63 firn cores could be used. Eight additional surface snow density values from density profiles of neutron density probe measurements and two from firn cores (Montgomery et al., 2018) were added after the tuning, and are thus only used for evaluation. For tuning the MO fits, only dry firn cores are used, as Eq. (3) describes dry snow densification only. A firn core is considered dry if its location experiences on average less melt than 5 % of the average annual accumulation. For the MO fits, 92 dry firn cores could be used, which improves upon Ligtenberg et al. (2011), who used data from 48 dry firn cores. In addition to the 94 dry cores used for the tuning of MO ratios, another 12 wet cores were used to evaluate the $z_{550}$ and $z_{830}$ model output. For evaluating FAC output, 31 firn cores could be used. Table S1 in the Supplement lists all measurements that have been used, the corresponding coordinates, method and citation."

15. line 182: This sentence has the word 'density' 3 times – can you rewrite it to make it a bit easier to read?
We suggest to change: "Eight additional surface snow density values from density profiles of neutron density probe measurements".
Into:
"Eight additional surface snow density values from neutron density probe measurements".

16. line 184: In some locations, this 0.5 m is likely snow (i.e. less than a year old), but in the interior it is several years of accumulation. Is this a concern, either for tuning or for interpreting model outputs?
For tuning and interpretation, we compared simulations and observations that are both of the upper 0.5 m. Furthermore, densification is still limited in the majority of the locations as either the time for densification is very short, due to the high accumulation, or the densification goes slow due to the very low temperatures. Therefore, we think this is not a concern for interpreting model outputs.

17. line 203/205 (elsewhere too?): fix the +- to +/- (may be done in manuscript typesetting?)
Thank you for noticing this, we corrected this.

18. line 219: Reference the equations from earlier section.
We added this.

19. line Table 2: It appears that the letters A,B,C,D do not correspond to equations 1 and 2 (e.g. the Lenaerts equation (2) does not have a D, but Table 2 lists a D). This makes the arguments in section 3.1 difficult to follow (though it appears that C is consistent).
Thank you for noticing. We improved, eq. (1) and (2) now have coefficients: A, B, C and E.

20. line 221: The MO fit bit here is distracting – move to end of section. Also, "impact of the different MO fits on the surface snow densities is negligible" – shouldn't there be zero impact, because the surface snow density is not a function of the MO? I.e., changing your surface density scheme will change the MO values after tuning, but changing the MO values will not change the surface density.
The fresh snow density is independent of the MO fit, however the surface snow density (top 0.5 m) has undergone some densification, so this is in fact dependent on the MO fit (although negligible), as mentioned above. We agree that this is confusing in this part of the section, so therefore we have moved this towards the end to avoid confusion.

21. line Section 3: I assume that the r2 statistic is calculated by regressing the modeled surface densities against the 1:1 line – is that true?
No, it is calculated by evaluating the scatter of the data points around a fitted regression line, because with this statistic we want to represent how well the model captures the spatial variation instead of how close it is to the 1:1 line.

22. line 227: You say: "This aligns with the fact that IMAU-FDM does not include densification by wind packing." I would agree that the subsurface densification scheme does not include wind packing as a densification, but that is also true in reality, isn't it? But, if you are using any of the Kaspers, Lenaerts, or modified Lenaerts equations in IMAU-FDM to set the surface density, I would suggest that wind packing is implicitly included in IMAU-FDM because those equations do include a wind term to account for wind packing.
For FDM with Kaspers and modified Lenaerts, this is indeed true. But the Lenearts parameterization is calibrated with fresh snow density observations of the skin layer, so wind-packing has likely not yet occurred there. This explains the underestimation with Lenaerts, and e.g. not with Kaspers.

23. line 230: "reduced with": do you mean "reduced by"?
Thank you for noticing this, we corrected this.

24. line 231: You provide a specific number that is the surface density reduction (i.e. 18 kg $m^{-3}$), which implies that the density was reduced by that everywhere; but then you point to different regions and imply that the density changes are different in those regions. I think you mean something along the lines of, "On average over the AIS, the surface density was reduced by 18 kg $m^{-3}$. The surface density decreased more in the high accumulation margins and less in windy escarpment regions." However you rephrase it, please be sure to use careful language to indicate what less and more mean, since you are dealing a with a reduction (i.e., does "less" mean reduced by 30 rather than 18, or reduced by 10 instead of 18?).
We agree that this might be confusing, therefore we suggest changing this:
"In general, the surface snow density is reduced with 18 kg $m^{-3}$, especially in the high accumulation margins, and to a lesser extent in windy escarpment regions (Fig. 2b)."

Into: "Overall the surface snow density is reduced with 18 kg $m^{-3}$, most notably in the high accumulation margins, and to a lesser extent in windy escarpment regions (Fig. 2b)."

25. line 235: remove comma after v1.2G
Done.

26. line 237: The idea that the surface density in Greenland is a function of annual temperature seems rather unphysical to me – I think that our knowledge of snow science is adequate to state with confidence that in reality, local conditions on relatively short temporal and spatial scales are the determinates of snow density (which is consistent with what you are finding in Antarctica). I suggest that the discrepancy is probably more due to lack of appropriate data in Greenland, and not significant differences in climatic conditions (there are many locations in Greenland and Antarctica that have very similar climates – so why in Greenland would the density instead be determined by the previous year's climate?).
We agree. We suggest changing this part: "The fresh snow density parameterization for Greenland used in FDM v1.2G, is only a function of yearly temperature (Brils et al. (2021). This is in contrast with Antarctica, where we found a strong dependency with instantaneous wind speed and temperature, which is in line with previous work

(Keenan et al., 2021; Lenaerts et al., 2012; van Kampenhout et al., 2017), and likely owing to the larger range in temperature and wind speed conditions during snow deposition in Antarctica."

Into: "The fresh snow density parameterization for Greenland used in FDM v1.2G, is only a function of yearly temperature (Brils et al. 2022), owing to a lack of co-located surface snow density, temperature and wind speed observations on the GrIS. This contrasts to the expression used here, which includes a dependency on instantaneous wind speed and temperature."

27. line 240: This is inconsistent – the first sentence of the section says you ran IMAU-FDM without MO fits (does this mean with the original Arthern equation?), and then then next sentence talks about the resulting MO fits.
Yes, this is indeed the original Arthern equation. To clarify, we suggest to replace this sentence by: "To tune the dry snow densification rate we first performed a simulation of FDM v1.2A without MO corrections. The resulting MO fits and statistics are listed in Table 4 and shown in Figures 3a,b."

28. line Table 3: Is there an error for your updated α value for $MO_{550}$ FDM v1.2A? It is 1000 times the others listed.
Thank you for noticing this, this should be 1.228 instead of 1288, we corrected this.

29. line Table 3: Why do you not get RMSE values for the first 3 rows?
We have not included this, since these statistics are taken from Ligtenberg et al. (2011) and based on a smaller dataset, which is therefore not fair to compare to $MO_{550}$/$MO_{830}$ FDM v1.2A. However, FDM v1.2A and FDM v1.2A-log have similar datasets, so these allow a comparison in terms of RMSE.

30. line 243: Do you mean that more densification is needed to match density profile measurements?
Yes indeed, we have added this:
"The fresh snow density is now independent of accumulation and generally lower, and hence more densification is needed at high accumulation sites to match density measurements".

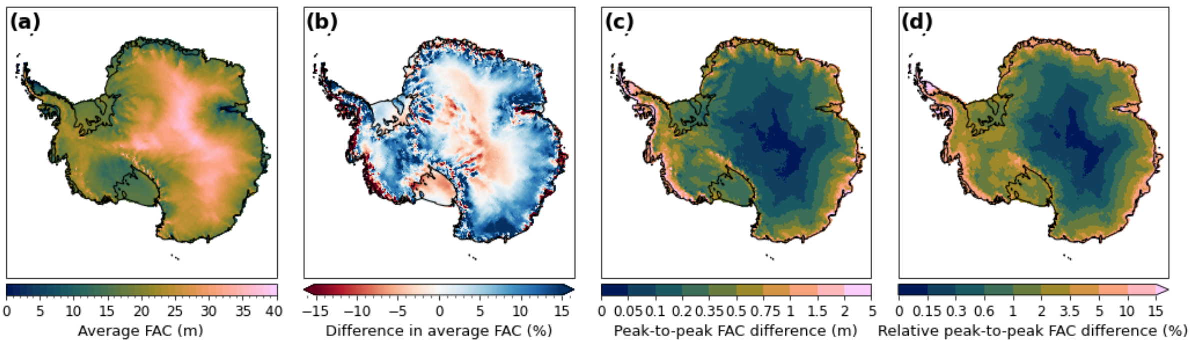
31. line 246: "When the locations are rerun …" sentence is confusing. Change to something like: "When the locations are rerun using the new MO values with the power fit for $MO_{830}$, the resulting RMSE of the modelled $z_{830*}$ and $z_{830}$ (firn thickness) are respectively 25 and 23 % lower compared to the logarithmic fit (Fig. 3c)".
That is indeed clearer, we have changed this according to your suggestion.

32. line Figure 3: It would be nice on panel (c) if the colors matched the same model run in the other panels (e.g. FDM v1.2A-log should be orange in all cases).
We agree that those colors should match, the color of FDMv1.2A in all panels is changed to blue, of FDM v1.1p1 to orange, of FDM v1.2A-log to red.

33. line 286: change to: 'contain the most air' and 'along parts of the coast'
We corrected this.

34. line 287: change to 'areas with very high accumulation'
We changed this.

35. line Figure 5: the high values on panels (c) and (d) are one the peninsula, which somewhat removes structure from much of the ice sheet. Consider changing the color bars to max out at a lower value (e.g. 2 m for panel c) and use an extended colorbar (as you do in panel b).
To improve the visibility of spatial detail, we suggest to update this to:



36. line 315: You say that snowfall is highest in winter, but I think that is only true in certain parts of the ice sheet.
We agree that this does not hold for the entire ice sheet, but 6a shows that this holds for the mean of the ice sheet. Therefore, we suggest to change this sentence:
"Snowfall is highest in winter, while firn densification, melt and sublimation peak in summer."
Into:
"Averaged over the ice sheet, snowfall is highest in winter, while firn densification, melt and sublimation peak in summer."

37. line 325: I am confused here: you say that the average seasonal amplitudes are defined as half the peak-to-peak values of the average seasonal cycle (and for FAC that is 2.4 cm). Then, you say that the average seasonal peak-to-peak values for FAC is 8.5 cm. I am not clear why, by your definition, the average season peak-to-peak is not just double (i.e. 4.8 cm) the average seasonal amplitudes. I am guessing that I am just not interpreting this correctly, so perhaps you can change the language to make it clearer.
We found that we used the words seasonal amplitude and seasonal peak-to-peak in the wrong way. As is pointed out, seasonal peak-to-peak is indeed twice the seasonal amplitude.

What we actually want to describe here are, firstly, the amplitude of the ice-sheet wide FAC/firn height, and secondly, is the average of all local FAC/firn height amplitudes.

Therefore, we suggest to change this:
"The spatially average seasonal amplitudes of the firn thickness and FAC are 3.5 and 2.4 cm, which amounts to Antarctic wide integrated volumes of 463 and 304 km$^3$, respectively. This agrees with the modelled firn thickness amplitude of 3.5 cm by Medley et al. (2020). The average seasonal peak-to-peak values of firn thickness and FAC, defined as the average difference between the highest and the lowest value of each year, are considerably higher, 13.5 and 8.5 cm, amounting to volume changes of 1784 and 1123 km$^3$, respectively (Fig. 6c). The difference between the seasonal amplitudes and peak-to-peak values is due to interannual variations in the timing of the seasonal maximum and minimum."

Into:
"The seasonal amplitudes of the ice sheet wide firn thickness and FAC are 3.1 and 2.1 cm, which amounts to Antarctic wide integrated volumes of 428 and 307 km$^3$, respectively. This agrees with the modelled firn thickness amplitude of 3 cm found by Medley et al. (2022). The average of all local seasonal amplitudes are considerably higher, 6.8 and 4.2 cm, amounting to volume changes of 892 and 562 km$^3$, respectively (Fig. 6c). The reason is the spatial variation in timing of the seasonal maximum and minimum (Fig. 6d)."

38. line 328: Change the word 'It' (starting the sentence) to 'This result' or similar ("It" acting as a pronoun does not have specific noun it is referring to).
"It" has been changed to: "This result"

39. line Figure 6d/335: It might be simpler to just say 'average day of year of maximum firn height' rather than "phase".
Yes I agree. We have the labels and caption of Figure 6d, and we have changed this:
"Figure 6d shows the average phase of the firn thickness maximum."
Into: "Figure 6d shows the average day of the year of maximum firn height."

40. line Figure 7: I am not sure of the best solution, but I think this figure is trying to convey too much information. It is rather difficult to look at the behavior of 7 lines. Additionally, I think it would be nice to pick a representative ~2 year period to plot with months marked to be able to see the seasonality (which you do a good job of discussing in the text) more clearly.
We agree that the figure is quite full. To avoid overlapping, therefore we already had shifted the FAC and temperature anomalies time series in the vertical direction. This section is about decadal variation, selecting two representative years would not show any decadal variation, and we already discuss the seasonal variation in Section 4.3 and Figure 6a. So we decided to keep the figure as is.

Another option would be to only include timeseries with a running average window of 5 year? But then you don't see the interannual variability in seasonal cycle.

The drawback of this solution is that the current figure nicely shows that interannual variability still relates to individual events. Smoothing would remove that information. We decided to keep Figure 7 but highlight more that it focuses on interannual variability and not on decadal variability.

41. line 382: Figure 8 caption specifies that 8d is model minus altimetry, but this is not as clear in the text. Consider change to: 'which is calculated by subtracting the altimetry elevation change from the FDMv1.2A elevation change'.
We agree that this requires some clarification in the text.
We have added: "Figure 8d shows the residual surface elevation trend of FDM v1.2A, which is calculated by subtracting the FDMv1.2A elevation trend from the altimetry trend".

42. line 383: Likewise, what is your subtraction for Figure 8e? FDMv1.2A minus FDMv1.1p1, or vice versa? Am I interpreting this correctly that blues are where FDMv1.2A is improved vis-à-vis FDMv1.1p1? I am a bit confused on 386-7 where you mention "less negative" – any negative (red) area in 8e is where FDMv1.1p1 is better than FDMv1.2A, correct? (Because you are subtracting the absolute residuals). So, to what are you referring that demonstrates the less negative? This figure does also include the altimetry product, however we acknowledge that this is not clearly described. The figure is calculated as the absolute residual of FDM v1.1p1 (|altimetry-FDM v1.1p1|) minus the absolute residual of FDM v1.2A (|altimetry-FDM v1.2A|).

We changed the figure label: "Difference in altimetry agreement between FDM v1.2A and FDM v1.1p1".
And added in the figure caption: "**(e)** Difference in altimetry agreement between FDM v1.2A and FDM v1.1p1. (absolute resdiaul FDM v1.1p1 compared to altimetry minus absolute residual FDM v1.2A compared to altimetry)".

In addition, we suggest to change: "In Figure 8e we subtracted the absolute residual surface elevation change of FDM v1.2A and of FDM v1.1p1 from each other, which yields the improvement of FDM v1.2A compared to FDM v1.1p1. In Figures 8d and e, regions with ice-dynamical imbalance are excluded. The surface elevation change of FDM v1.2A has improved in most regions, especially in Dronning Maud Land, Wilkins Land and Adelie Land, where the FDM v1.2A trend has become either more positive or less negative. The residual absolute trend of FDM v1.2A has been reduced by 17 % compared to FDM v1.1p1 (from 2.6 to 2.1 cm/yr)."

Into:
"In Figure 8e we show the difference in altimetry agreement between FDM v1.2A and FDM v1.1p1, which is calculated by subtracting the absolute residual surface elevation trend compared to altimetry (Fig. 8d) of FDM v1.1p1 from the similarly derived absolute residual of FDM v1.2A. The blue areas indicate an improvement in altimetry agreement of FDM v1.2A compared to FDM v1.1p1. The average absolute residual trend of FDM v1.2A compared to altimetry has been reduced by 17 % compared to FDM v1.1p1 (from 2.6 to 2.1 cm/yr). The improvement is most notable in Dronning Maud Land, Wilkins Land and Adélie Land. In Dronning Maud Land the FDM v1.2A trend has become more positive, and in Wilkins Land and Adélie Land the trend has become less negative (Fig. 8b,c, for region names see Fig. 1)."

43. line 389: trends → trend
This mistake has been improved, thank you for pointing it out.

**References**


Adusumilli, S., Fricker, H. A., Siegfried, M. R., Padman, L., Paolo, F. S., & Ligtenberg, S. R. (2018). Variable basal melt rates of Antarctic Peninsula ice shelves, 1994–2016. *Geophysical Research Letters*, *45*(9), 4086-4095.

Carter, J., Leeson, A., Orr, A., Kittel, C., & van Wessem, J. M. (2022). Variability in Antarctic surface climatology across regional climate models and reanalysis datasets. *The Cryosphere*, *16*(9), 3815-3841.

Medley, B., Neumann, T. A., Zwally, H. J., Smith, B. E., & Stevens, C. M. (2022). Simulations of firn processes over the Greenland and Antarctic ice sheets: 1980–2021. *The Cryosphere*, *16*(10), 3971-4011.