Response to Reviewer 2:

First of all, we would like to thank Baptiste Vandecrux for his time for reviewing and editing this study. We are happy with the constructive feedback we received. Based on the feedback we have revised the sensitivity analysis, and improved the remote sensing evaluation, and improved the clarity of the text. We believe that the given suggestions will improve the quality of the results and manuscript in many ways.

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 comment:
1. The authors present a new offline run from their firn model, using both updated forcing (RACMO2.3p2) and updated firn model (IMAU-FDMv1.2A). The changes in the model are clearly described. The model output is thoroughly presented, including how it differs from previous version of the IMAU-FDM model. The output is compared to a multi-mission altimetry product over 1992-2015. The model is also run under various scenarios to test its sensitivity to uncertainty in the forcing data, model parameters, or in the choices made for the spin-up procedure. The manuscript is nicely written, and the figures are of very good quality.

However I have major concerns on the science output of the study and how it increases our understanding of the firn characteristics in Antarctica. The comparison to altimetry (although necessary and much appreciated) is non-conclusive because of the uncertainty of the altimetry product, and reveal that other, more precise, datasets should be used to evaluate the model. The sensitivity analysis shows that the firn model tuning procedure allows to fit equally well observations with different forcings. This makes it hard for other research teams to use the parametrizations developed here with any other forcing than RACMO2.3.p2. Testing the sensitivity on the spinup setting can be of broader interest for regional climate modelling community, but the study does not conclude on a best practice on this matter.

Nevertheless, the dataset it produces is of great interest for the science community, and actually has more value in itself than the science output being presented in the manuscript. This makes the manuscript perfectly suited for a data-oriented journal. For the Cryosphere, I would encourage the authors to strengthen their findings. This could be done by exploring one of these three options:

- using new and hopefully more precise observation datasets to gain more insights on what the model is doing good or bad (coffee can experiments, GPS records, IceBridge laser measurements...)

- looking in more detail at regions where the model indicates changing processes.

- detailing the sensitivity analysis so that the study can conclude in best practices that can benefit other research teams.

Maybe such scientific findings are already in the manuscript, and I have just missed them. Then I much apologize for my comment. More light should then be given to few key findings and some of the side analysis should be removed to keep the focus on the main insights. The findings should be highlighted in the abstract and in the conclusion. Once the interesting science findings properly highlighted, the manuscript will be a great candidate for publication in the Cryosphere.

Thank you for these recommendations. We appreciate all the given options. To add further relevance to our scientific findings, we add more detail to the sensitivity analysis by estimating the uncertainty in the FAC and surface elevation trend for every location on the ice sheet. We have done this by extrapolating sensitivities for specific locations to the entire ice sheet by regressing them against various climatic variables. In addition, we included analysis using spatially variable uncertainties in temperature and precipitation. Notably, the spin-up sensitivity analysis reveals large uncertainties in the surface elevation trend. As differences in accumulation, and to a lesser extent temperature, during the spin-up period (compared to 1979-2020) result in large surface elevation trends over the Antarctic Peninsula and Ellsworth Land, the altimetry product provides a reliable source of evaluation here.

Detailed comments:
2. - l.5 "observations" of what? We added: “in-situ firn core observations”

3. - l.8 "with altimetry" Please replace by "with previously published multi-mission altimetry product for the X-Y period" Thank you for this recommendation, this has been changed: “with a previously published multi-mission altimetry product for the period 2003-2015”.

3. - l.9 "reasonably well" Please quantify the agreement. This sentence will be omitted in the revised manuscript. Quantification in other parts of the text is improved in the revised manuscript.

4. - l.19 Where does the 98% come from? Is it on average on the firn-covered area? At some locations, significant runoff can occur from the firn (e.g. ice slabs, perched or perennial firn aquifer regions). This value comes from Medley et al. 2022, where it is calculated from a firn model, however in the revised manuscript the value has been changed to 94%.

5. - l.63 "measurements" of what? We added: “firn core observations”

6. - l.66 "remote sensing altimetry" please replace by "a multi-mission remote sensing surface height change product for the X-Y period (REF)" We changed this into: “with a previously published multi-mission altimetry product for the period 1995-2015 (Schroder et al. 2019).”

7 - l.79 "further improved" do you use v1.2G as starting point? meaning do you use the same thermal conductivity as Brils et al. ? Yes, indeed. We propose to change this into: “We further improved upon FDM 1.2G for Antarctica to FDM v1.2A by updating ..”

8. - l. 83 "Kaspers et al. (2004)" How do they define the fresh snow density in their study? How long after deposition do they consider the snow to be fresh? Or if they look at surface snow, for which depth range? Was this study in Antarctica? Thank you for this recommendation. We clarified: “FDM v1.1 used the fresh snow density parameterization of (Kaspers et al. 2004), which has been calibrated over Antarctica and yields density values that typically represent the first 0.5 m of the snowpack.”

9. - l. 94 Same question as above for Lenaerts et al. (2012) We added that the parameterization has been calibrated over Antarctica. The depth range was already described.

10. - l. 108 "average surface temperature" is it a fixed, long-term average or the average for the past x years? Is it surface skin temperature or near surface air temperature? It is the long-term average surface skin temperature, we have specified this in the revised text.

11. - l.115 "z_830*" Here you mention z_830* but in the next lines you only mention MO_830. Is there a MO_830* ? I don't fully understand how z_830* is used. Please explain how MO_550 and MO_830 is used in Eq3. You are right, MO_{830} should actually be MO_{830^*}.
To explain how z_{830^*} is used, and to explain how MO_{550} and MO_{830} are used in Eq. (3) we suggest to replace lines 112-116 by:
“By comparing simulated and observed depths (m) of the 550 and 830 kg m-3 density levels (z550 and z830, respectively) Ligtenberg et al. (2011) found that Eq. (3) requires correction terms MO550 for p < 550 and MO830* for 550 < p < 830, which are defined as the ratio of modelled and observed values of z550 and z830*, where z830* = z830 – z550. The correction terms MO550 and MO830* are added as a multiplier to Eq. (3). MO550 and MO830* 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:”

12. - l. 120 What do you mean "optimize densification"? To make the model match observed firn density? Please describe explicitly what is your objective function when fitting the alpha and beta parameter.
Yes, we mean making the model match observed firn density. In the revised manuscript, we discuss differences between Antarctica and Greenland in Section 2.2, therefore this sentence has been omitted.

13. - l. 126 Please mention if there is any difference from v1.2G on this point. No, there are no differences between v1.2G and FDM v1.2A on this point.

14. - l. 127 "tipping-bucket" I think it is called simply "bucket". When the retention capacity of a layer (=the bucket) is full, the excess water flows to the next layer without the bucket/layer to "tip" or empty itself. It overflows, but it does not tip. Thank you for this recommendation, we have changed this to "bucket method".

15. - l. 129-130 "if the latent heat..." Please rephrase to " and that has subfreezing temperature" or something alike. Thank you, we have changed this to: "The retained meltwater refreezes when it reaches a layer with a temperature below the freezing point."

16. - l. 134 "the reference period" Since you haven't specified it so far, maybe change to "a" reference period. Thank you for this recommendation, we have changed this accordingly.

17. - l. 142 "a minor trend" Please give its magnitude to show that it is minor. Please explain briefly its origin. Is it because ice with air bubbles continue to replace the dense (917kg m-3) bottom ice prescribed at the initiation of the spinup? Later in the manuscript, you mention that there is no trend in average surface height because you assume a steady state. Is the removal of the "minor trend" your way of prescribing steady state? I am wondering how this minor trend look compared to the trend you get when using your alternative spin-up strategy.

In the alternative spin-up strategy that is used for Greenland, which was developed after our simulations were performed, the trend in FAC between 1 January 1979 and 31 December 2020 that is allowed is 0.5 mm/yr. In our simulations, the FAC trend is 0.6 mm/yr. However, as the output is stored at a 10 day interval, we compare 10 January 1979 with 30 December 2020. This means that part of the trend is caused by this time gap in summer. By removing this minor trend we make sure that the 830-917 kg/m3 density part of the firn layer is also in steady state. We have added the magnitude of the dH/dt trend to show that it is minor and added an explanation.

"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. This trend is removed before further analysis of the results."

18. - l.147 Here I am a bit confused by the v_ice and v_by (for ice shelves only?) terms and by the frame of reference used. Is v_ice the vertical velocity of the pore close-off (PCO) depth? Do you separate the compaction of deep firn into ice (which changes the PCO depth) from dynamical thickening or thinning of the underlying ice? Where is your height reference point? at the PCO depth? at the bottom of the ice? At sea level?

We explain this as follows: "Vice represents the downward movement of the surface by the local divergence of the ice flow, driven by the long-term vertical mass flux through the lower boundary of the firn column (density = 917 kg/m3). In a steady-state firn layer, this equals the mass flux at the upper boundary. Vice is therefore equal to the mean SMB (kgm-2) times ρi of the reference period, but of opposite sign."

If it is at PCO or bottom of the ice, then the buoyancy is outside of the system and shouldn't be included. If it is at sea-level, the isostatic rebound, and bedrock movement should also be in the equation. If you indeed placed yourself in a reference system where isostatic rebound is not important and where buoyancy is, then please define how you calculate buoyancy and ice thickness. Eq. (6) **only** describes the impact of firn and SMB processes on the firn height, therefore bedrock movement, basal melt and ice dynamical imbalance are not included, but it **does** include the impact on buoyancy of the changing firn mass, because this is an impact that the firn/SMB processes have on the firn height (compared to sea level). If we want to include buoyancy, we need to define our frame of reference at sea level.

We include the buoyancy impact only of the changing firn mass, (which is only relevant for the ice shelves, as those are floating), which does not require ice thickness estimates. We clarify: "(Vbouy) represents the vertical motion associated with the changing ice shelf draft when the mass of the firn layer changes."

19. - Section 2.5: Please use the same order as above to introduce your observations: i) Surface snow density, ii) Densification, iii) Surface height. We agree, the order has been changed.
You could even consider moving the observations' description at the end of sections 2.1.1, 2.1.2 and 2.3. So the reader finds out about the data right after it read about the fitting method or the surface height definition. For clarity we decided to keep the observation description separately.

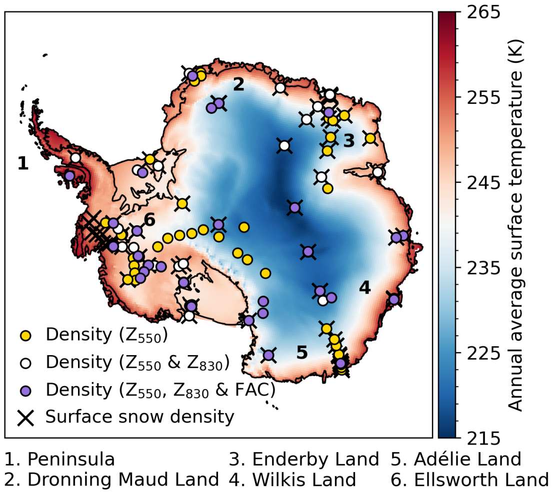
20. -l.181 "122 firn cores" I'm guessing this includes the 104 cores used for tuning MO. Please rephrase into "In addition to the dry cores used for the tuning of MO ratios, another 18 cores that could not be considered dry were used to evaluate the model output." This has been changed into: "In addition to the dry cores used for the tuning of MO ratios, another 12 cores from areas with regular melt were used to evaluate the model output."

There still 11 profiles missing from the 125 + 8 profiles mentioned in l.174. This section has been rewritten to clarify, see our answer to comment 14 of Reviewer 3.

21. - l.183 "Montgomery" Please give the original source/reference of the profiles, along with the SUMup citation. References and characteristics of firn core data will be added as supplementary material. We refer to this in the data availability statement.

22. - l.184 "top 0.5 m" This top layer thickness does not match with the 3 cm thickness of the top layer, nor the 1h time step of the model. We also calibrate with simulated 0.5 m, which is now explained in Section 2.1: "The fit coefficients A, B and C in Eq. (2) for the fresh snow density are retuned to improve the fit of the simulated with observed surface snow densities, defined as the top 0.5 m of the firn column, as this matches the thickness of the sampled layer."

23. Figure 1: Here a bit of creativity could help to avoid symbol overlapping each other. For example, z550 and z830 could be shown by half disks, and sensitivity analysis could be shown as a box surrounding the markers. Z550 and Z830 do not overlap, as the white dots indicate that both Z550 and Z830 are available. The final figure is shown below, which we think is sufficiently clear. We propose to add an additional figure for the sensitivity analysis locations in Figure S1 (See our response to comment 25).



24. - l. 197 "105 observations" From the legend in Figure 1 I thought that sensitivity analysis would be only done at the gray dots. See comment above.

25. - l.199 "10 additional locations" Now I am concluding that the gray dots are sensitivity analysis locations where you have no density observation available. Please change the marker of the sensitivity analysis to a different one than the observation location and remove the gray fill to show that the marker does not mask anything underneath. See comment above.

I also see that there are 105 sensitivity analysis location while you use 133 profiles for tuning and evaluation of the model. We need to see which locations have observations but are not part of the sensitivity analysis. I am also unsure how you summarize the sensitivity analysis at these 105 locations into Table 4. Are the results averaged? How do we know that the average is not biased due to over/underrepresentation of certain climate zones in the 105 sites selected? We added the 11 locations, to better represent all the climatic conditions and avoid underrepresentation of certain climatic zones. See the Figure S1 shown below for a map and temperature and precipitation histograms of the locations. In addition, in the revised manuscript, we now expand the sensitivity to the entire ice sheet, so there should be no spatial bias in the final results. We also specify this in Table 4 (Table 5 in the revised manuscript).

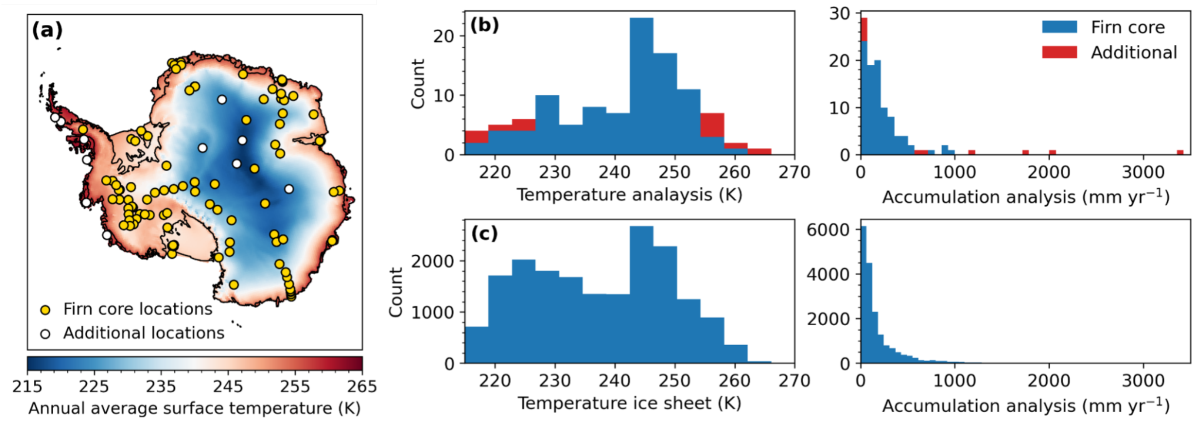
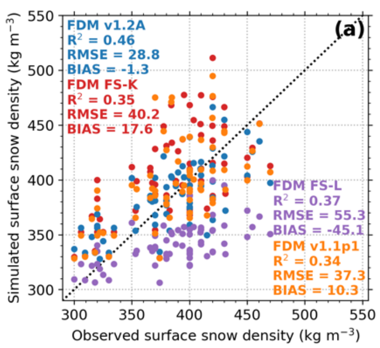


Fig. Locations used in the sensitivity analysis.

26. - l. 204 "-1.5K" Does this evaluate the air temperature or the surface skin temperature (which is that actual input of IMAU-FDM, I assume?)? This is evaluated with 10-m firn temperature, which is a good approximation equals annual average surface skin temperature. However, this comment is not relevant anymore, as this analysis has been changed to distributed uncertainty in surface skin temperature based on the spread in surface skin temperature between RCM's and reanalysis datasets.

27. - l. 212-213 I thought that you had no data before 1979? what are these years corresponding to? Please update to something like "To mimic this increase in precipitation, we create a spinup in which model loops three times over the 41-year-long reference period where we decrease precipitation in the first loop by 10%, by 6.66% in the second and by 3.33% in the third". Make something similar for the temperature.
Thank you for this suggestion. We have changed this into: "To mimic a gradual increase in precipitation, we performed a spin-up experiment in which the precipitation up until the third-to-last loop (the 42 year long reference period) is decreased by 10 %, in the second-to-last by 6.66 %, and in the last by 3.33 % with respect to the mean precipitation of 1979-2020. To mimic the increase in temperature, we create a spinup 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 by 0.33 K."

28. - l.224 Using different symbols and reducing the size of markers, I believe you can display this third comparison in Figure 2a. Please mention that the FDM FS-K shown in Figure 2a are calculated with RACMO2.3p2 surface climate. In IMAU-FDMv1.1, FS-K was used but the surface climate was also different, therefore we still don't know if the fresh snow density in the new model is greater or lower than in IMAU-FDMv1.1.
Thank you for this suggestion. We added FDM FS-K as well as FDM v1.1p1 in Figure 2a to also evaluate the impact of the different forcing.



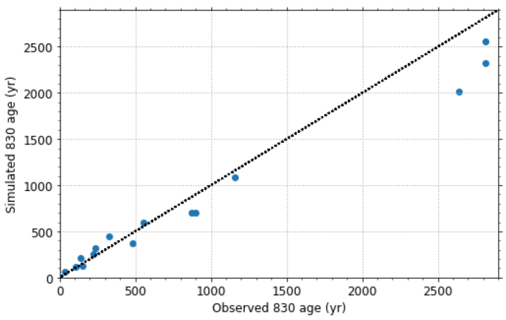
29. - l. 234-238 Brils et al. uses yearly temperature because it is the only type of parametrization of surface snow density in Greenland. There has not been any evaluation of the impact of wind speed in Greenland because there is simply no instantaneous and collocated measurements of temperature, wind-speed and surface snow density. In turn, people used yearly temperature, because they thought it was more robust and less model dependent. Also, a surface snow density parametrization built on simulated T and WS will only be valid when using T and WS from the same model because it accounts for potential model biases in the T and WS values. This should be mentioned somewhere. Consequently, I don't think that the difference between the parameterizations in

Greenland and Antarctica tell anything about the snow or the climate and might be just arbitrary.
We agree, we suggest to change Lines 234-238 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.”

30. - l.241 "the optimal MO is less steep" Please add, at least at the beginning of the paragraph, plain word explanation of what "lower/higher MO830" mean (i.e. overestimated/underestimation modeled densities for the top/deep firn). I'm still not sure if the MO correction enhances or inhibits the densification. Thank you for pointing this out, we understand that this can be confusing. We now clarify this in Section 2.1.2 as follows: “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.”

31. - l. 253 "Differences between" Please mention what are these differences. In terms of MO fits but also in general terms: Is densification faster/slower in than in Greenland or less/more responsive to accumulation or temperature? Good point. We clarified as follows: “In FDM v1.2G, the version for the Greenland ice sheet, a logarithmic MO_{830} fit was applied. Dry firn cores from the GrIS cover a smaller range of average annual accumulation, 80-680 mm/yr compared to 20-960 mm/yr in Antarctica. If we only include cores from the 80-690 mm/yr accumulation range, we find a similar r-squared value for the logarithmic vs power fit (0.67 vs 68). Another difference is that in FDM v1.2G, an almost constant value of 0.67 for the MO_{550} fit was found, which implies a linear correlation between the densification rate and the accumulation rate. In FDM v1.2A we do find a moderate (r=0.37, b = 0.12) correlation of MO_{550} with accumulation, i.e. AIS densification rate depends more strongly on accumulation than temperature compared to the GrIS. Again, if we only include cores from the 80-690 mm/yr accumulation range this correlation weakens (r=0.08, b = 0.07).”

32. **- l. 275** "Figure 4d shows the age of firn" This part could be more quantitative. What is the median/maximum firn depth, what is the median/maximum age at PCO? What is the fraction of the ice sheet covered by firn? Since you mention the age of the firn at PCO and its usefulness for ice core interpretation, could you present the age of the firn at PCO depth for ice core locations and compare to the values found in literature? Since you are trying to show that the model output can be used for this purpose, you might as well show how well it does.
Otherwise this paragraph and the figure could be removed to leave the focus on the surface height and FAC discussion.
I have compared the simulated PCO age to measured 15 delta age values from Breant et al. (2017) (See figure below). Based on that favourable comparison we added more quantitative information to this section, following your suggestions as follows: “Nearly all (>99 %) of the AIS is covered by a layer of firn (i.e. the ice fraction where average SMB is positive (observations suggest 99 %; Winther et al. (2001)).”
&
“The highest z500 and z830 values (30.6 and 114.2 m, respectively) are found in the interior, where the densification rate is low due to low temperatures, and the surface snow density is low.”
&
“Figure 4d shows the age of firn at the pore close off depth, here assumed to equal z830, which is on average 754 yr. The combination of strongly increasing firn depth and decreasing accumulation towards the interior leads to firn ages at close-off depth up to 3240 years in central East Antarctica, whereas the pore close-off firn age in warm and wet coastal margins can be as low as 20 yr. If we compare this to observations, we find an RMSE = 231 yr, R^2 = 0.985. On average RMSE is 25% of the observed values. In comparison, RMSE of z_{830} is 15% of the observed values and R2 = 0.73. These simulations add to our understanding of the large spread in timescales of the firn.“



33. - l.288 "somewhat" Please avoid this word and replace by quantitative information. We replaced it by 5 % and 7 %: “These average FAC estimates are 5 % and 7 % lower compared to values simulated with the CFM forced with MERRA-2 climate data of 24.0 and 17.0 m respectively (Medley et al. 2022).”

34. - l.289-290 "higher accumulation rates" Can you quantify this difference? We decided to remove this part, and reformulate as mentioned above.

35. - l.300 "also contributes to the general pattern..." Higher melt could indeed participate to the lower FAC at low elevations. But increased snowfall in the interior does not show in the new FAC pattern (FAC decrease in the center of the ice sheet in Fig5b). This could mean that the MO update overprints the effect of higher accumulation. That is indeed a good point, we included this in our explanation as follows: “However, in the highly elevated and low accumulation region, we find a reduction in FAC, which implies that the MO_{830} update outweighs the effects of higher accumulation and lower fresh snow density in this region.”

You also mention that the fresh snow density is decreased since the last version. This should have enhanced the FAC increase due to increased accumulation. Yes indeed, we also include the effect of fresh snow density.

36. - l.318 "Vice is constant" Why is it constant by definition? How do you define it? The description of V_{ice} in Section 2.3 (lines 164-165) has been clarified, see our response to comment 18.

37. - l.319 "Vbuoyancy is negligible" I'm still not sure how you define this and why is it negligible? Buoyancy is simply calculated with Archimede’s principle, we explain in Section 2.3: “vby represents the vertical motion associated with the changing ice shelf draft when the mass of the firn layer changes.”
and: “Vby, only relevant over ice shelves, is equal to the negative change in firn mass divided by the density of sea water.”

As can be seen from Figure 6a, these displacements are small. We added: “, while the AIS average impacts of snowdrift (v_{snd}) and buoyancy (v_{buoy}) are small (<1 %).”

38. - l.326 "in the timing" and magnitude? This section has been rewritten based on previous comments.

39. - l.327-328 "63 to 68%" I am not sure how you calculate this number. The lines above these lines have been rewritten. This explanation will be rephrased as: “The difference between the firn thickness and FAC seasonal amplitudes (6.8 and 4.2 cm) suggests that around 62 % of the seasonal surface elevation fluctuations are caused by a change in air content rather than actual mass change.”

40. - l. 332-335 Isn't it redundant to the discussion of Figure 5c? Could be moved elsewhere or removed for concision. See typo "ànd". In Fig. 5c we only include the peak-to-peak variability, “The high peak-to-peak variability (on ice shelves) is caused by a combination of temporal variability in accumulation and melt.” In Fig. 6 we describe the seasonal peak-to-peak variability in more detail, which is caused by high snowfall and melt, and not necessarily temporal variability in accumulation and melt.

41. -l.335-338 Does not add information. Consider removing for concision. We decided to maintain this figure. Differences in timing of the firn height maximum contributes to our understanding about the spatial variations in seasonal cycles and hence the interpretation of satellite altimetry. We also refer to this figure to explain the difference between the ice sheet-wide seasonal amplitude and the average of all local seasonal amplitudes.

42. - l.345 "5 to 10 years" Why does it matter that it last 5-10 years? How does that relate to frequency of El-Niño or annular mode? That is an observation from Figure 7. It does not directly relate to frequency of SAM (weeks) or El-Nino (3-7 years). A lot is still unknown about decadal variability in El nino and SAM. Therefore, we decided to remove this sentence.

43. - l.340 "Above 2000 m ..." I don't fully agree with this sentence. The variability in V_fc is much lower than below 2000 m. a.s.l. "larger temperature variability" than what/where? We agree that this was not clear. We have rewritten this: “Above 2,000 m a.s.l. the relative contribution of firn densification to the total cumulative dH anomaly is larger (32 %) than below 2,000 m a. s. l., where this is only 14 %. This difference can partly be explained by the larger interannual temperature variability above 2,000 m a.s.l. (sd in annual means = 0.78 K compared to 0.48 K) and the absence of melt.”

44. - l.352 "sd" Spell out what it stands for the first time you use it . This has been added.

45. - l.353 "67%" It sounds like a repetition from l.327-328, but calculated slightly differently. Maybe keep only one of the two formulations. Lines 327-328 is only about the seasonal variability. This number is about the decadal variation.

46. - l.358-364 This is methods. We agree, and moved this part to the methods (Section 2.5; Observational data).

47. - l.366 "Ligtenberg et al. ..." Please present your results first, then bring in previous simulations to put your results in perspective. Thank you for this recommendation. This is a good point. We will first present our own results.

48. - l.368 "likely explained by altimetry errors" This deserves more details. What was the motivation to say that? What altimetry errors was it?
To comply, we propose to replace "altimetry errors" by "seasonal variation in radar penetration depths, as these may amount to up to several centimeters in amplitude (Nillson et al. 2022)."

49. - l.369 "The altimetry observations prior to 2003..." This should be moved to the method and further justified. What is the measurement precision threshold that you use? The measurement precision threshold is highly variable in space, mainly depending on the surface slope, therefore we don't use a single threshold. Instead, Table 1 in Schroder et al. 2019 shows that the satellites from 2003 onwards have a higher precision, where the noise level is a function of the slope. We do include 1992-2003 in Figure 9. Therefore, we propose to add this to the method section:
"We will focus on the observations from 2003 onwards, as satellites before that period have a lower measurement precision (Schroder et al. 2019)."

50. - l.374-376 Since you mention that the improvement is due to better melt forcing in the coastal margin, can you present the same seasonal surface height amplitudes for the altimetry product and two model runs below and above 2000 masl (or any elevation that is relevant). This would provide the evidence that the improvement comes from these coastal regions. The increased melt of the forcing dataset is described by Van Wessem et al. (2018). Since melt almost exclusively occurs in the summer months, this obviously results in a higher seasonal amplitude. However, not only the increase in melt impacts the seasonal amplitude. Also, the fresh snow density over the AIS is lower, and there is increased accumulation in the interior. Therefore, we only find a slightly larger increase below 1,000 m a.s.l. than above 1,000 m a.s.l. (13% compared to 10%), which is therefore not included in the revised manuscript.

51. - l.379 "linear regression" Please give more detail: a linear regression is fitted to...annual/monthly/daily values... for the period ... Changed accordingly.

52. - l.385 "excluded" I understand that high dynamical imbalance regions should be masked out of the altimetry product. But there is no reason to remove it from the FDM to FDM comparison in Fig 8e. This Figure does also include the altimetry product, however we acknowledge that this was 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 have explained this in the text as follows: "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."

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

53. - l.385-388 Please merge these two sentences and show that the "17%" is what make you say that there is improvement.
We changed this as follows: "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)."

"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 regions are not indicated on any maps and you don't mention which ones became "more positive or more negative", neither why the simulated trend became more positive or more negative in these areas. Either give a proper description or remove .
This has been improved, see comment above.

54. - l.388-393 Please explicit how you rule out long term trend in ice dynamic thickening?
That is a good point, we propose to change this into: "In these regions, a substantial residual positive trend remains (>10 cm yr-1) (Fig. 8d), which can either be due to an increase in accumulation or ice dynamical thickening. Smith et al. (2020) found that the mass gain is mainly located along the steep slopes of the Antarctic Peninsula and decreases with distance from the ocean, which is indicative of an increase in snow accumulation. The increase in accumulation on centennial time scales in these regions (e.g., Thomas et al., 2015, 2017) implies that the actual firn column is not in balance with the 1979-2020 climate, as assumed in our spinup. As a result, the vertical downward ice flow (Eq. 6) is overestimated, which results in underestimated surface elevation change. This is confirmed by the results from the spin-up sensitivity analysis, in which the precipitation during the spin-up period was reduced, as indicated by the red shaded areas in Fig. 9."

55. - l.394-402 After briefly introducing Fig9, you end up discussing residual trends (l. 399). You might as well refer to Fig8d (which should be annotated with the names of the places mentioned in the text).
We agree. When discussing the residual trends, we will refer to Figure 8d.

Eventually in this paragraph you only mention location 5, 1 and 6. Which means that most of Figure 9 is not described in the text. The only piece of information given by Figure 9 is this variability in the altimetry product. It should be described and discussed better in the previous paragraph, where this variation actually matters. In that paragraph (Section 5.1) please give a metric for the this change in inter-annual variability: please show that on the majority of the ice sheet the standard deviation of detrended altimetry height change is significantly higher before 2003 than after. I suggest removing figure 9 and merging this paragraph with the two previous ones.
We will discuss the short-term variability of the altimetry product in Section 5.1, including the quantification in terms of standard deviation. We would like to keep Figure 9, as it shows temporal detail of the surface elevation time series. It also allows us to show the similarity FDM v1.2A and FDM v1.1p2, as we only have simulations of this model version on a limited amount of locations. Based on feedback from Reviewer 1, we also decided to include uncertainty bands in Figure 9. This shows whether the offset with altimetry trend falls within the uncertainty bands of IMAU-FDM for several locations.

56. - l.401 What is the proposed explanation for this increased variability before 2003? Could it be the frequency of the altimeter that was more sensitive to change in penetration into snow? The frequency of ENVISAT and ERS-1/ERS-2 is the same. We clarify in the revised manuscript, as follows: "as satellites before that period have a lower measurement precision (Schroder et al. (2019)."

57. - l.409 "+/-6%" Is this the temperature uncertainty or the magnitude of the FAC change? See my suggestions for Table 4 to avoid misunderstanding. We have redesigned Table 4, where we also resolve this confusion. See our response to major comment 1a of Reviewer 1.

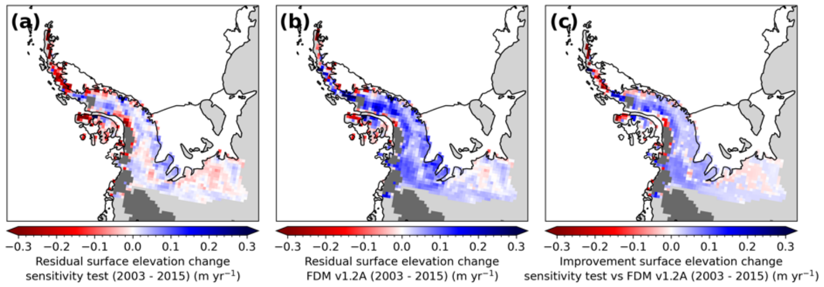
58 - l.409 Can you be more specific than "somewhat increase"? This part will be omitted.

59. - l.410 "robust" What do you mean by robust?
I would rather say that the MO fitting allows to produce realistic firn densities even if an arbitrary bias is applied to the forcing. It is also important to mention here whether the RMSEs showed here are from the measurements that are also used for the fitting of MO or if they are independent. If they are the one the MOs are being fitted to, then it is normal that they remain low after re-fitting the MOs. See comment 122 of Reviewer 1. As said above, we will focus on the sensitivity on FAC and surface elevation changes. Instead of using uniform changed precipitation (-/+8%) and temperature (-/+ 1.5K) in our revised manuscript we will use spatially variable perturbations in the sensitivity experiments, as a result there is a larger sensitivity of the simulated FAC (2- 4%) than with the uniformly changed precipitation and temperature (<0.5%). We will discuss the role of MO fitting

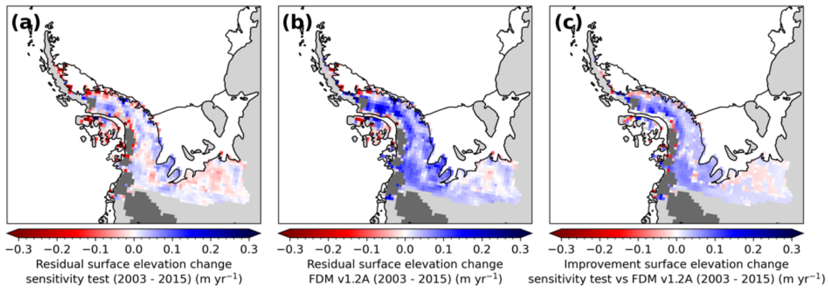
in reducing the sensitivity compared to without MO fit (4-7% FAC sensitivity) and contextualize the model sensitivity.

60. - l.411 This shows the limitation of using the entire period as a reference period for spinup.
Since no obvious strong long-term trends have been detected in Antarctica's surface climate and SMB during the period 1979-2020, there is no reason to choose a shorter period. The reason for this is that you want to capture the climate of an as long as possible period. If you take less than 42 years, you have a bigger chance that you are in an anomalous climatic period.

61. - l.416 "yields an average surface elevation trend" Does that help to get closer to the altimetry results in some areas? This is an interesting point, which we decided to explore further. We found that it reduces the residual surface elevation change over the Antarctic Peninsula and Ellsworth Land with 25% and with 38% if we only include firn regions whereby the age of PCO is >42 years (See figures below).



All locations



Only locations age PCO <42 yr.

Locations with PCO <42 yr show no increasing surface elevation trend as the firn from before 1979 has already been refreshed. We will include this analysis in the revised manuscript.

62. - Table 4. Please add the magnitude of the change applied in each run (not just the sign).
Are the relative changes in FAC and RMSEs in m or in %?
Is the FAC for all the ice sheet or the average change at the density profile locations?
The design of this table will be improved in the revised manuscript to avoid these confusions. See our response to major comment 1a of Reviewer 1.

63. - l. 417 "linear regression" I am wondering if this linear regression is necessary, since you have access to the model run. Could it be rephrased into "We notice that this trend (+3.3 cmyr-1) is more pronounced in high accumulation areas and find that, in this scenario, the average height change of areas with accumulation >1000mm yr-1 is 5.x m over the 1979-2020 period." ? We have extended our sensitivity analysis, and this part will be removed.

Once again, would this scenario get the model closer to the altimetry observations? Do you deem it more realistic than the reference run? See our response to comment 61.

64. - l.426 "partly explains" Please be quantitative. What was the improvement given by the inclusion of accumulation and temperature trends in the spinup? How much of the observed surface height trend is left unexplained even in this scenario? See our response to comment 61.

65. - l.430 I'm not sure what you mean by "robust". Please spell out. See our response to comment 122 of Reviewer 1.

66. - l.431 "error range of altimetry observations" The high uncertainty on the altimetry product (variability before 2003, uncorrected penetration, high relative uncertainty in the interior) poses the question of the suitability of this altimetry dataset for the evaluation of the firn model. There are many other datasets available (coffee can compaction experiments, multi-year gps records, ice bridge flights and as you mention ICESat 1 and 2 which have no penetration issues).

My main concern is that since the only observations presented in the manuscript are either used to tune the model (density profiles) or are too uncertain to truly assess the quality of the model (altimetry), I am wondering what knowledge we gain about the model or the processes it tries to describe. Where is its weakness? Where should it be improved next?
To partly overcome these concerns, we included 10-fold cross evaluations for surface snow density, z_{550} and z_{830}. Besides that, we also include FAC evaluation, and focus on the altimetry comparison on 2003-2015.

We acknowledge that the remote sensing dataset has substantial uncertainties, but it is the only dataset available that has a large spatial and temporal coverage. We have looked into other datasets, but they also have their disadvantages. ICESat-1 has poor spatial-temporal coverage over Antarctica, and ICESat-2 overlaps only 2 years and 2 months with our dataset, and it does not overlap with our FDM v1.1p1 simulation. IceBridge is mainly available in ice dynamical imbalance regions, which also makes the comparison difficult. Also, in this region the decadal and seasonal signals are relatively large (and thus less sensitive to the altimetry errors), which gives us confidence to use the dataset of Schroder et al. (2019).

We will articulate what knowledge we learn from this study in terms of model weakness. In the sensitivity analysis and remote sensing comparison we show that the spin-up can be a substantial model weakness, which provides an opportunity to improve the model next.

67. - l.438-442 This is a repetition of what was said before. Could be removed or merged with the conclusion. This section will be rewritten, to avoid repetition, and partly moved to the conclusions.

68. - l.463 please replaced "somewhat improved" by a more quantitative "improved by X%" We agree and we will include quantification throughout the manuscript. This paragraph will also be rewritten.

69. - l.464 Please spell out what you mean by robust? See our response to comment 122 of Reviewer 1.

70. - l.465 This last statement is rather loose and arbitrary. You could move your outlooks to the end of the conclusion to finish on future perspectives. We agree. We have removed this sentence, and we will rewrite this section.

71. - Code and data availability: Note that the Cryosphere is in favor of an open code and open data science: "We recommend that any data set used in your manuscript is submitted to a reliable data repository and linked from your manuscript through a DOI. Please see our data policy. Please also consider other assets like software & model code or video supplements." (TC instruction to authors) I know that it is something that has been done before by your research group and it would be great to continue in that direction. As a minimum, the scripts to reproduce the analysis and the figures should be made available to reproduce the study's results from the model output. Please also mention where to find the density profiles that are not in SumUp and the altimetry data.
References and characteristics of firn core data will be added as supplementary material. We refer to this in the data availability statement. The code of FDM v1.2A will be shared on Github. IMAU-FDM data can be obtained from the authors without conditions.

References

Brils, M., Kuipers Munneke, P., van de Berg, W. J., & van den Broeke, M. (2022). Improved representation of the contemporary Greenland ice sheet firn layer by IMAU-FDM v1. 2G. *Geoscientific Model Development*, *15*(18), 7121-7138.

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.

Nilsson, J., Gardner, A. S., & Paolo, F. S. (2022). Elevation change of the Antarctic Ice Sheet: 1985 to 2020. *Earth System Science Data*, *14*(8), 3573-3598.

Van Wessem, J. M., Van De Berg, W. J., Noël, B. P., Van Meijgaard, E., Amory, C., Birnbaum, G., ... & Van Den Broeke, M. R. (2018). Modelling the climate and surface mass balance of polar ice sheets using RACMO2– Part 2: Antarctica (1979–2016). *The Cryosphere*, *12*(4), 1479-1498.