Response to Reviewer 1:
First of all, we would like to thank dr. Vincent Verjans for his time for thoroughly reviewing this study. We appreciate the constructive feedback we received. By following your suggestions, we have, amongst other things, revised the sensitivity analysis, model evaluation, and various parts of the text. We believe these changes have led to an improvement in the quality of the results and manuscript.

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

Kind regards,

Sanne Veldhuijsen

Major comments:
1a) Many of the possible experiments presented in Section 2.6 are not even mentioned in the results section about uncertainty (Section 6). In Section 2.6, the authors mention various scenarios of accumulation and temperature reduction for the spin-up period (three for each variable). However, they only discuss a single test per variable in Section 6, and it is not specified which one it is. This obviously requires more clarity.
Thank you for pointing this out. We think there has been a misunderstanding, as our experiment description was unclear. We only tested three scenarios within the spin-up forcing uncertainty experiments, one with a lower temperature during the spin-up, one with a lower precipitation during the spin-up and one with a combination of lower temperature and precipitation during the spin-up period. In the revised manuscript, we also test the sensitivity to a higher temperature, higher precipitation and a combination of higher temperature and precipitation during the spin-up period within these experiments. We propose to clarify this as follows, in Section 2.6:

"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."

To clarify this further, we added the table shown below in Section 2.6, in which we give an overview of all the experiments.

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
ρ_s	ρ_s	+ and -30 kg m^{-3}	entire run	RMSE evaluation Section 3.1
T_s	T_s	+ and - spatial variable K	entire run	σ from Carter et al. (2022)
b	b	+ and - spatial variable %	entire run	σ 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.

Similarly, the sensitivity tests using "the 95 % confidence intervals of the MO fits", and the "uncertainty of the fresh snow density" are not discussed. I strongly recommend to clarify the links between Sections 2.6 and 6. For example, each sensitivity test should be given a name. And there should be a Table that specifies the sensitivity tests, their corresponding variable adjustments, and results.
Thank you for this recommendation. In the revised manuscript, we will include a discussion of the MO fits and fresh snow density results. To clarify the link between Sections 2.6 and 6, we added Table 2 in Section 2.6 and Table 5 in Section 6, which both include the names of the experiments.

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
ρ_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.

The Table 4 needs some further adjustments. I suppose that the FAC values in Table 4 are calculated only over the dataset of firn cores, and not over the entire ice sheet, which should be specified in the caption. Table 4 should also include the results from the spin-up perturbation experiments: 3 per variable (see Section 2.6), and the experiments combining temperature and accumulation perturbations (not explained in Section 2.6). Finally, the sensitivity experiments use ice sheet wide averages of precipitation spread (+/- 8 %) and of temperature uncertainty (+/- 1.5 K). However, local uncertainties can differ strongly from the ice sheet averages. Where possible, I recommend using spatially variable uncertainties in the sensitivity analysis.
Thank you, we agree that Table 4 (divided between Tables 2 and 5 in the revised manuscript) can be improved. We have added your suggestions. In addition, we have used spatially variable accumulation and temperature sensitivity uncertainties based on the standard deviation between RCMs and reanalysis datasets, which we obtained from Carter et al. (2022).

1b) Ice-sheet wide method for uncertainty quantification
The authors have investigated the results of the sensitivity experiments only at the firn core locations. However, this is of little interest to the community. A method to compute uncertainty estimates over the entire ice sheet in the different components of Eq. (6), and thus ultimately on dh/dt, should be developed. A straightforward approach could be to sample perturbations in all uncertainty sources mentioned in Section 2.6, compute

simulations at key locations, and regress uncertainty estimates against climatic variables. However, I leave the choice to the authors of how to best estimate uncertainty from their model, and I note that they have already worked on similar issues (e.g., Kuipers Munneke et al., 2015). The outcome of a more thorough uncertainty analysis should be:
- uncertainty bands in time series of Figure 7 and Figure 9
- maps of total uncertainty in modeled surface elevation change (and possibly of the uncertainty components)
- a quantification and discussion of the different components to uncertainty in dh/dt across Antarctica

Thank you for this recommendation. We agree that ice-sheet wide uncertainty estimates are of greater interest to the community. We expanded the uncertainties estimates by regressing the sensitivity of the uncertainty sources mentioned in Section 2.6 against climatic variables for the average FAC and dH/dt. We propose to include details of the sensitivity analysis in the supplementary material. As there are many interacting processes in the model, which can be variable in time, we propose to regress dH/dt instead of the different components of Eq. (6) separately. In addition, we propose to estimate the ice-sheet wide average FAC sensitivities as well in a similar manner. The empirical relations for each sensitivity experiment are included in the Figures S3. An example of these empirical relations is given in the Figure below.

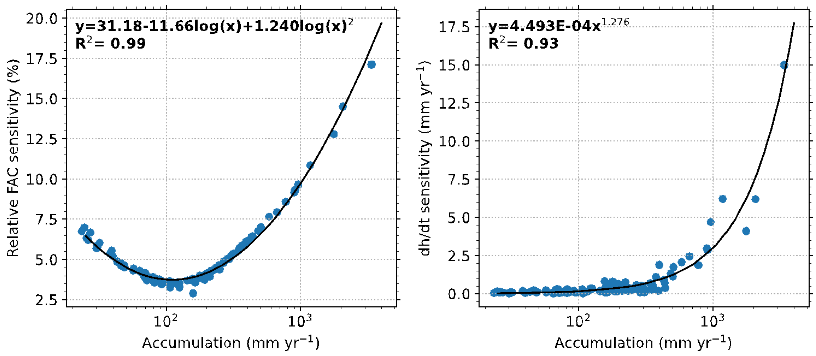
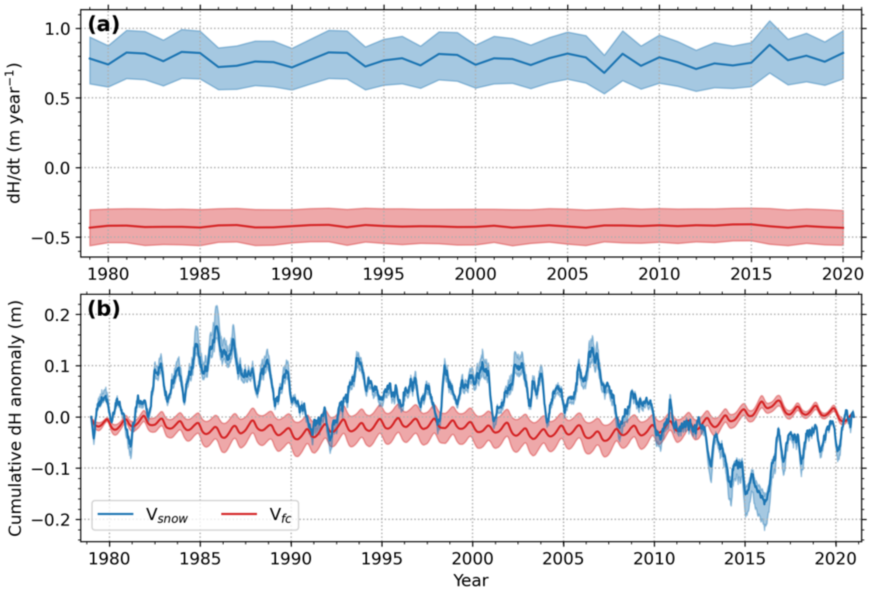


Fig. Examples of the MO fit empirical relations for the relative FAC and surface elevation change uncertainty.

These empirical relations are used to expand the sensitivity of the sample locations to the entire ice sheet. We have added the resulting maps of the total FAC and surface elevation change uncertainty and the contribution of each uncertainty source in Figures S4.

In order to show the variation in uncertainty of FAC and surface elevation over time, we plotted time series of surface elevation change and FAC including the sensitivity for the sample locations (Figs. S5a,b). This shows that nearly all uncertainty (>98%) of the ice sheet averaged surface elevation change and FAC is caused by the spin-up experiments. As the spin-up experiments cause a relatively uniform spatial and temporal response in FAC and surface elevation change across the ice sheet, the uncertainties can be expanded over the ice sheet for each timestep. In Figure S5c we only show the dominant sensitivities of the total AIS surface elevation change and FAC time series for the spin-up experiments. To expand FAC change for the combined accumulation and temperature spin-up experiments, we obtained additional empirical relations, see Fig. S3f. We include a discussion of these new sensitivity analysis results in Section 6.

We decided not to include uncertainty bands of the components of eq. (6) in Figure 7. In the Figure below we show time series of the uncertainties of (a) yearly rates and (b) cumulative anomalies of V_{snow} and V_{fc} of all sensitivity experiments for the sample locations. However, in our manuscript, we rather focus on the total surface elevation change, and the contribution of each uncertainty source, as this is of greater interest to the community, instead of on the processes separately.



Results from our sensitivity analysis are now also included in the comparison with satellite altimetry, by including uncertainty bands of dH/dt in Figure 9. In addition, we compare dH/dt results for the Antarctic Peninsula and Ellsworth Land over the period 2003-2015 of our sensitivity analysis to satellite altimetry, and we found that the residual has been reduced by 25% compared to the reference FDM v1.2A simulation. See our response to comment 61 of Reviewer 3 for more details.

2) Statistical procedure

In statistical calibration of parameters such as the MOs, it is well-known that validation should be, to some degree, independent of calibration. Here, the authors use the same dataset to calibrate their new parameters (MOs and surface density) as to evaluate their calibration. As such, the validation is not meaningful. I realize that firn data is sparse, and it can be argued that excluding a part of the dataset from calibration might be detrimental to model fitting. But that does not exclude a form of k-fold cross validation to better evaluate the fits. Furthermore, this approach would provide more robust uncertainty estimates on the MO and surface density values. For the evaluation, the authors use only 10 additional measurements in the evaluation dataset, in addition to the ones used in the calibration dataset.

Thank you for this recommendation. We have included k-fold cross validation for the MO fit and surface snow density. We found that the RMSE for the surface snow density amounts to 30.4 kg/m3 (compared to 28.8 kg/m3 for the reference FDM v1.2A), the RMSE for Z830 to 9.11 m (compared to 8.83 m for reference FDM v1.2A), and the RMSE for Z550 to 2.49 m (compared to 2.44 m for the reference FDM v1.2A). This additional evaluation will be included in the manuscript.

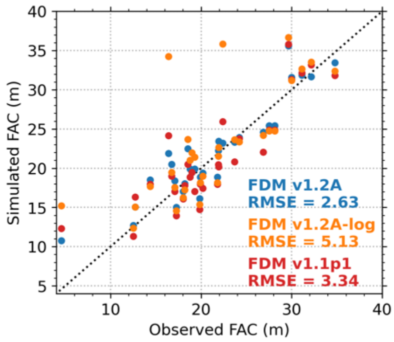
This modest introduction of independent data in the evaluation results in (line 243): "a slightly deteriorated correlation for the MO550." when compared to the FDMv1.1p1 parameterization.

I think there is a misunderstanding here. The deteriorated correlation of the MO_{550} fit is not because of the introduction of the independent surface snow density data, as the independent surface snow density data has no impact on the MOfit. The r-squared value of the FDM v1.1p1 MO_{550} fit is higher (0.43) than the FDM v1.2A MO_{550} fit (0.37). The reason for this could be that the previous fresh snow density parameterization was based on accumulation, which can give a higher dependency of accumulation and MO_{550}. We propose to rephrase Lines 242-244 as: "The optimal MO550 fit is less steep than the one found by Ligtenberg et al. (2011). The fresh snow density is now independent of accumulation and generally lower, and hence more densification is needed at high accumulation sites to match the density profiles. This might explain the slightly deteriorated correlation of the MO_{550} fit."

Furthermore, in Section 6, the authors demonstrate that the calibration of the MOs is sensitive to realistic uncertainty in climate forcing, and that this sensitivity strongly impacts results in FAC, z550, and z830 (Table 4). These two aspects thus show evidence that the statistical fits performed in this study are probably sensitive to noisy features inherent to the observations used in the calibration.

This is indeed true; therefore, we include k-fold cross validation and a sensitivity analysis with the MO fits.

In Table 3, I recommend providing the RMSE values in all the rows, to give the bias values in addition to the RMSEs, and to compute these fit statistics also with respect to the FAC values of the dataset. Finally, if I understand the process correctly, the comparison of fits between FDMv1.1p1 and FDM v1.2A for z550, and z830 is unfair. FDMv1.1p1 was calibrated with climate output of a previous RACMO version and another surface density parameterization. As such, it is obvious that it performs worse than FDM v1.2A when it is used with RACMO2.3p2 forcing and another equation for surface density.
In this table we only show the statistics of the MO fits and not the evaluation of the density, which is done in Figure 3. The FDMv1.1p1 MO fit statistics are taken from Ligtenberg et al. (2011), and are thus entirely based on old calibration, parameterizations and input, hence fair. We include characteristics of all simulations used in Table 1, to avoid raising the impression that the comparison is unfair. In addition, the RMSE column is about the RMSE with the fit, and only valuable to compare for FDM v1.2A-log and FDM v1.2A, as these have the same set of data points. We have added FAC evaluation of FDM v1.2A-log, FDM v1.2A, and FDM v1.1p1 in Figure 3e (See figure below).



Furthermore, it is unclear how much of the improved fit to observations is due to the re-calibration of IMAU-FDM versus the updated climatic forcing and the updated surface density parameterization. All these aspects should be addressed in the manuscript, and identified as caveats in the comparison.
This is an important point. To overcome this, we now include FDM v1.1p1 in the surface snow density evaluation (Table 3 and Figure 2a).

The same holds when comparing FDMv1.1p1 and FDM v1.2A to the altimetry product (Figure 8): it remains unknown what part of the improvement is due to changes in the densification equation, changes in the climatic forcing, and changes in the surface density parameterization. As pointed out by the authors, Figure 9 suggests that most of the improvements are due to the update of RACMO, which raises questions concerning the improved performance of FDM v1.2A in simulating firn processes. This should be discussed more in depth.
Regarding the surface elevation change, the forcing indeed explains almost all of the difference as is shown by Figure 9. However, Figure 2b and 3d show that changes in updates in the fresh snow density and firn densification parameterizations have substantial impact on the time averaged firn profiles. These outputs are important when IMAU-FDM is used for instance in the input-output method to estimate AIS mass balance, and to assess the meltwater buffering capacity of the firn layer. This distinction will be discussed in Section 7 (Implications, remaining limitations, and outlook) following suggestions from Reviewer 3.

3) Neglect of melt areas
The MO parameterization is constrained only in dry firn areas, but the model is used in wet firn areas also. One can reasonably expect errors to be much larger in modeling wet firn densification. While this topic is not the focus of the study, any ice-sheet wide study of firn evolution should at least discuss this limitation. Ideally, I would encourage the authors to evaluate IMAU-FDM in the wet firn areas also, by comparing modeled FAC to observed FAC from firn cores. They could also provide uncertainty estimates that are valid for melt areas.
Thank you for pointing this out. It can indeed be expected that densification rates are different between dry and wet firn. Unfortunately, firn cores from melt areas are sparse in Antarctica, and we only found two that were of sufficient quality to calculate FAC. Especially in melt regions, forcing uncertainties have a large impact on the simulated FAC. In Greenland more cores with FAC observations are available, see Brils et al. (2022), their Fig. 5 (see below). Based on that result, the model appears quite capable of simulation wet firn densification.

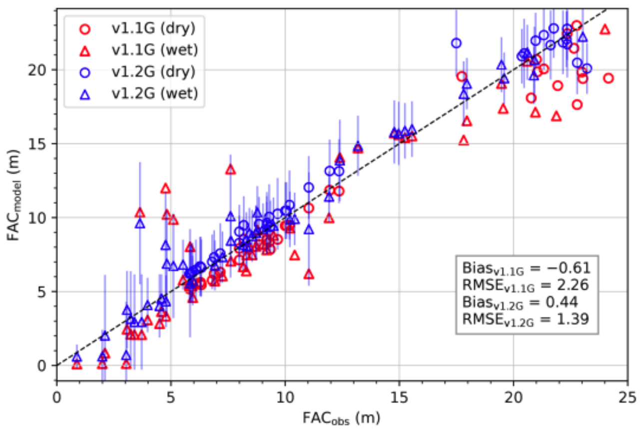


Figure 5. Modelled vs. observed firn air content in metres. Dry locations are indicated with circles, whereas wet locations are indicated with triangles. A location is labelled as dry if it experiences 5 % less melt than accumulation during the spin-up period. The blue lines indicate the uncertainty in the v1.2G results.

(Brils et al. 2022).

Minor comments

1) There is a general lack of quantification. I encourage the authors to identify all the uses of words such as "reasonably", "somewhat", "roughly", "improved", and "substantially". These should be complemented by quantitative values.

We agree and have included quantitative values throughout the manuscript.

2) In the densification equation (Eq. (3)), it seems to me that the mean long-term accumulation rate is used. An alternative approach is to use the mean accumulation rate over the lifetime of each specific firn layer, which is more representative of the effect of overburden stress (Li and Zwally, 2011). This aspect could be important given that decadal snowfall variability can be large on the Antarctic ice sheet. And there is no physical reason for the densification of a firn layer to be a function of past accumulation rates. Why did the authors choose to use the mean long-term accumulation rate?

We acknowledge that taking ƀ as a constant is an approximation that simplifies the densification rate, but the uncertainty this introduces is minor. Brils et al. (2022) demonstrated for two locations in Greenland that the error in the load experienced by a layer of firn is <3.2% at Summit and <1.9% at Dye-2. We included this information as follows: "ƀ is used as a measure of the overburden pressure, thereby assuming that the accumulation rate is constant over time. Using FDM v1.2G for Greenland Brils et al. (2022) shows that assuming a constant ƀ introduces a minor error in the load experienced by a layer of firn (e.g. <3.2 % at Summit and <1.9 % at Dye-2)."

3) Lack of clarity about the firn core dataset

Depsite re-reading several time Section 2.5, it is still unclear to me how the authors selected their dataset.

- "We used 125 density profiles from firn cores and 8 density profiles from neutron density probe measurements": but the wet firn cores were discarded for the calibration, so how were they used in this work?

Wet firn cores were used for the evaluation and calibration of the surface snow density, and used for firn densification evaluation. We clarified this in the revised paper and have rewritten this section, see our response to comment 14 of Reviewer 3.

- "For the MO fits, 104 dry firn cores could be used (…) To evaluate the firn density profiles from the simulation using the derived MO fits, 122 firn cores could be used.": what explains this difference of 18 cores?

Some of these cores are wet cores, and others have only been used for the surface snow density. We clarified this in the revised paper.

- in Section 2.6 "105 observational locations shown in Figure 1": 105 is not even mentioned in Section 2.5. We clarified this in the revised paper, and plotted the sensitivity analysis locations separately in Figure S1 (See our response to comment 51).

- in caption of Figure 1 “The grey circles indicate ten additional locations that were included in the sensitivity analysis”: give more details about these 10 additional firn cores in Section 2.5.
We have rewritten these sections by taking the comments above into account, and feedback from the other reviewers.

- Concerning access to the firn core data, please see the Data policy section of The Cryosphere: “Authors are required to provide a statement on how their underlying research data can be accessed. This must be placed as the section "Data availability" at the end of the manuscript.”
References and characteristics of firn core data will be added as supplementary material. We refer to this in the data availability statement of the revised manuscript.

Specific comments

1. l4 “improved”: specify that this refers to a previous model version. We have changed this into: “We have improved previous fresh snow density and firn densification parameterizations.”

2. l5 “observations”: change to “firn core observations”. Thank you for noticing this, we have changed this accordingly.

3. l15 The reference cited shows that firn thickness can exceed 100 m. Thank you, this has been changed accordingly.

4. l17 – 19 “Firstly, firn depth and density estimates are required to convert altimetry observed volume-to-mass changes, which remains a major source of uncertainty in mass balance studies”: please also reference the work of Morris and Wingham (2015). Thank you for noticing this, we have included the reference.

5. l27 – 29 Please reference the work of Arthern and Wingham (1998). This reference has been included as well.

6. l27 “are used as measures of its dynamics and mass balance”: clarify and rephrase. Thank you for this recommendation, this has been changed to: “Changes of the AIS surface elevation are an expression of changes in the firn layer and of dynamical change of the underlying ice and bedrock.”

7. l30 Change “in a mass and density component” to “in a mass- and density-change component”. Thank you for this suggestion, we have changed this accordingly.

8. l32 – l37 Is this paragraph only about ice shelves. If yes, this should be specified in the first sentence. If not, the reference to ice shelf hydrofracturing is confusing. We have rephrased this into: “In a warmer future climate enhanced firn compaction, melt, and refreezing can all potentially lead to firn air depletion, thereby limiting the meltwater storage capacity of the firn. This is especially important over the ice shelves, where meltwater accumulation can lead to hydrofracturing-induced ice shelf collapse (Munneke et al. 2014; Datta et al. 2019) thereby accelerating future Antarctic mass loss and sea level rise (Gilbert et al. 2021).”

9. l33 I do not believe that there is evidence of “reduced accumulation” for a warmer future climate in Antarctica. We agree. This has been omitted.

10. l33 Use “potentially lead to”. This has been changed.

11. l40 Please reference the work of Medley et al. (2015). This reference has been added.

12. l44 “Firn models can roughly be divided in two classes: physically based and semi-empirical models.”: please reference the work of Morris and Wingham (2011). We do not see how this paper is related to this sentence, therefore we have not included it.

13. l46 It is not clear to me what “which” refers to. We clarified this: “use a larger degree of tuning to observations representative of the past or present climate, while these conditions might not be representative of a future warming climate.”

14. l47 Change “less poorly known parameters” to “a smaller number of poorly constrained parameters”. We have changed this.

15. l49 Please reference the work of. The reference is missing in the reviewers’ comment.

16. l56 Start the sentence as: "This study shows that, at the basin scale, (…)". Thank you, done.

17. l58 Add comma: "climatic conditions, firn densification". The comma has been added.

18. l63 Specify the sort of "field measurements". We have specified that this is about in situ firn core measurements.

19. Table 1 In the column "Other", please refer to Equation numbers. Good idea. We have included this.

20. l75 "FDM V1.2G" has not been defined yet. Maybe, refer earlier to Table 1. Thank you for noticing this, we changed this.

21. l77 "comparable": with respect to what metric? This sentence will be omitted.

22. Section 2.1.1 and in the remainder of the manuscript. Please do not use the same symbols for different parameters. If the authors want to preserve a connection between related parameters, I recommend the use of subscripts. Thank you for noticing this. D has been changed to E throughout the manuscript.

23. l83 Change "average" to "averages of". This has been changed.

24. l86 Cite a reference for "Snow crystal size and therefore fresh snow density indeed increase with increasing temperature.". We have included the reference of Judson and Doesken (2000).

25. l94 Specify the frequency of "instantaneous": hourly/daily/… This is a single time step of RACMO, which is 6-min, this has been added.

26. Equation 2 I believe that there should be no overbar on Ts and V10 in this equation. Agreed, we have adjusted this.

27. l101 "defined as the top 0.5 m": am I correct that the authors calibrate the density of the modeled upper 0.5 m to the density of the observed upper 0.5 m, but that the calibrated surface density is then used only for the top layer of IMAU-FDM? If so, please clarify the approach as well as the slight discrepancy between calibration and usage of the surface density parameterization.
Yes indeed. We changed this into: "The fit coefficients A, B and C in Eq. (2) for the fresh snow density are retuned to improve the fit of 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. These fit coefficients are than used to simulate the fresh snow density."

28. l102 Specify the model time step here or elsewhere. The model time step is described in Section 2.1.3.

29. Section 2.1.2 All equations should be specified as two different cases for rho<550 and rho>550 kg/m^3
This has only been done for eq. 5, because this distinction is not made in the other equations.

30. l110 I suggest replacing "processes" by "mechanisms". This has been replaced.

31. l111 – 112 A citation is needed for this sentence. We added Herron and Langway (1980) as a reference.

32. l114 Change "turn out to depend on the accumulation rate" to "are chosen as functions of the long term mean accumulation rate". We changed this accordingly.

33. l120 Why was the power-law function not tested for MO_{550}? This was indeed tested, resulted in poorer results, and, therefore, was not applied. In order to keep the manuscript somewhat short, this test was not mentioned.

34. l124 Provide formulation of the thermal conductivity. We refer to Brils et al. (2022), to somewhat limit the size of the manuscript.

35. l128 Provide formulation of the irreducible water content. See comments above.

36. l129 The description of the refreezing algorithm is unclear and confusing. It suggests that no meltwater refreezes if a layer cannot accommodate all the incoming meltwater. We agree and have changed this to: "The retained meltwater refreezes when it reaches a layer where the latent heat can be released."

37. l131 Typo: remove "the". Thank you for noticing this, we have removed this.

38. l131 Specify the amount of melt. There is actually no limit, we have improved this.

39. l131 Do not use "significant" because it does not refer to statistical significance here. We agree. This part has been omitted, see comment above.

40. l134 – 135 What is the depth of the model domain? And what are the boundary conditions at the lower boundary? This is defined in Section 2.1: 3000 layers of 3 to 15 cm thickness. The lower boundary conditions for the thermal conduction are described in Section 2.3.1 of the revised manuscript.

41. l137 I believe that "total thickness" should be replaced by "total mass". You are right. Done.

42. l142 Please quantify the "minor trend". Done.

43. l142 Change "ice" to "material with ρ>830 kg m^{-3}". This has been changed.

44. Equation 6 Format of the variables in the equation does not correspond to format of the variables in the main text. Thank you for noticing this. The format in the text has been adjusted.

45. l152 SMB units are wrong. The units should indeed be kg/m2/yr, changed.

46. l160 Change ";" to "and". Done.

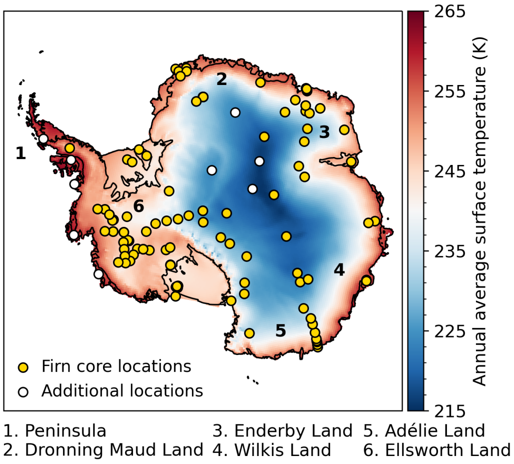
47. l161 Explain briefly the notion of "upper-air relaxation". We added: "Upper-air relaxation is the indiscriminate nudging applied to the upper part of the atmosphere only."

48. l168 Split this sentence in two: "This results in an improved forcing. For example, (…)". Done.

49. l169 Remove "e.g.". Done.

50. l178 Typo: "describes". This has been changed.

51. Figure 1 As I understand it, all the firn cores are used for the sensitivity analysis. For this reason, I recommend changing the label for the grey dots in the legend to "Sensitivity analysis only". We propose to add a separate figure for the sensitivity analysis locations in Figure S1:



52. Figure 1 caption Replace "on top of" by "in addition to". See comment above.

53. Section 2.6 Provide a name for each sensitivity experiment (see Major comment 1). This has been done in Table 2 and 5 of the revised manuscript.

54. l197 "To improve the representation": I do not see the causal link between improvement and the rest of the sentence. This might be confusing, what we mean is that we aim to have a better representation of all climatic conditions across the ice sheet within the sensitivity analysis. This has been changed: "To improve the representativeness of the climatic conditions over the AIS, we used 10 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 (Fig. S1)."

55. l203 – 204 Change "+-" to +/-". Thank you, done.

56. l204 Does +/- correspond to the RACMO2.3p2 RMSE? If so, please specify this. This comment does not apply anymore, as we have changed the temperature uncertainty experiments.

57. l206 Specify that 30 kg m-3 corresponds to the RMSE in surface density. Same as above: We added: "The uncertainty of the fresh snow density was based on the RMSE from the evaluation with in-situ measurements"

58. l206 "Section 3.2" should be "Section 3.1". Thanks.

59. Table 2 There is no "D" parameter in Eq. (2) for FDM FS-L and FDM v1.2A. We fixed this.

60. l218 – 219 Refer to Equation numbers. We fixed this.

61. l231 Change "reduced with" to "reduced by". Thank you for noticing this.

62. l240 "a simulation of FDM v1.2A without MO fits": does that mean a simulation with the FDMv1.1 MO values? This has been changed to: "To tune the dry snow densification rate we first performed a simulation of FDM v1.2A without MO corrections."

63. l241 "less steep": specify that this is with respect to accumulation (bdot). This has been changed to: "The optimal MO_{550} fit is less steep with respect to accumulation than the one in FDM v1.1p1."

64. l253 Replace "approaches zero" by "decreases asymptotically towards zero". This has been changed.

65. l255 Provide references for the accumulation rate values. Thank you, a reference to Brils et al. (2021) has been added.

66. Table 3 In the column "Version", I believe that MO_{550} and MO_{830} should be replaced by z_{550} and z_{830*}. In the column "Fit", please refer to Equation numbers. Add a Bias column (see Major comment 2). Provide RMSE values for all models (see Major comment 2). Add FAC rows (see Major comment 2).
See response to major comment 2. We refer to the equation number in the fit column.

67. l258 Specify "27 km horizontal resolution". Thank you. Horizontal has been added.

68. l259 Typo: "in Figure 4". We fixed this.

69. l261 What does "calm" mean? The definition of calm is the absence of winds, therefore we assume this is clear.

70. l262 Change "On top of this" to "In addition to this". Thank you, we changed this.

71. l264 – 265 "The spatial pattern of the depths of the critical density levels z_{550} and z_{830} are shown in Figures 4b and c, and are roughly the inversed pattern of the surface snow density": not in melt areas, please discuss (see Major comment 3). This is also the case for melt areas, we find high surface snow density and low z_{550} and z_{830}.

72. l266 – 267 "The patterns vary spatially across climatic regions with temperature as a primary driver and accumulation as a secondary driver.": please discuss the impact of surface melt (see Major comment 3). We agree that surface melt is also an important driver, we have changed this into: "The patterns vary spatially across climatic regions with temperature as a primary driver and accumulation and surface melt as secondary drivers." In line 268-268 we discuss the impact of surface melt: "Generally, the z_{550} and z_{830} values are low along the warm coastal margins due to faster densification, higher surface snow densities and surface melt."

73. l270 – 271 Please rephrase this sentence with more formal language. We propose to change this into: "The high accumulation rates in these regions result in less densified snow with depth and thus thick firn layers despite relatively high temperatures."

74. Figure 3 I believe that there is a mismatch between the numbers of points shown in the subplots, and the number of cores mentioned in Section 2.5. For example, it seems to me that there are less than 100 data points shown in Figure 3a. Also, please show scatter plots of the match between modeled and observed z830*, as well as between modelled and observed FAC. See responses to minor comment 3) Lack of clarity about the firn core dataset. We will rewrite this section. We have included FAC evaluation in Figure 3 (See figure major comment 2), and we already have z_{830*} in Fig 3c.

75. l281 – 282 Change "can be used" to "must be used". It can also be done with surface elevation change in combination with SMB data. However, we have changed "can be used" to "as it is used."

76. Figure 4 caption Change "firn age of the crtical density level" to "firn age at the crtical density level". Thank you, this has been changed.

77. Figure 5 Define "peak-to-peak" in the caption. Please also include a map of average seasonal amplitude. We have included the definition of the peak-to-peak FAC. We decided not to include a map of average seasonal amplitude, as we focus on seasonal variability of surface elevation change and FAC in Section 4.3. In Section 4.2 and thus Figure 5, we only introduce the temporal variability of FAC.

78. l286 Change "most" to "more". Based on feedback from reviewer 3, we have changed this into: "contain the most air."

79. l286 Typo: "parts of the coast". Thank you, this has been changed.

80. l292 If statistical significance has not been tested for, please do not use "significant". We agree, significant has been omitted.

81. l294 Replace "closer to the mean" by "less spatially variable". Thank you, that is indeed better.

82. l303 Please remind the reader about the study period. We added: ".. during the study period (1979-2020)."

83. l303 – 304 "Large values indicate that seasonal and interannual climate variability cause large temporal variations in FAC.": this statement is not supported by the map of peak-to-peak variability, which depends only on two single values in the entire time series. In my view, peak-to-peak does not characterizes temporal variability well. We agree that peak-to-peak is no direct measure of seasonal and interannual variability, therefore we have omitted this sentence.

84. l313 Vtot does not appear in Equation 6. Maybe simply replace dh/dt by Vtot in Equation 6. We added this.

85. l320 Specify that dh/dt in summer months has contributions from sublimation and melt. We added: "which is driven solely by snowfall except for the two summer months, which also have contributions from sublimation and melt."

86. Figure 6 caption Change "indicate the standard deviations" to "indicate the inter-annual standard deviations". Thank you for this recommendation, done.

87. l329 Change "biases" to "errors". Thank you for pointing this out, we changed this.

88. -Section 4.4 In general, this section requires much more quantitative assessments (see Minor comment 1). We agree and have made this section more quantitative. E.g. we added this in the part below: "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."

89. -l340 “cumulative surface temperature anomaly”: is this the cumulative anomaly in surface temperatures from the long-term mean? Please clarify. We specified this.

90. -l341 “the seasonal firn thickness and FAC variability is driven by”: please discuss why there is no one-to-one correspondence between firn thickness and FAC variability. As the FAC and firn thickness are different characteristics, we also do not expect a 1:1 comparison, therefore we have not discussed this in the manuscript.

91. -l341 Typo: “is” should be “are”. Thank you for noticing this, we have changed this.

92. l343 – 344 “firn densification, despite the long time scale, reduces these snowfall-induced fluctuations by about 15 %”: please clarify how this is calculated. This is calculated from the difference in standard deviation between the cumulative anomalies of Vsnow and Vsnow+vfc. We have specified this.

93. l356 Specify: “captures the strong spatial variation in firn thickness and density observed in our firn core dataset.”. We changed this into: “In Sections 3.1, 3.2 and 4 we presented and evaluated the spatial variation in firn thickness and density.”

94. l363 – 364 Specify: “shown in grey in Figs. 3d and 3e”. We moved this to the methods (Section 2.5), and therefore we do not refer to Figures 8d and e. However, it is explained in the figure caption.

95. -Section 5.1 Please provide maps of seasonal amplitude and of discrepancy in seasonal amplitude.
To keep the size of the manuscript within reasonable limits, we decided not to include figures of the seasonal amplitude in the remote sensing comparison, and to focus on long-term surface elevation trends in the altimetry comparison instead, which is less impacted by the altimetry uncertainties.

96. -l368 – 369 Provide references to support that lower seasonal variability in IMAU-FDM can be explained by altimetry errors.
We agree and we refer to Nilsson et al. (2022).

97. -Figure 7 I find the color codes in this figure confusing. I suggest to show FAC in a color other than blue.
We assume the reviewer means Figure 8. As we dont see which other colors would lead to an improvement, we decided to keep it as is. We made sure the colors combination used is colorblind friendly.

98. -Figure 7 caption Change the caption to: “Time series from FDM v1.2A of FAC, of the cumulative anomalies of surface temperature, of the vertical firn surface velocity, and of the separate components of the vertical velocity from Eq. (6). Time series are shown for (a) the entire ice sheet, (b) the part of the ice sheet situated above 2,000 m a.s.l. and (c) the part of the ice sheet situated below 2,000 m a.s.l.”. Thank you for this recommendation, that is indeed clearer, we have changed this accordingly.

99. -Figure 8 Remove ice shelves from the maps, or show them in a separate color.
The altimetry product we use does not include ice shelves, steep topography regions and the polar gap. However, we still plot FDM v1.1A and FDM v1.2A (Fig. 8a,b) for these areas, as we do not see a reason why not. Ice shelves are indicated by the grounding line, and omitted from all the plots that include RS data.

100. -Figure 8 caption I suggest changing “Maps of average surface elevation change” to “Maps of trends in surface elevation”. Thank you for this recommendation, we agree and have changed this.

101. -l384 “which yields the improvement of FDM v1.2A compared to FDM v1.1p1”: I do not understand why the authors call this an “improvement”. As I understand it, the residual of FDM v1.2A is calculated as altimetry minus FDM v1.2A. The residual of FDM v1.1p1 is calculated as altimetry minus FDM v1.1p1. Thus, subtracting the residuals results in FDMv1.2A minus FDMv1.1p1. In other words, it is simply the difference between both models because the altimetry term cancels out in this operation. If I misunderstand something, please clarify. Otherwise, please revise the use of “improvement” when referring to the difference between the residuals throughout the manuscript.
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 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."

102. l389 Typo: "trends" should be "trend". We have adjusted this.

103. l394 – 395 Please provide a quantitative justification for why the 11 glaciers chose are "representative locations". We remove representative, and explain the reason why we selected these locations. "These locations were selected, as they cover the main distinct patterns from Figure 8, and have continuous observations."

104. l401 "(+9 % sd)": is that compared to post-2003 altimetry? Yes, we have added this: "The altimetry observations prior to 2003 exhibit relatively stronger short-term variability (+9 % sd compared to after 2003)."

105. l401 "likely related to the measurement imprecision": please provide a reference. This we found in Table 1 of Schroder et al. (2019). We will add the reference.

106. l402 "(+13 % sd)": is that compared to FDM v1.2A? Yes indeed, we have added this: "The agreement with FDM v1.2A also appears to increase after 2003, however the altimetry variability remains higher than the simulated variability (+13 % sd compared to FDM v1.2A)."

107. l402 "the altimetry variability remains higher than the simulated variability (+13 % sd)": please discuss possible reasons. Thank you for this recommendation. In the revised manuscript we have discussed possible reasons: The higher simulated variability can partly be related to altimetry errors, as these newer observations still contain noise (<11 cm) and have variations in radar penetration depth, while on the other hand, it can also partly be related to errors in the firn model.

108. Figure 9 Why do the authors use a 6-months running average? This masks out all the seasonality. I recommend using a shorter averaging window, 3 months for example. The seasonality amplitude is discussed in Section 5.1, therefore we chose a 6-month average running period in this section to focus on decadal patterns and reduce the impact of seasonal variation in radar penetration depth. We motivated this in the text: "The time series have been smoothed using a 6-monthly moving average window, in order to focus on decadal patterns and reduce the impact of seasonal variation in radar penetration depth."

109. Figure 9 caption Change "altimetry observed" to "altimetry observations". Done. Thank you.

110. Section 6 This entire section should be thoroughly reworked (see Major comment 1). Done, see major comment 1.

111. Table 4 This table should be thoroughly reworked (see Major comment 1). This has been done, and is divided between Table 2 and Table 5 in the revised manuscript, see major comment 1.

112. l430 "that our results are robust": what do the authors mean here? In contrast, I understand from the results that the MO fits are not robust to realistic uncertainties in climate forcing, and that this sensitivity induces strong discrepancies (Table 4) in FAC estimates (see Major comment 2).
See our response to comment 122. Instead of using uniform changed precipitation (-/+8%) and temperature (-/+ 1.5K) 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 uniform changed precipitation and temperature (<0.5%). In the revised text we now discuss the role of MO fitting in reducing the sensitivity compared to without MO fit (4-7% FAC sensitivity) and contextualise the model sensitivity.

113. l430 – 432 Repetition: "A difficulty of the data to model comparison is that (…) but also hampers the comparison.". We propose to change this into: "Especially in the interior, both the long-term trend and seasonal variability are small compared to the error range of the altimetry observations, which explains part of the remaining discrepancy that exist there, but also hampers the comparison (Verjans et al., 2021)"

114. l434 Typo: "dependent" should be "to depend". Thank you, we changed this.

115. l440 – 441 Repetition: "in these regions (…) in these regions". We omitted the second "in these regions".

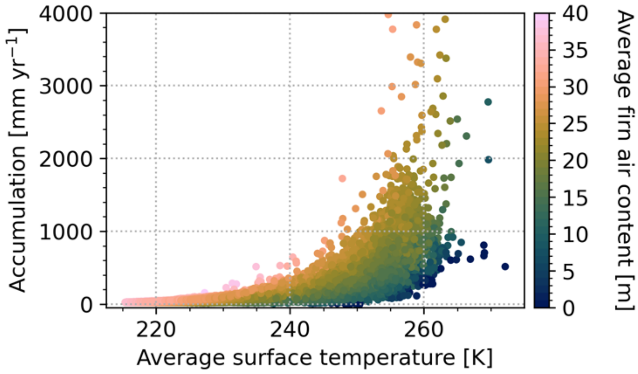
116. l438 – 440 Please reference the work of Medley and Thomas (2019). We added this reference.

117. Section Conclusions Make sure to be consistent in using past or present tense.
Thank you for this recommendation, we will be consistent with this.

118. l446 – 448 This statement of improvement is misleading, because FDMv1.1p is forced in this study with climatic forcing that was different than the climatic forcing used for its calibration (see Major comment 2).
As we explained above, I think there is a misunderstanding, FDMv1.1p2 is not used as comparison to our performance, only in Fig. 9 to show that the improvement in dH/dt is mainly due to updated climatic forcing. Regarding the Z_{830} and FAC: the logarithmic MO_{830} fit has been retuned with the new forcing. Also, the previous fresh snow density parameterization was taken from Kaspers et al. (2004), and therefore not calibrated with the previous forcing from FDM v1.1p1. In addition to that, we now also include FDM v1.1p1 in the surface snow density evaluation and FAC.

119. l450 Typo: "it has" should be "they have". Thank you, we have adjusted this.

120. l451 – 452 "the firn thickness and density patterns vary spatially across climatic regions with with temperature as a primary and accumulation as a secondary driver": Here, I believe that more nuance is needed. Spatial variability in firn thickness is primarily dictated by accumulation rate patterns. Also, the sensitivity tests have shown that the trend in surface elevation change is more sensitive to accumulation uncertainty than temperature uncertainty.



The plot above shows that temperature is a primary driver and accumulation a secondary driver for the spatial variability. Indeed for the surface elevation change, accumulation uncertainty is higher, but that is something different than what we state.

121. l455 "As variations in firn air content and firn thickness align": the meaning of this statement is not clear to me, please clarify. Thank you for this recommendation. We changed this to: "As variations in firn air content and firn thickness have a similar phase."

122. -l463 – 464 "our model in general is robust": again, what do the authors mean here?

We mean that: "In general, uncertainties in the model formulation, forcing or forcing during the spin-up cause rather small changes in simulated FAC (<5.2%). On the other hand, uncertainties in the accumulation forcing, and in the accumulation and temperature forcing during the spin-up period can lead to substantial differences in simulated surface elevation change (between 4-19 mm/yr)."

We will specify this in the revised manuscript.

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.

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.

Herron, M. M., & Langway, C. C. (1980). Firn densification: an empirical model. *Journal of Glaciology*, *25*(93), 373-385.

Judson, A., & Doesken, N. (2000). Density of freshly fallen snow in the central Rocky Mountains. *Bulletin of the American Meteorological Society*, *81*(7), 1577-1588.

Morris, E. M., & Wingham, D. J. (2014). Densification of polar snow: Measurements, modeling, and implications for altimetry. *Journal of Geophysical Research: Earth Surface*, *119*(2), 349-365.

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.

Ligtenberg, S. R. M., Helsen, M. M., & Van den Broeke, M. R. (2011). An improved semi-empirical model for the densification of Antarctic firn. *The Cryosphere*, *5*(4), 809-819.