Review of (revised version of)
*Improved representation of the contemporary Greenland ice sheet firn layer by IMAU-FDM v1.2G* by Brils et al.

Reviewer: C. Max Stevens

General comments:
I appreciate the work the authors have done to address the comments from the other reviewers and me on the previous version of the manuscript, and I think the current version is a significant improvement. I recommend that the paper be published after minor edits. I have provided line-by-line comments only. The gist of most of my comments has to do with writing clarity: in numerous places I found explanations to be unclear, and in some cases I was looking for more details about methods or assumptions.

21: the word 'by' occurs 3 times in this sentence

The sentence has been reworded to avoid too much repetition.

27: I think you mean changing the mass, not changing the mass balance – mass balance in my mind refers to the mass sum of accumulation and ablation processes.

Thank you for pointing this out. We indeed mean mass and not mass balance.

34: Not sure if it is worth mentioning here or following paragraph, but recent work by Rennermalm et al. (2022) suggests that pore-space loss (in SW Greenland, at least) is not entirely irreversible.

This is indeed true. However, while pore space loss is not strictly irreversible, in practice the rate at which pore space is recovered is much lower than at which it is lost, especially in regions of low accumulation. Nevertheless, we agree that "irreversible" is too strong a word and have reworded the sentence accordingly. We have also added a reference to the paper of Rennermalm et al. (2022).

50: be specific of what kind of observations – density? Temperature? Depth-age?

Density, temperature and depth-age relation are all commonly interpolated with firn models. These have been added to the text as examples.

52: can you give an example or two of a coupled RCM/firn model?

Examples of such climate models are RACMO and MAR. This has been added to the manuscript.

59, Section 2.1, and Section 3: You define v.1.2G on line 59, but you do not in the paper explicitly define what v1.1G is – is that model version described in Kuipers Munneke et al. (2015)? Given that section 3 is mostly comparing the outputs from the two versions, I think that more detail is needed at the start of section 3 describing the comparison. For example, at the end of the paper I was still not sure if the v1.1G and v1.2G results were both produced using the same version of RACMO (i.e. the same forcing) with different FDM physics, or if the v1.1G is uses a different, older version of the RACMO forcing. A short paragraph describing the two model runs that are being compared will help.

IMAU-FDM v1.1G is indeed Kuipers Munneke et al. (2015). Both models have been ran at the same resolution and are forced with the same forcing, the only difference between the two runs is the model physics. This is now explicitly mentioned in Section 2.1 and Section 2.2, which discuss IMAU-FDM and the RACMO forcing. This way, no entirely new paragraph is needed.

71-72: this neglects to mention Section 5. If you are outlining the paper with this amount of detail I think it is worth mentioning that section as well.

Thank you for pointing this out. We have added a mention of Section 5 to this paragraph.

159: This is written as if you have already introduced the new set of observations; I suggest "In order to calibrate Eq. 3 to a new, expanded set of observations (Section X.X), we …". Also, consider rewording here to indicate that MO is changed from the previous model version but retains the same general form.

The sentence has been changed according to the reviewers suggestion and we have added a sentence to indicate that the form of the MO stays the same.

169: I think there is an issue of confusing wording – I would think that the previous calibration used 22 cores and the new calibration used 29 – but this implies that previous calibration used an expanded data set? Reading forward to section 2.4, it is a bit unclear also – there are 123 observations. Those are used just for evaluation? I think it would be helpful if you added 1-2 sentences in section 2.1.2 describing the new observations used in the present work, or add a bit more detail in section 2.4 (in which case make a reference to 2.4 in section 2.1.2) about the data that were used for the new calibration. (I am not suggesting you list all the calibration cores; rather, just add a bit of text differentiating the calibration data and the evaluation data. E.g., are the calibration cores a subset of the evaluation data, or do you keep them separate?)

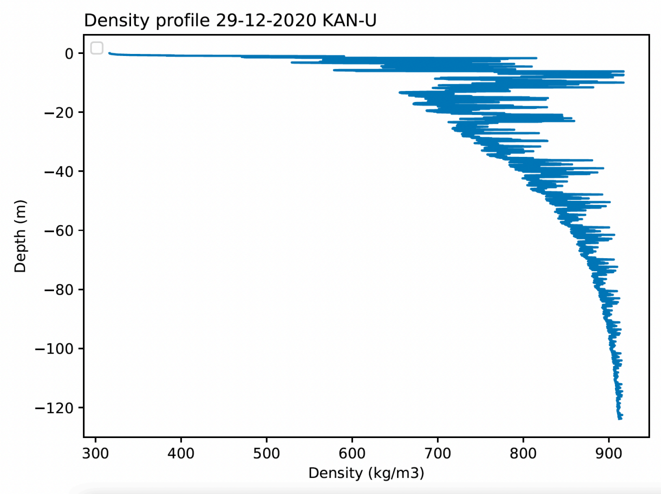
As suggested by the reviewer, we added some sentences to section 2.4 in which we more clearly explain that we use 123 observations, 92 of which are firn cores (compared to the 62 cores used in Kuipers Munneke et al. (2015)), one measurement year of liquid water GPR measurements and the remaining observations being measurements of the temperature at a depth of 10 metre. We refer to this section in section 2.1.2.

176: I think this is what you get at in the following paragraph, but you could be more explicit here: the very low $r^2$ value for $MO_{550}$ indicates that the linear model is not any better than just using the mean of the data (0.67). So: does the new model use the $\alpha_{new}$ and $\beta_{new}$ for $MO_{550}$, or do you just use a value of 0.67? If the former, can you further justify your choice given the low $r^2$?

We use both the $\alpha_{new}$ and $\beta_{new}$ for $MO_{550}$ instead of using the mean of 0.67. The main reason for this is that in this way we retain the same general formulation for both MO550 and the MO830.

239: I agree with you that this is probably a fine assumption to make on the spatial scales you are looking at – but I would appreciate a bit of discussion on the implications of that vis-à-vis the discussion of ice lens formation in your introduction: can thick ice lenses/slabs form while using this assumption, or is it necessary to be able to include ponded water?

While it is true that our model does not simulate melting water, it does not prevent the formation of ice lenses in Southwest Greenland. This can, for example be seen in this profile from KAN-U in 2020:




It must be noted, however, that the modelled ice slabs consist of multiple thin layers of high density ice, often interlaid with a thin layers with slightly lower density. In reality, a thick ice slab was formed at KAN-U in 2012 (see for example Rennermalm et al. (2021)). The absence of ponding may be what is causing this behaviour. We have added this discussion to the manuscript, and the firn density profile has been added to the supplementary material.

298: Ok, now I see that you are using 92 cores and the fitting cores are included. I think it would clarify your paper if you add a bit more detail in Section 2.4 about the observations and how you are using them. For example, I was expecting that you used 123 cores based on reading the start of section 2.4. Now I see that the number 123 refers to observations in general – it would be much clearer if you specified e.g., ‘we used 92 depth-density profiles, X depth-temperature profiles, and one observation of meltwater intrusion’, or something along those lines.

See our response to the comment on line 169.

306: “up to a depth” is colloquial and somewhat self-contradictory (up is the opposite of depth perhaps?).

This has been changed to “to a depth”

310: comma after FAC, and remove word ‘obviously’

We have taken over this suggestion.

Figure 10: is it possible for you to label the study sites on the figure?

Although it would definitely be possible to add the names of each study site, we feel that it would clutter the image too much. Instead, the name of every site can be found in the supplementary file. We have added a reference to this list to the caption of the figure.

331: please add units on 2.0x10^3. (And note: I think these should be formatted with a latex \times rather than a dot.)

The units have been added to this sentence.

345: It seems that you are initially talking about Das 2 (336), and then it seems here you are talking about Summit? Which is it? Looking at the figure, I am guessing that you may have mislabeled Summit as Das 2? If it is indeed Das 2, please consider changing the figure to be for Summit rather than Das 2 to be consistent with your section 4 (i.e., you provide specific information about Summit, but not Das 2.) Likewise, consider changing Figure 1 to be density at one of your 3 case study sites.

Thank you for pointing out the confusing wording. This is a mistake in the manuscript: we should refer here to Das 2 and not to Summit. This has now been fixed. We intentionally did not opt to show only results from the three case study sites discussed in Section 3, since that may give of the impression that these three sites are the main focus of our work and of this manuscript. Instead, we want to demonstrate the model's performance over a wide range of sites. We opted for Das 2 and FA-13 to demonstrate the model's performance at two different climates (wet vs dry). However, we do agree that the density profiles can be insightful and may help interpreting our results. Therefore, we have decided to add the density profiles of Summit, KAN-U and Dye-2 to the supplementary material.

351: The new model does fit the upper density better, but it is still a rather large misfit in the upper firn. I would like to see a bit more discussion of what is causing that misfit, or at least acknowledgement that there is a deficiency in the model at this sort of site – and I don't mean to pick on this model in particular, because it is a deficiency in firn models in general probably.

The modelled density in the upper layers of the firn layer fits the observed density profile better. The old model clearly overestimated the density near the surface. The imporved performance can be attributed to the lower surface density and the new MO fit. Despite the improvement, the densification rate in the upper region is still too high. This may be attributed to the lack of a description of microstructural properties on the firn. In the presence of liquid water the rate at which snow grains grow is increased. Firn with larger grains lead to a lower densification rate. This feedback is currently not present in the model. The presence of liquid water may also reduce the densification rate in a different way: it reduces the effective stress felt by the firn layer, which is the driving force for densification. This process is often observed in soils, where it is called consolidation: initially water takes up a change in stress before the soil matrix. To our knowledge, however, the influence of the pore water pressure on the effective stress has not been investigated in the context of firn densification. We have added this discussion to the manuscript.

360: I know this was picked out in the previous reviews, and it is still not entirely clear: first you say that RACMO2 has a cold bias, but then you say it has a warm bias. Are you saying that the RACMO biases vary spatially? I think changing the text here a little bit will clarify this significantly. Perhaps the issue is the word 'model' – RACMO2 is a model (RCM), and the FDM is also a model – so when you say "a persistent warm model bias remains", it is not clear if you mean a persistent bias in RACMO or resulting from FDM physics.

The bias indeed seems to vary spatially: in some regions the 10 m temperatures are too warm and in others they are too cold. In the manuscript we mention first that there seems to be a cold bias at the warmer locations and a warm bias at the cold locations. We admit that the wording was confusing, and we therefore reworded this section. In the dry interior the error is likely due to an error in the RACMO forcing. In wet areas the error is likely due to missing physics in the way the model handles meltwater and thus refreezing. We have made this more clear in our discussion.

371/Figure 8: It is not clear to me that the temperature maximum is the refreezing depth – can you justify further why this maximum is assumed to be the refreezing depth? Further, the width of the blue summer DYE-2 temperature curve at ~1m depth (the maximum) would indicate to me that diffusion has happened rapidly, not slower as you posit; i.e., a melt event would cause the firn to be at the freezing temperature, causing a large temperature gradient, which will diffuse much faster than a smaller temperature gradient. I am willing to believe your explanation, but in their current form the explanations seem incomplete.

The maximum in the liquid temperature profile is assumed to be the refreezing depth because from the liquid water measurements we see that liquid water usually penetrates and refreezes around these depths (see Figure 9). Moreover, Summit does not show a similar peak in the temperature profile at that time, indicating that it is likely not caused by the temperature at the surface. Furthermore, the new formulation yields of the conductivity is lower than the old formulation at densities below ~500 kg/m3 (see Figure 3). This, together with the lower snow density, leads to slower diffusion rates. Indeed, the temperature gradients have become larger in both the Summit profile as well as in the Dye-2 profile in the upper 1$^{st}$ metre.

392: How are you calculating RMSE of penetration depth and volume fraction? E.g. is one time step in the model compared with one observation over that time period? What is the temporal resolution of the upGPR data? Are you including all of the zero-water periods in your RMSE calculation?

The zero-water periods are included in the calculation of the RMSE. Since the observations have a higher temporal resolution than the model output, we compare each model time step with the nearest observation. We chose for this approach because we would like to compare the liquid water in the firn column over the whole measurement period. If there is no water detected by the GPR, then we would also like the model to simulate no water in the firn column and vice versa. Therefore, computing the RMSE over the whole period gives the most complete view of the model's performance. We have added this information to the text.

438: change 'like': "… in exceptional years; for example, 1983 was …" (and change to past tense "was")

This has been changed.

459: I appreciate the work done for these uncertainty analyses. Can you provide a bit more information that summarizes these sensitivity runs? It is not clear to me from the text for example how many runs were done, and I am curious if the accumulation, melt, and temperature variations were applied in simultaneously in single run? You say one-by-one, but the ensuing sentences do not make it clear if each sentence describes one of those runs or several runs. Perhaps a table in the supplement might work? E.g.

| Run # | Variation |
|---|---|
| 1 | Increased Density by X |
| 2 | Decreased accumulation by Y% |

In order to make it clearer how we conducted the sensitivity analysis we added a table listing the experiments, as suggested by the reviewer. We also slightly altered the text for clarity. During each sensitivity test, only one of the variables is changed at a time. Then, the resulting uncertainties are added together quadratically.