Reviewer #3

The manuscript "Runoff from Greenland's firn area – why do MODIS, RCMs and a firn model disagree?" by Machguth et al. describes a study where the authors developed an improved algorithm to estimate the runoff limit using MODIS. This result is compared to simulation output from firn models in RCMs. A mismatch is found in the extent of the runoff area between MODIS and the modelled runoff area, as well as discrepancies that are present between firn models. This is then investigated further by analyzing detailed firn model output along a transect, which reveals that the way water retention is treated is the main cause for the discrepancies.

The study is very relevant: meltwater runoff from the ice sheets is a major souce of sea level rise, and is in fact quite uncertain (as clear from this study). The study is informative for the further firn model development acitivites in the firn community, and I think it's well suited for publication in The Cryosphere. Nevertheless, revisions are necessary, to improve the substantiation and clarity of the results. I think that a bit more analysis may be required to make the study more relevant. Now, the detailed analysis of possible causes for the model discrepancy is restricted to 1 transect only, which makes it uncertain how well the findings translate to the Greenland ice sheet runoff area as a whole.

We thank the reviewer for their comments. Below we answer all of the reviewer's comments.

Main concerns:

1. My biggest concern with the study is that basically only 1 transect is used to substantiated the conclusions. The argument that this transect has been used very frequently in studies, given the wealth of detailed field observations available is a bit weak, because very little field observations are in fact used. For example, no density-depth profile is shown to indicate if the density profile simulated by FDM or by MAR is in closer agreement with field observations. The majority of the results section describes the discrepancies, rather than explain them (as the title "why..." would suggest). RACMO FDM shows a strong transition from basically full ice to firn with substantial pore space around -47.5 longtidude. Immediately, this raises the question if this is general behaviour from FDM, also found further north? I think that the authors should try to find a way to make the discussion and results more robust, by analyzing larger parts of the ice sheet. Similarly, Fig. 2, A1 and A2 basically show one flowline. Are the results robust and can be extrapolated to other flowlines?

We agree that it would be good to analyse more profiles. However, we feel that we cannot do this. The reviewers also request additional explanations and interpretation of our K-transect analysis. To keep the manuscript at a manageable length we would rather address these requests than show more transects. However, we now discuss our findings in the context of the study by *Glaude et al.* (2024) which shows that the two RCMs studied here strongly divert in their predicted runoff areas for the year 2100 and that the differences are large in all regions. These independent results (i) confirm that under strong melting the two RCMs divert strongly in simulated extent of their runoff areas and (ii) that all regions of the ice sheet show this effect.

2. An aspect that I think is slightly difficult to grasp for readers not familiar with the topic, is how the different definitions of runoff are used. When it comes to modelling, a very clear definition is water that leaves the firn column. But here, a crucial difference is that in FDM, water can only leave at the bottom, whereas in MAR, it can leave at the bottom or laterally. Furthermore, bottom in FDM is much deeper down than in MAR. When it comes to MODIS, runoff is estimated from the slush limit, an assumption that is reasonable, but doesn't come without caveats.

Maybe a sketch can help to establish clear definitions. Note that in the text, it is is not clearly defined what runoff is for the used firn models. For FDM, it is not mentioned, and for MAR, only lateral runoff is explained. But it should also be explicitly mentioned that water leaving the firn column at the base is considered runoff.

We thank the reviewer for pointing out this inconsistency and we have added text to clarify the definitions.

3. Fig. 5: Here, I with panels l and k (what happened to panel i and j?) should also include water retained in the firn column as liquid, not only the refrozen part. I think this needs to be included to better explain how the differences in water retention parameterization are responsible for the differences. Moreover, I wonder to what extent it plays a role that the FDM simulated firn column is much deeper. For example, the additional refreezing in FDM shown in 5l is partly caused by the fact that the firn column is deeper in FDM than in MAR. Maybe these figures should be standardized to only show the uppermost 10, or 20m.

It is panel "i", though the way the character is rendered makes it a bit difficult to distinguish from "l". Furthermore, we omitted the letter "j" as is often done in labels due to its visual similarity with "i". The figures are already standardized and show data only for the top 20 m. No data is available for IMAU-FDM liquid water content below 20 m depth as these data are not written to output. We have now clarified to which depth the summed values refer. Liquid water contents for the top 20 m are already shown in Figs. 6c,d (RACMO-FDM) and 6i,k (MAR). The plots show that liquid water content is substantially higher in MAR. We note that Fig. 5 shows time integrated values and the meaning of time integrated liquid water content would be less clear than e.g. time integrated refreezing or runoff. For this reason, we prefer to show the temporal evolution of liquid water content in Fig. 6. We noticed that there was no indication to what depth the summed liquid water content and refreezing refer, we have added this information.

4. A bit in line with my previous comment: I find it hard to understand how the thermodynamics are different between MAR and FDM. Given that both are driven by ERA5, I assume that the overall energy balance should be more or less similar. The 10m-depth temperature, however, varies largely between MAR and FDM (Fig. A3). I find it very hard to grasp if this is only due to the percolation scheme, or that the firn is generally warmer, or if the surface energy balance is higher in MAR. Also, I think it easily gets confounding that the firn layer in FDM is so much deeper than MAR. This would allow for much more cold content to be stored at depth in FDM than can ever be captured in MAR. I wonder if the authors can give a bit more insight in this. Maybe calculate the uppermost 10m cold content of the firn layer, to see to what extent surface energy balance, and firn temperatures in general differ?

Both models simulate similar amounts of surface melt (Figs. 5e,f). This means that overall, the surface energy balance between the two models cannot differ strongly (the evidence based on melt is valid for summer, but Fig. A4 shows that winter surface temperatures are similar too). In 2012, there is a remarkable jump in RACMO albedo, directly above the 2012 runoff limit (Fig. 5c). This sudden increase in albedo also leads to a sudden drop of melt (Fig. 5e). This sudden change in RACMO albedo has been discussed in detail in the manuscript and it is not present in e.g. 2017 (Fig. 5e).

The thermodynamics of firn involve a complex interplay of thermal conductivity and latent heat release from percolating meltwater, which are both strongly influenced by firn density. Positive and negative feedback mechanisms are also present, such as the percolation depth of water depending on firn temperature and density. If for some reason a model transports meltwater in summer to too deep a depth, then latent heat release cannot be fully undone in winter through heat conduction (porous snow and firn having relatively poor heat conduction). The amount of surface melt, controlled by the surface energy balance, plays an important role in controlling the temperature of a firn pack. However, as shown in the manuscript, the differences in melt are relatively small between the two models while firn temperatures differ very strongly (Fig. A4). For these reasons, we state that the water percolation scheme is mainly responsible for the observed differences. As mentioned in our replies to Reviewers #1 and #2, we now analyse differences in melt in more detail and provide an extended discussion of the reasons behind the large differences in simulated runoff limits.

We agree that the thickness of the firn pack plays a role. However, the depth of zero annual temperature amplitude in firn is typically around 15 m. Firn deeper than 20 m contributes to near-surface firn temperatures only through slow heat conduction.

- I would check for consistency in using the terms "saturation" and "water content". For me, saturation is the part of the pore space taken up by water. Thus, 100% saturation means all pore space filled by water. In contrast, liquid water content is most often defined per volume. Thus, 100% saturation in firn with density ~450, would mean a liquid water content of ~50%. So for example, in L111, both water content and saturation are used in the same sentence, and I'm not sure if the 7% refers to saturation, or to liquid water content. In L126, a percentage of 13% is mentioned, but it's not clear if it refers to saturation or liquid water content, since it's only written "irreducible water". Given the numbers, I think 13% is a value for saturation, not liquid water content. Anyway, I would encourage the authors to thoroughly check this throughout the manuscript, because I think now the percentages given are a mix between saturation and liquid water content values.

We have revised the text to improve clarity.

Minor comments:

- L8: "where meltwater is routed" --> "to route meltwater vertically"

Done

- L25: "found to perform well" is too general. Performs well in terms of? Or on what variables did it perform well? In terms of mass balance, or calculated melt?

We have added more details to the statement.

- L36: "over which mass loss takes place". I would add "mass loss through runoff", because sublimation and wind erosion are also mass loss terms.

Agreed.

- L39-40: Obviously, the choice of forcing model can have an affect as well. Forcing both RCMs with the same boundaries removes a large part of uncertainty that would come from the GCMs. This is not really a drawback, because it allows to compare RACMO and MAR on more equal footing. But maybe a brief remark on how well ERA reproduces Greenland climate and is suitable to use as forcing along the model boundaries could be justified here.

Certainly, the choice of driving re-analysis is important. However, we do not fully understand why an assessment of the various ERA products over Greenland is relevant here, with respect to the study by *Tedstone and Machguth* (2022). Our study shows that the RCM simulated runoff limit either overestimates the MODIS runoff limit (in the case of MAR) or underestimates (in the case of RACMO-FDM). Thus, there is no systematic bias between RCM (modelled) and MODIS (observed) runoff limits. If there were a systematic bias then the RCM forcing (ERA-5) would need to be investigated, among other potential issues. We have carried out additional analysis with focus on simulated meltwater input and added a more detailed argumentation, explaining the role of meltwater and why the main issues are the differences in the firn parameterisations.

- L74: I suggest to specify that these concern output time resolution. Like: "MAR output and RACMO 1 km downscaled data" and "Output from RACMO and IMAU-FDM are at ..."

Done

- L134: "150 vs 3000 layers". Maybe specify if this covers equal total firn depth?

Both models cover the entire firn column, whose depth varies in space and time. The difference in the number of layers thus reflects their different vertical resolution, which is most pronounced at greater depths. We have added this information to the text.

- L153: "is rather insensitive": this is actually an interesting point. A figure and a full blown sensitivity study is not necessary, but maybe a statement like: using a threshold of X compared to Y affected the runoff limit only by Z kilometers would be useful here.

A sensitivity analysis on this subject is already included in *Tedstone and Machguth* (2022), visualized in Extended Data Fig. 8 and discussed therein in the Methods section under the subsection "Runoff Volumes". The sensitivity analysis shows a very low sensitivity for MAR and a somewhat larger sensitivity for RACMO. However, the latter is mainly due to the 1 km resolution of the downscaled product used in *Tedstone and Machguth* (2022) while MAR is at 10 km resolution. Hence, the low sensitivity of MAR is more representative for this study where RACMO-FDM is at 5.5 km resolution. We have added more quantitative information from *Tedstone and Machguth* (2022) rather than only a reference.

- L205: "The plateau is shorter in the north than in the south." is a too general statement given what is shown in the figures. In fact only one transect in the north or in the south is shown. Can this indeed be generalized?

We have removed this statement as it was also questioned by reviewer #2. The duration of the plateau is not relevant for the present study.

- L171: The way it was written, this actually confused me a bit, because I was looking for the masked points in Fig. 1. I would maybe write it more explicitly, like "Fig. 1 does not show retrievals from 60 to 68 ...."

Thank you, we have changed the wording.

- L245-246: "and thus could not warm further" is a bit poorly phrased, since this would also limit refreeze once temperatures reach 0 degC.

Agreed, it is obvious that 0 °C firn cannot warm further, we have modified the sentence.

- L311: Is it really inertia? The example given in the next sentence sounds to me more like non-linear behaviour.

The term was also criticised by reviewer #1. We have changed the wording, avoiding the confusing term inertia.

- L347: Please correct: "A secondary reasons"

Done

- L393: "later water flux" --> "lateral water flux" (I assume?)

Done

- Please avoid the use of green and red in one figure for color blinds (Fig. 1 and A4)

Changed

- Generally, I think the figure captions are too short.

We have added more information where needed, also with respect to comments from the other two reviewers.

References

Glaude, Q., Noel, B., Olesen, M., Van den Broeke, M., van de Berg, W. J., Mottram, R., et al. (2024). A factor two difference in 21st-century Greenland Ice Sheet surface mass balance projections from three regional climate models under a strong warming scenario (SSP5-8.5). *Geophysical Research Letters,* **51**(22). https://doi.org/10.1029/2024gl111902

Tedstone, A., & Machguth, H. (2022). Increasing surface runoff from Greenland's firn areas. *Nature Climate Change,* **12,** 672–676. https://doi.org/10.1038/s41558-022-01371-z