We thank the three reviewers for their comments, and we hereby submit a modified study which hopefully addresses all reviewer comments.

- We have rewritten and rearranged larger sections of the text, mainly in the description of the RCM bucket schemes and in the discussion.
- We have added additional analysis (i) comparing MODIS and LANDSAT runoff limits, (ii) quantifying the differences in MAR and RACMO melt and accumulation at the runoff limit, and (iii) of the position of the ELA and runoff limit in order to explain certain albedo patterns.
- We have redrawn all but one figure and we have added three more figures. One figure has been moved to the appendix to keep the main text shorter.

We are aware that the manuscript has become relatively long but the reasons behind the different modelled runoff limits are complex and pinpointing them requires to analyse potential other factors which were not addressed before. Below we reply to all individual reviewer comments. Note that this document has been updated from the initial reply to reviewers, submitted on 27 February 2025. This updated document reflects further changes that were implemented while revising text and figures.

In the following the reviewers' comments are in black, our replies are in blue, and the description of our changes is highlighted in red.

Reviewer #1

Review of Machguth et al. 2024

This paper provides an update to a methodology to determine runoff limits on the Greenland Ice Sheet from MODIS and in addition investigates differences in runoff from these observations as well as MAR and IMAU-FDM models, partially in RACMO.

While there is clearly some new science in this paper I would argue it needs some major revision if it seeks to answer the question posed in the paper title. The research goal of the paper is unclear, is it to introduce an improved method for using satellite data, or to compare different methods for calculating runoff? I would argue that for either improvements are needed, due to the following reasoning.

Thank you for your comment. Our main goal is to answer the question, to what degree do RCM simulations and MODIS observations of the runoff limit agree? This required us to make some improvements to our remote sensing methodology originally introduced in *Machguth et al.* (2022), mainly increasing its flexibility to be used ice-sheet-wide. The modifications are not substantial enough to merit a dedicated publication. We have made our motivation clearer throughout the text and modified the wording, where needed.

1. The new methodology for improving detection of runoff limits from MODIS is not compared with other remote sensing methods, or with the previous method in Machguth 2022. Although the improvement here is that the method can now be used in more areas, without any validation or comparison it is impossible to judge the validity of the method.

   We considered our changes as minor in the sense that the basic method of detecting the runoff limit remains the same. The actual runoff limit retrievals will not be influenced by whether they are being derived along east-west transects or in flowline polygons. However, we understand that these considerations are not obvious to the reader, so we now provide a similar comparison to *Tedstone and Machguth* (2023) as was provided in *Machguth et al.* (2022). This comparison is also discussed.

2. The authors make suggestions as to why MAR and RACMO/IMAU-FDM may differ but can't actually evidence this. An RCM and a firn model are very different things, and while it is clear that there is a difference in the implementation of vertical water percolation between the two, that is not the only difference between them. A firn model is run on a very different vertical resolution and will be dependent on RACMO forcing in this case, which will also influence the runoff. The bucket method may only be part of the story as the paper does not compare like for like, or truly compare the two RCMs, or explain why they differ from MODIS completely.

   We believe there are two aspects that need to be distinguished.

   (1) We agree that an RCM is not the same thing as a firn model. However, the polar RCMs we investigate here have firn models embedded. These embedded firn models do not differ fundamentally from dedicated firn models (such as IMAU-FDM). We compare an RCMs firn module to a firn model. We made this clearer throughout the text.

   (2) We disagree that differences such as model vertical resolution prevent simulations of firn from being compared to one another. For instance, two given firn models can be more different from each other than two firn modules in two RCMs (such potential is indicated in *Vandecrux et al.,* 2020). Specifically concerning vertical resolution, even if we were able to make the resolutions of both models identical, this would likely create new issues that would severely hamper comparison. For example, MAR includes a parameterization that a certain percentage of water runs off immediately when it encounters a layer of the density of ice. If we were able to run MAR at the same high resolution as the RACMO (or IMAU-FDM) firn module, this parameterization would only allow water to reach the bottom of the model domain very rarely thanks to thin layers of ice in the simulated firn column. The parameterization is thus optimized for the coarse MAR firn layers. There are more parameterizations in MAR, RACMO and IMAU-FDM that are optimized to work in the context of the modelling framework they are in.

   Furthermore, we emphasize that MAR's and RACMO's firn modules, as well IMAU-FDM, are used to simulate the same processes, one of them being the interaction of firn and meltwater. Given that it is not possible to run all of these models with, for instance, the same common resolution and inputs, our solution is to compare model output and to assess qualitatively which parameterizations likely contribute most to differences in output.

   While we believe that our findings show that differences in the implementation of the bucket scheme strongly influence simulated runoff limits, we agree that interpretation needs to be done cautiously. To do so, we have carried out further analysis with a focus on simulated melt and how MAR and IMAU-FDM simulated runoff limits react to changes in melt. We present and discuss these additional analyses. We also more clearly distinguish RACMO and IMAU-FDM. Finally, we explain in more detail why the findings related to IMAU-FDM are also relevant for RACMO.

Below I provide line by line comments to add detail to the above.

Title- The paper doesn't really compare RCMs (plural). It jumps from sometimes including RACMO, sometimes IMAU-FDM forced by RACMO, but to really compare RCMs a full comparison of MAR and RACMO would be needed.

We agree, such a comparison would be favourable. However, RACMO detailed firn output is not written to output files. RACMO does provide depth-integrated variables at a daily resolution (firn air content, total water content). However, other variables such as vertical density, temperature and liquid water profiles are only available from IMAU-FDM. IMAU-FDM outputs these data at 10-daily (for the uppermost 20 m) or 30-daily resolution (for the entire firn column) because otherwise the file sizes become unmanageable. Furthermore, also the IMAU-FDM firn output is used in numerous studies. Hence, we decided to compare MAR and IMAU-FDM. We use plural RCMs because certain parameters need to be compared between MAR and RACMO. We explain more clearly why we need to use IMAU-FDM for a detailed comparison to MAR.

Line 9- The paper does not demonstrate that the difference in implementation of the bucket scheme are responsible for the disparity, only that this is a possible explanation posited by the authors.

We refer to our detailed answer to the reviewer's major comment #2. We have added Appendix B to analyse whether differences in firn model/module input (accumulation and melt) are responsible for the observed differences in runoff (see also our answers to reviewer #2). We have revised the text to more clearly argue that the percolation schemes are the main reasons for the observed differences. To do so, we have also added an analysis of the position of the ELA with respect to the position of the runoff limits.

Line 18- Suggest addition of 'among' our most advanced tools here. Remote sensing, firn models etc. are all advanced tools that contribute to our understanding.

The statement has been removed (see our reply to the following reviewer comment).

Line 19- Do all these papers actually say RCMs are our most advanced tool? E.g. are they not a part of some of the methodologies described in IMBIE?

These papers were cited as examples of assessing "past, present and future surface mass balance of the Greenland and Antarctic ice sheets". As this was unclear, we revised the sentence to simply say "are widely used to assess …"

Line 46- Confused by the use of the word 'oppose' here, how do you oppose the runoff limits?

Replaced by "compared"

Line 46- Landsat is mentioned here- why would MODIS be used instead of Landsat? The context is missing here, and a reference for any work done with Landsat as well as justification for why the results in this paper are not compared to any results using Landsat.

This was unclear. We now introduce our MODIS study (*Machguth et al.,* 2023) earlier in the introduction and then state that we use MODIS runoff limits in this study because of their higher temporal resolution.

Line 75- This is further evidence that MAR vs IMAU-FDM is not a straightforward comparison and that there may be other reasons for differences between them.

We agree that it would be better to have all data at the same temporal resolution, optimally 1 day. Unfortunately, the data are provided at different temporal resolutions. Nevertheless, we believe that a comparison is valid (see our detailed answer to the reviewer's major comment #2) and that we provide clear evidence that differences in the implemented of the bucket schemes contribute to the observed differences in modelled runoff limits. We have clarified our argumentation, also to address comments by reviewer #2.

Line 80- Again this makes me more confused why Landsat isn't used, or at least compared against for the MODIS methodology.

We hope this has now become clearer as we have modified the text to better explain why we here use MODIS data.

Line 94- Does the new method give the same results for the areas covered in Machguth 2022?

We now show the comparison of MODIS (improved methodology) to Landsat (*Tedstone and Machguth,* 2022) Landsat runoff limits.

Line 105-'used no more' or better might be 'not used'.

Agreed.

Line 105- Is the difference between clean and dirty ice a function of water depth? This doesn't quite make sense, clean ice, dirty ice and ponded water are different things.

The aim of this parametrisation was to represent the supraglacial lakes (the albedo of water) as explained in *Lefebre et al.* (2003).

Originally, the bare ice albedo could vary between 0.15 and 0.55 as a function of water depth but in view of the spatial resolution of MAR (not explicitly resolving the presence of lakes) and the lack of observations to validate such a parametrisation, we have reduced a lot the possible range of bare ice albedo values (0.5-0.55) to decrease a lot the importance of this parametrisation in the recent MAR versions. We have added an explanation in the text.

Line 137- Could the differences in albedo scheme also contribute to the differences in runoff found between this model and MAR?

The different albedo schemes certainly have an influence. For this reason, albedo is compared in Figs. 5c and d and melt (directly affected by albedo schemes) and runoff (not directly affected by albedo schemes) are compared in Fig. 5e and f. The figure shows that at higher elevations the differences in runoff are much larger than the differences in melt. This is one of the key reasons why we conclude that differences in parameterizations of firn meltwater percolation, rather than melting, are mainly responsible for the differences in runoff. At this occasion we also point out that for low elevations MAR shows higher surface albedo and higher melt rate (Fig. 5e and f). This is contradictory but exploring this issue is beyond the scope of this study. We have updated the entire discussion why the two models differ. Thereby we now clearly state that the differences in albedo are also contributing to the differences. We highlight how the position of the ELA influences RACMO albedo fields and how this helps to reduce mobility of the runoff limit.

Section 3.2 Please state clearly how IMAU-FDM does (or doesn't) deal with ice lenses given this is a key difference with MAR.

The description of the bucket schemes has been expanded. We hope this is clearer now.

Line 172- Could the lack of masking for smaller aquifers influence the results?

Where smaller aquifers are present, the reliability of MODIS visible runoff limits is negatively affected. In such areas, our detection method based on MODIS (as well as the approach based on Landsat) probably underestimate the *actual* runoff limits as optical remote sensing can only detect the *visible* runoff limit. We speak of a probable underestimation as there is not other "ground truth" available. The effect is especially clear in Figure 3 and is mentioned on lines 195-196.

Line 184- How do the detections in the Tedstone paper compare to those made here?

We have added Figure A2 which compares MODIS runoff limits to LANDSAT and a discussion of the comparison. We hope this answers the reviewer's question.

Section 4.2.1 I found this section and the jumping between RACMO and IMAU-FDM confusing. I would suggest either having a full comparison of MAR, RACMO and IMAU-FDM, or removing the RACMO here as it's not clear what it adds when it's only partially included.

We have modified the text, trying to make the arguments clearer. We agree with the reviewer that including the downscaled RACMO 1 km adds to the complexity, but we would like to keep it as these data are widely used. Furthermore, these data help to show that the behaviour of the IMAU-FDM runoff limit (small inter-annual variability, at relatively low elevation) is very similar in RACMO.

Figure 7- This figure to me is the one that really made me question the comparisons made here. It shows very clearly that MAR and IMAU-FDM are working on very different scales and thus capturing different processes, and the suggestion that differences are due to bucket methodology over simplifies this.

As we note in our response major comment #2, we believe that the two models *need* to be compared despite (and indeed because of) their scale differences. Ultimately, they both numerically calculate the same quantity: Greenland surface mass balance including simulating firn and its reaction to melt. They both are used in numerous studies to assess how the Greenland ice sheet is currently changing and will react to climate change, they both are widely used to assess runoff, sea level rise contribution and so on.

A recently published study (*Glaude et al.,* 2024) shows that under identical forcing, three RCMs differ very strongly in simulated Greenland runoff extent by the year 2100. The difference in simulated runoff areas is particularly striking. The study points out that there is a factor two difference in simulated mass balance. Hence, models need to be compared to understand why their output differs, whether due to e.g. spatial resolution (for which we find no evidence) or firn parameterizations (for which we find evidence). We have addressed the reviewer's general critique as outlined in our replies to their general comments. We have also added reference to the study by *Glaude et al.* (2024).

Line 311- Is this a lack of inertia or just that processes are not accounted for?

The lack of inertia is the result of processes not being accounted for. We have accordingly changed the wording.

Line 318- This paragraph is confusingly written. It starts by comparing RACMO and IMAU-FDM, then compares MAR and RACMO, but not IMAU-FDM and MAR. Stating that RACMO and IMAU-FDM show similar temporal patterns here is hardly surprising given that one forces the other. The RACMO firn model is also mentioned here but hasn't been detailed before.

We agree that the paragraph was unclear and have modified it for clarity. It is correct that RACMO forces IMAU-FDM, but the fact that both show the same reduced temporal variability of the runoff limits is not due to identical meltwater input but due to their firn modules being very similar, apart from a substantial difference in the number of vertical layers. If meltwater input would control temporal variability of the runoff limit, then RACMO, IMAU-FDM and MAR should all show similar temporal variability because the amount of melt in MAR's accumulation area is similar to RACMO (Fig. 5e and f). We investigate and discuss these aspects now in our revised analysis. We also have added more detail explanations of IMAU-FDM's and its similarities and differences to RACMO's firn module.

Section 5.2- As stated above I don't think this paragraph shows the bucket scheme is the main cause of deviations. Several other differences between IMAU-FDM and MAR are mentioned but without justification as to why they are less important.

Please see our answers to earlier comments. It is potentially impossible to *quantitatively* compare the effects of all differences between the two models. We qualitatively estimate that the way the bucket scheme is implemented is of major importance to the simulated runoff limits. At the same time, we state that the two models have substantial differences unrelated to firn simulation.

Also in reply to comments of reviewer #2 we clarified our argumentation why the implementation of the bucket schemes is a major contributor to the observed differences. We also now provide an extended analysis to investigate the contributions of other differences between the RCMs.

Line 386- Proof that this is an improved method is missing. It might cover more areas but it is not validated.

Indeed, we do not want to claim that the method is improved in the sense that its results are more accurate. But it is improved in the sense of being applicable much more flexibly. As already mentioned, we have added a validation and we also have revised the text to more accurately reflect the nature of the improvement.

Section A2- Please make clearer how this differs from the 2022 paper.

This was unclear, we have revised the text to make the differences clear.

References cited

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

Lefebre, F., Gallée, H., van Ypersele, J.-P., & Greuell, W. (2003). Modeling of snow and ice melt at ETH Camp (West Greenland): A study of surface albedo. *Journal of Geophysical Research: Atmospheres,* **108**(D8). https://doi.org/10.1029/2001jd001160

Machguth, H., Tedstone, A., & Mattea, E. (2023). Daily variations in western Greenland slush limits, 2000 to 2021, mapped from MODIS. *Journal of Glaciology,* **69**(273), 191–203. https://doi.org/10.1017/jog.2022.65

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

Vandecrux, B., Mottram, R., Langen, P. L., Fausto, R. S., Olesen, M., Stevens, C. M., et al. (2020). The firn meltwater Retention Model Intercomparison Project (RetMIP): evaluation of nine firn models at four weather station sites on the Greenland ice sheet. *The Cryosphere,* **14**(11), 3785–3810. https://doi.org/10.5194/tc-14-3785-2020

Reviewer #2

Review of: "Runoff from Greenland's firn area – why do MODIS, RCMs and a firn model disagree?"

By Horst Machguth et al.

Summary

This paper examines and compares meltwater runoff limits in Greenland (i) derived from MODIS imagery and (ii) predicted by two regional climate models (RCMs), RACMO and MAR. The authors find that in general the runoff limits predicted by RACMO are lower than he observations, and those from MAR are higher than observed. The variability in the MAR limits is more closely aligned with the observations, while the RACMO results show comparatively little interannual variability. The authors attribute much of the difference between the models to differences in the meltwater schemes (bucket schemes). The higher runoff limits in MAR leads to higher predicted runoff volume (up to 29% along the KAN-U transect) than in RACMO.

In general, I found this paper to be scientifically sound and well written. It will make a quality contribution to our understanding of meltwater processes on ice sheets (and limitations in modeling them. There are several issues that should be addressed prior to publication, but I expect that most of these should be quite tenable. I have split my comments into "general comments" and "specific and line by line comments"; note that the line-by-line section includes both small comments (e.g., typos) and larger questions I had as I read the paper.

We thank the reviewer for their feedback. We reply to all comments below.

General Comments

1. My primary comment is that the discussion (particularly Section 5.2) needs to be a bit more substantive and based on the results presented. One of the chief takeaways of the paper seems to be that the formulation of the bucket scheme, but this section only sparsely uses the results to support their claim that the bucket scheme is responsible for the differences. For example, is there information that I can glean from figure 5 to support this?

In part this concern comes from the fact that the comparison uses two RCMs, and there is much more in RCMs than the bucket scheme. For example, are RACMO and MAR predicting similar amounts of snowfall and similar winter temperatures to each other, which could change the melt dynamics? The authors also present on albedo differences between the two, but are there other differences in the terms in the energy balance (e.g. calculation of the turbulent fluxes, downwelling longwave, etc.) that could be different between the two? What about how the models handle heat transfer, especially with phase change and associated latent heat?

I don't think that the authors need to do a ton of additional analysis to this end, as I agree that the formulation of the bucket scheme is likely to make a difference. But, I do think it would be appropriate to include a paragraph or two about what other factors might contribute to the differences between the model, and why the bucket scheme formulation is the most important one.

This comment is similar to the overall critique of reviewer #1. We refer to our answer there, please also see our response to the major comment number 4 of reviewer #3. Differences in melt input, caused by differences in the surface energy balance, would obviously trigger differences in meltwater percolation. This was partially considered in the manuscript (mainly Fig. 5 left and right columns) but differences in melt input have now given more attention with a dedicated figure and discussion (Appendix B), showing that both RCMs have similar melt. Furthermore, we now also investigate differences in accumulation as they would also impact simulated firn structure. Again, we find only small differences which cannot explain the strong differences in position of the runoff limit (Appendix B). Albedo is also similar but there are differences in 2012 albedo directly above the RACMO-FDM runoff limit. This sudden change in RACMO-FDM albedo is now investigated and discussed in more detail in the manuscript. We now show how a frequent superposition of ELA and runoff limit in RACMO causes this jump in RACMO albedo. Together, the extended analysis now shows more clearly that the differences in the elevation of the runoff limit are mainly caused by disparity in refreezing, which happens in the subsurface.

2. This could be my ignorance, but throughout the paper the authors seem to use RCM and firn model somewhat interchangeably. I've previously operated thinking that the RCM is an atmospheric model, which is coupled to a subsurface snow/firn model. A bit of language (in section 2.2 perhaps?) clarifying this may be useful.

Agreed, we have revised the text to use a clearer terminology

3. The implications of the paper seem to be mostly limited to a vague paragraph at the end of section 5. Would it be a reasonable amount of work to include the total GrIS runoff from both MAR and RACMO and discuss the uncertainty in runoff with a more detailed discussion of the implications on our understanding of GrIS SMB?

We agree that the importance of runoff in Greenland's firn area for the ice sheet's total mass balance needs to be estimated. However, we would like this study to focus on exploring the reasons behind the models disagreement and quantifying the relevance of firn area runoff to total mass balance is beyond the scope of our study. Nevertheless, we have rewritten most of the discussion and rearranged the text. There is now a dedicated subsection "Implications". We now highlight the study by *Glaude et al.* (2024) in the introduction and discussion as their work shows that there is very substantial uncertainty in simulations of future mass balance of the Greenland firn region. They demonstrate large differences in RCM simulated future melt extent which leads to pronounced uncertainties in future projections of Greenland mass balance.

4. Clarity: mostly the paper is well written and clear, but there are a number of instances (especially in the discussion) that were not written clearly. I note some of these in my specific comments below. My recommendation is to try to avoid writing in passive voice.

We thank the reviewer for pointing out these issues which we have corrected as described below.

5. Model settings: can you provide more detail about the firn model settings that were used? How were the firn columns spun up? Are model settings the same to the extent possible, e.g. surface snow density, etc.? Do both models use similar surface energy balance schemes?

We have added additional detail and references to publications describing the models.

6. The figures are creative and well made, but I generally found that soft colors were hard to see and that text (e.g., legends, axis labels) is too small. (There are also specific comments below that I made while reading the paper.)

We have modified axis fonts and colors, were possible.

Specific and Line by line comments

L31 paragraph: It may be worth mentioning here that the ELA and runoff limit vary from year to year (or calling it a "zone" to encompass that variability?), to differentiate between shifts in ELA and runoff limit that change the long term mean SMB.

Agreed, added.

46: "oppose" – I would choose a different word here. Intercompare? Or just "compare"? "analyze the differences between the runoff limits…"?

Agreed, done.

47: remove period after 2021

Done

Section 2.2 (related to my general comment 2 above): Can you provide slightly more detail/clarification about the differences between RACMO2.3p2 and IMAU-FDM, especially in the context of how you use them in this paper? My previous impression was that IMAU- FDM was coupled to RACMO as the subsurface scheme – is this not the case? Here, is the only reason you are bringing in the online version of IMAU-FDM to evaluate model physics and outputs that are not provided from the coupled model, or are you running IMAU-FDM as well for comparison? (Table 1 clarifies this somewhat, but I think it would be helpful to clarify the text slightly as well.)

We have added more detail, as also mentioned in our reply to point 2 above. Furthermore, the description of the firn models and modules in Section 3.2 has been extended.

83: "agree to" -> "agree with"

Done.

90: I get the gist of what you are doing as described here; however, is it possible to use Figure 1 to help illustrate the polygons?

Showing all polygons directly in Fig. 1 will make the figure unreadable and the polygons would be small. However, we added a supplementary figure (Fig. A1) to illustrate the concept.

Section 3.2.2 – regarding the IMAU bucket scheme – can you provide more specific detail about how it handles ice lenses and slabs, as you do for MAR? (this will help a lot with clarity of discussion section.)

Agreed, we have rewritten this paragraph in order to provide more detail.

151: Please clarify: Is the annual maximum runoff limit the highest elevation where runoff occurs for each year? Given the definition of runoff limit provided in the introduction, why is the maxY!"# not simply where runoff is greater than zero? Also, would it make more sense to consider a maxY!"# threshold as a percentage of the annual snowfall than just a value?

On the ice sheet, the runoff limit is indeed defined as the elevation where runoff starts. However, this definition cannot be directly applied to the RCMs. The RCMs simulation of runoff is fundamentally different from the actual processes and there can be numerical issues where tiny amounts of runoff always occur. In discussion with RCM modellers we decided to use a fixed threshold. Using a threshold that is a function of the amount of annual snowfall would make it very difficult to understand RCM runoff positions. Furthermore, the larger the amount of annual snowfall, the less likely near surface lateral runoff and instead aquifers will form, which are not the topic of this study.

183: can you speculate why the approach does not work well in that terrain?

Yes, we added a brief discussion in the appendix.

Figure 2: The caption and accompanying text needs more detail (text especially) of how to read the various lines. It may help to label the pale lines "IMAU-FDM daily" and "MAR daily". I initially missed the fact that you say where the transects are at the end of the caption. Perhaps include that detail in the first section of the caption, and make the flowlines bolder in figure 1? Figure 1 legend could also include the flowlines. Also, the axis labels on this figure are too small, and the figure would be easier to view if the colors were bolder. I don't quite understand the lat/lon labeled at the tops of the figures, as this is a transect and not a point?

Thank you for this feedback, of which some refers to Fig. 1, if we understand correctly? We have addressed these issues and hope that the new figure is more readable. The coordinates are provided as a rough indication to where the transect is located. This allows the reader to immediately understand whether a transect is in the far north or south. We have mentioned this now and also reduced the coordinate precision to one digit.

200: What difference are you referring to specifically here?

We have removed this sentence.

Figure 3: This is a neat figure. However, the text size throughout and the soft colors make it difficult to read. In panel a, is it possible to darken the flowlines as well?

Done.

Figure 3/Line 200 – It was surprising to see RACMO1k and IMAU-FDM differing so much. I wrote this comment before seeing the text about this later in the paper. I don't think you need to add more in this paper, but it is surprising to see and I hope you will investigate further in the future.

Thank you for the comment.

Figure 4: perhaps include the region along with the panel label at the top to add clarity, i.e. a: NW and b: K-transect

Done.

204: can you quantify this variability with a simple correlation?

Unfortunately, we do not fully understand the reviewer's comment. There are two variabilities mentioned on this line, (1) variability with intensity of the melt season and (2) differences in interannual variability between MODIS and MAR. Both were based on qualitative assessment, which we have now replaced with a quantitative assessment.

208: "shorter in the north than in the south" – is this robust, or just the case for the areas you picked?

This was a qualitative assessment, we agree it would need to be undermined with statistics. As it is not of particular relevance in the context of this study, we have removed the statement.

212: It May be useful in some cases when talking about a specific RCM to use e.g. Y$%& or &%'!"#

Yes, this has been changed

217: "appears more step-wise": Is this just an artifact of the gridding?

We removed this statement as our analysis does not demonstrate this statement quantitatively.

219: This looks like it is only true for MAR?

The challenge is the different temporal resolution of the available data: MAR at 1 day, RACMO-FDM at 12 days. Unfortunately, the latter are unavailable at 1-day resolution. We now comment on this in the text now more clearly and also do this twice, as this is an important difference.

Figure 6: the direction of the y-axis is opposite what I would expect (I would expect time to proceed downward in the direction of reading) – so I recommend switching that or noting this in the caption.

We originally planned to comment on this in the figure caption, however, we decided it appears logical to us that the y-axis shows growing values from bottom to top. While there are different opinions on how to handle dates along figure axes, we believe the figure is sufficiently clear as every subplot has y-axis labels.

Figure 6: It is not clear to me why the 10m firn temperature at the lower elevations much cooler than at the runoff limit (e.g., Figure 6g). Why does the melt cause 10m temperature at runoff limit to increase to near the freezing point, whereas at lower elevations this does not occur? I suspect this is due to there not being firn at all (figure 7), but it would be good to clarify here. (Perhaps label the figure T()# instead of T*+!, 10 m, as there isn't actually firn there?

This is correct, there is no firn at elevations much lower than the runoff limit. Consequently, meltwater runs off without releasing any latent heat in the subsurface. We have relabelled the figure as suggested.

245: Can you explain the odd (discontinuous in distance) 10-m temperatures in MAR? I would expect a (more or less) monotonic change as distance increases.

The snow layer of each pixel is evolving independently of the other ones from the initialisation. Moreover, the snow pack vertical discretisation could be very different following the considered pixel. There are thin snow layers close the surface but thick layers (of several meters) at depth. Therefore, the 10 m temperature shown here is the temperature of the snow layer including the 10 m depth. For some pixels, it could be the temperature of a unique ice layer of 15 m thick or to a 1 m thick snow layer close to 10m depth. To reduce this problem of spatial discontinuity (identified in this paper) in the MAR snow pack, a very light spatial smoothing/filtering is now applied over the 1st layers (the deeper layer) of each pixel in the latest version (3.14.1) of MAR (this study uses 3.14.0). We have commented on this issue in the text.

250: does this mean that in IMAU-FDM, ELA and maxY&'$ are effectively the same thing?

We have now investigated the ELA as simulated by the RCMs.Thereby we noted that for RACMO the ELA coincides with the runoff limit roughly every fourth year. All other years, the ELA is lower than the runoff limit. In MAR, ELA and runoff limit coincide only in 2012. The relatively frequent superposition of ELA and runoff limit in RACMO is relevant to understand the sudden change in albedo at the runoff limit in 2012 and in other years of intense melting where ELA and runoff limit are identical. We now mention and discuss this analysis.

266: "MAR simulates runoff between the two runoff limits" – not exactly clear what you mean by this. I think you mean the additional runoff simulated by MAR above the IMAU-FDM runoff limit and below the MAR limit, but perhaps there is a clearer way of stating this.

Thank you for pointing this out, we have tried to clarify the wording.

273: "the larger the difference in total runoff simulated by the two RCMs" – even with the below paragraph I think is phrased as too strong of a claim, as I don't see anything in this regression based on the total runoff in each. I think the claim needs to be a bit more nuanced, along the lines of "the amount of MAR melt above the IMAU-FDM runoff limit increases exponentially as a function of the MAR melt below the IMAU-FDM runoff limit. Assuming the MAR and IMAU total melt below the IMAU-FDM runoff limit are similar, this implies that the difference in predicted melt between MAR and IMAU increases in high-melt seasons". (or rearrange the section a bit so that the comparison of common area is not at the end.

The total runoff in RACMO-FDM is represented in the regression by MAR runoff below the runoff limit of RACMO-FDM. The RCMs' similarities in "ablation area runoff" are discussed in the following paragraph. We have changed the wording in line with the suggestion of the reviewer.

279: is this maxY-$%.?

Yes, this information was missing, updated.

285: "16 % out of which almost four fifths" (and previous sentence) – these statistics are hard to follow – i.e., 80% of 16% - can you make it a bit clearer for the reader (perhaps adding the actual volumes would do this?)

Adding volumes can lead to more confusion as the analysis focuses on a transect where runoff has, in a strict sense, the unit m^2. We have reworded to clarify the message.

288: "regardless of fundamental differences between runoff processes detected from remote sensing and their simulations." – clarify the sentence structure here.

Done.

306: Aren't bucket schemes instant (all routing occurs within a model time step)? That could be mentioned here instead of "RCM vertical routing is much faster".

Correct, done.

311: It is not clear to me how what is described in this paragraph is lack of inertia – can you elaborate what you mean?

The word "inertia" does not add to clarity. In line with the above comment, we also use "instantaneous" here.

315: "This feedback mechanism, by which ice slabs thicken, is challenging to mimic through a relatively instantaneous bucket scheme." I don't necessarily agree with this – in the model domain, once there is meltwater percolating to the slab all snow/firn above the slab is temperate. If the slab is below freezing, some of the meltwater can refreeze (according to the heat transfer scheme in the model). Why would an instantaneous bucket scheme prevent the model from simulating this process correctly, albeit in a single time step rather than over some timescale? If the slab is being buried in IMAU-FDM and not in reality, a simpler explanation to me is that there is too much snowfall and not enough melt in RACMO at KAN-U. Wouldn't this be consistent with your findings that IMAU-FDM has lower-elevation maxY&'$?

Our statement was misleading or incomplete. The issues lie in the RCMs limiting meltwater storage in firn to irreducible water content. No slush formation is allowed, which by definition is meltwater exceeding the irreducible water content. The ice slabs thicken primarily by accretion of superimposed ice on top of the slabs, which requires slush to persist on top of the slabs for longer time periods. This cannot be simulated by the RCMs at the moment. Too much snowfall could also cause the effect of the slab getting buried, but this is not the problem here. We have modified the text for the message to become clearer. Furthermore, we refer to a recently published study by *Tedstone et al.* (2025) which explains the mechanism of ice slab growth by superimposed ice formation, which cannot be modelled by means of a normal bucket scheme.

316: "In particular, both RCMs do not permit any slush formation and even thick ice layers must remain "permeable" for meltwater to be routed vertically." I am not sure what your point is with this sentence – can you expand a bit to clarify what you are saying?

If the RCMs would treat layers of the density of ice (or any other threshold, such as density of pore close-off) as impermeable, then this would also require to have percolating meltwater ponding on top of such an impermeable layer. The ponding water would completely fill the pore space of the overlying porous layers which would mean that slush is formed in an RCM's firn layer. Both RCMs do not form slush because any water exceeding the irreducible water content is instantaneously routed downwards.

We have reworded the criticised sentence to make the statement clearer in line with the above explanation.

330: "in the absence of pore space, even moderate amounts of melt will run off" – isn't this what would be expected in reality? If there is no pore space available to store the water shouldn't all water run off?

Yes, this behaviour is intended. The question is more whether the pore space, as simulated by RACMO-FDM, corresponds to the spatial distribution of pore space in reality. We changed the wording to make this clearer.

332 – 335: in general this paragraph is hard to follow, in part because (as I noted earlier) the description of the bucket scheme in IMAU-FDM is not fully described. Here it is implied that any meltwater is allowed to pass through ice slabs – is that correct?

Yes, this is correct. As explained above, no water can pond on any ice layer regardless of who thick that layer is. Hence the water passes through ice layers. However, it does this without any interaction with the ice. For any interaction of percolating meltwater with a layer in the firn model, a layer needs to contain pore space and existing water content in the layer needs to be below the maximum irreducible water content. We have added text to clarify this behaviour.

336: "pronounced step change in surface albedo in 2012" Is this shown anywhere? Is this in time or space? Not exactly clear what you mean here.

Yes, this is shown in Figure 5c. We have added a reference to the figure for clarity.

337: Why does higher albedo reduce likelihood of “meltwater percolating to the bottom of the firn where it would run off”? Or are you trying to say that the higher albedo above that step change reduces melt and thereby reduces the volume of meltwater percolating to the bottom of the firn and running off?

Yes, this is what we are trying to say. We changed the wording to make the message clearer.

340: on one hand/on the other – colloquial phrase

We have adapted the wording.

358: “As the primary cause we identify the discrepancies between the two maxYrcm”: this could be rephrased to be clearer, e.g. the primary cause is the fact that MAR consistently predicts a higher runoff limit (and thereby a larger area producing runoff) than IMAU-FDM”

Done

359: why is this surprising? If both predict similar runoff in the ablation zone, wouldn’t the one with higher runoff limit be expected to produce more runoff?

Our message was unclear. The “surprising” refers to the fact that melt is rather small in the vicinity of the runoff limit as compared to the ablation area. Hence, one could expect that differences in modelled ablation area runoff dominate total difference in runoff. As this is unclear, we have reworded the sentence, avoiding the word “surprising”.

393: “strongly later water flux” – can you clarify what this means, or perhaps a typo (layered?)?

Thank you for pointing this out, it should read “strongly lateral”, changed.

Figure A4: this is pretty wild to see this much disagreement – I hope you’ll continue work like the present paper to figure out what is going on with the models here.

Thank you for the comment. Yes, the disagreement is substantial. It is challenging to simulate the interaction of firn and meltwater. We hope our work motivates further research on this topic.

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., Clerx, N., Jullien, N., Picton, H., Ducrey, J., et al. (2025). Concurrent superimposed ice formation and meltwater runoff on Greenland’s ice slabs. *Nature Communications,* 16(4494). https://doi.org/10.1038/s41467-025-59237-9

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. We have included additional analysis and discussion to better narrow down the reasons for the large differences in simulated runoff limits at the K-transect. 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. We now provide a more detailed description of the models' bucket schemes and their runoff conditions.

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. Furthermore, we have conducted additional analysis of the IMAU-FDM percolation depth at the runoff limit and we found that there is not percolation to below the shown 20 m depth. This is now also mentioned in the text.

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 and new Fig. B1) as well as similar amounts of accumulation (new Appendix B). 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 is now discussed in more detail in the manuscript emphasizing its link to the position of the RACMO ELA in years of strong melting.

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 and accumulation in more detail (Appendix B) 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.

Thank you for this comment. We have revised the text to use these two terms consistently. These changes also clarifies the differences in irreducible water saturation between MAR and IMAU-FDM, which contributes to their different runoff limits. We also now refer to their different quantities of irreducible water saturation in the discussion.

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 and accumulation 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. Furthermore, we have corrected the maximal number of firn layers in RACMO which is 100 and not 150.

- 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 the 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