Review #1:

Suggest publication as is.

Reply: We thank the reviewer for their valuable contribution with the earlier revision. This has resulted in substantial improvement to our manuscript.

Review #2: Suggestions for revision or reasons for rejection

(visible to the public if the article is accepted and published)

The authors presented a thoroughly revised manuscript, with most of my concerns taken into account. I appreciate the improved description of how models define runoff, the improved explanations and expanded figure captions.

However, I do find the response to my concern about the huge discrepancy between MAR and FDM firn temperatures a bit underwhelming (page 18 in the rebuttal document). The authors basically provide some hypothesis and argumentation, but it feels like that they have to conclude that they can not explain this. However, it feels like there should be more explanation possible, and such explanations would be key to understanding the results. Can it be validated somehow that it is not an effect of spinup, or lower boundary conditions? Or maybe that there is not a bug in one of the models? I don't know. Maybe if the other 2 reviewers don't mention this, it's not too much of a concern, but I'm not really convinced.

Reply: We agree with the reviewer that we provide hypotheses and argumentation. We cannot exactly point out the reason behind differences in firn temperatures. We added the comparison of melt amounts to rule out that the disagreement is simply due to differences in the amount of simulated melt (it was also requested by a reviewer). It could well be that spin-up plays a role or bugs in the model code (see below). However, this would require series of dedicated model runs, potentially even modifications to the model codes (of IMAU-FDM and/or MAR) to enable an optimal comparison. It would also require an in-depth analysis of the code of both models. Such analysis is outside the scope off our manuscript which aims at pointing out previously unnoticed discrepancies in simulated runoff limits by MAR and IMAU-FDM. Finally, it has been shown in Vandecrux et al., (2024) focusing on the temperature at 10 m below the surface that models disagree substantially, and that there seem to be problems in MAR. It is currently assumed that they are due to numerical instabilities (Vandecrux et al., 2024) but the exact reason is not yet known. We have further emphasized this point in the manuscript.

So the authors write in the rebuttal document: "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."

This implies that MAR routes meltwater down faster than FDM. Is this really the case? I tried to trace back a bit the values for liquid water content, because intuitively 2% saturation (not volumetric content), as mentioned in L120 is very low. Intuitively, I would think it's a volumetric contents. But it is not easy to find in the literature. Can you maybe add a citation to L120? It looks like there is no specific citation for MARv3.14?

Reply: The larger in the firn model the allowed liquid water content in the firn, the faster the warming of the firn, the faster the densification of the snowpack and the conversion from percolation zone to bare ice zone. In Glaude et al. (2024), it has been shown that the climate sensitivity of MAR (and HIRHAM) seems to be higher than RACMO. Therefore, the irreducible water saturation has been reduced in MAR below 1 m knowing that it does not impact the SMB simulated by MAR. 2% remains within the range of acceptable values.

So higher firn temperature could be caused by refreezing of deeper percolation. But first of all, MAR retains more water in the surface layers (7%, L120) than FDM (5.8%, L300). Second, in MAR, the runoff from the middle of the domain over ice layers would remove water available for downward percolation. On top of that, in Alexander et al., 2019, it is written: "The value of the irreducible water saturation in MAR is relatively high (7–10%) compared to the RACMO2.3p2 RCM (Noël et al., 2018; 1%), and could potentially lead to overestimated liquid water retention." But that paper is quite a bit old now, and maybe MAR has been modified since that?

Reply: The values indicated in the present study are correct. Alexander et al. (2019) used a former MAR version (MARv3.12). We have made this clearer in the text.

So, does MAR really route meltwater down faster, such that freezing heats up the deeper firn? If this was the case, then one would expect higher firn densities in MAR, which is not really the case in the regime where MAR is much warmer than FDM (comparing Fig. 7 with Fig. 6). Note that the mentioning from Alexander et al., 2019 of 1% for RACMO in the quote above, is not consistent with the 15% of pore space for 800 kg/m3 mentioned in L147 of the manuscript, which would be 2% volumetric content. Also, the quote from Alexander et al. 2019 is not consistent with L120 in the manuscript. Is this because of later model changes? It is actually a bit difficult to trace back the exact values and definitions, because it is not always clear from published literature, and irreducible water content/saturation probably is a parameter that modellers use to tune their models (and is also often inconsistently defined or confused in meaning). So it's not that I want to blame the authors for the confusion, but that doesn't remove the fact that I'm a bit confused. From the presented information, I simply cannot put the pieces together and understand how MAR and FDM really differ, such that MAR would be so much warmer than FDM.

Reply: As mentioned above, the value in Alexander et al. (2019) are outdated for MAR and wrong for RACMO. If MAR is too warm in depth, it is also due to the problems identified in Vandecrux et al., (2024) which concerns the whole ice sheet and not the difference in irreducible water content. In the dry snow area, for example, MAR is up to 5 °C warmer than the other models while there is no melt in this area. We have further emphasized these issues in the revised manuscript (so far, they were already mentioned on line 423). At this occasion, we have also replaced erroneous citations of the preprint of Vandecrux et al. (2024) with citations of the final published study.

So how much mm of meltwater must be refrozen in MAR to explain the higher temperature? That is an easy calculation when density and temperatures are known. Is that amount consistent with what can be expected from the produced melt, and the faster downward percolation?

I think that the warmer firn temperatures are key to understanding the differences in runoff limits, but currently, I do not manage to understand the connection with how the models treat vertical percolation. Again, I'm not sure about the view of the other reviewers, but maybe the authors can try to substantiate this aspect a bit more. It maybe also depends if the other reviewers have concerns here as well, which would then obviously make it a more substantial problem.

Reply: Figs. 6 and Fig 7 in Vandecrux et al., (2024) convincingly show that the difference of temperature between MAR and FDM is not only linked to the irreducible water content threshold as there are also discrepancies in the bare ice area (where there is no percolation) and in the dry snow zone (where there is no surface melt).

Minor comments

L111, Maybe check the range 0.55 to 0.5, because in Lefebre 2003, it is written: If snow depth is zero, surface albedo varies exponentially between the ice (α = 0.58) and water (α = 0.15)

Reply: We confirm that it is well between 0.5 and 0.55. The values of Lefebre (2003) are outdated. We also clarified this in the text.

L144: Maybe use (i), (ii) and (iii) to separate the three conditions and to make the sentence easier to follow.

Reply: done

Typo caption Fig. 2 "viible"

Reply: done

References

Vandecrux, B., Fausto, R. S., Box, J. E., Covi, F., Hock, R., Rennermalm, Å. K., Heilig, A., Abermann, J., van As, D., Bjerre, E., Fettweis, X., Smeets, P. C. J. P., Kuipers Munneke, P., van den Broeke, M. R., Brils, M., Langen, P. L., Mottram, R., and Ahlstrøm, A. P. (2024). Recent warming trends of the Greenland ice sheet documented by historical firn and ice temperature observations and machine learning, The Cryosphere, 18, 609–631, https://doi.org/10.5194/tc-18-609-2024.

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, e2024GL111902. https://doi.org/10.1029/2024GL111902

Review #2: Suggestions for revision or reasons for rejection

General comments:

I thank the authors for their care in addressing my previous concerns about the paper. While many of those concerns have been adequately addressed, I think the paper needs additional revisions to be ready for publication. The major remaining issues I identify:

1) Introduction of RACMO vs. IMAU-FDM vs. RACMO 1km. In my opinion, the introduction of the RACMO/IMAU products in sections 2.2 and 3.2.2., as well as Table 1, confuse the overall message of the paper. For instance, line 64 implies that the analysis will include runoff limits from IMAU-FDM, RACMO2.3p2, and RACMO 1km. In reality, the analysis focuses on results from IMAU-FDM, with brief mentions of RACMO 1km. To simplify and clarify the message, I suggest paring down these sections and simply state that you are comparing outputs from IMAU-FDM forced with RACMO2.3p2 to MAR. For example, the discussion of differences between RACMO and IMAU-FDM are not actually germane to the message of the paper and thus only serves as a distraction to the reader. For simplicity I would also suggest removing the RACMO 1km results – there are no detailed analyses of why it performs as it does.

Reply: We have completely removed the RACMO 1 km results from the manuscript. However, we do not believe that removing RACMO2.3p2 parameters from the analysis would increase clarity. In the previous rounds of reviews, for example, the question was asked whether the differences between the firn models are due to different amounts of melt or accumulation. Answering this question, for example, requires analysing parameters which are only available from RACMO2.3p2, and not from IMAU-FDM. The same is the case with albedo.

As an example of a lack of clarity in these sections: on line 73: *"Various parameters are unavailable from RACMO2.3p2 and are instead obtained from the offline firn model IMAU- FDM v1.2G henceforth IMAU-FDM. The model is forced in offline mode by RACMO2.3p2 and is run on an identical spatial grid. In the following we refer to 'MAR' for MARv3.14, to 'RACMO' for RACMO2.3p2 at native resolution of 5.5 km and we use 'RACMO 1 km' when we refer to downscaled and bias corrected RACMO2.3p2 data."* But, it turns out in your analyses that RACMO2.3p2 results are not actually used, correct? And on line 80, *"whose output is not available at a sufficient level of detail for the present study."* Again, what does this mean? The above text had indicated that RACMO2.3p2 was part of the study, and here there it seems contradictory that it is not being used.

Reply: As mentioned above, this is incorrect as we use parameters from RACMO2.3p2. The RACMO2.3p2 simulated parameters (α, C and M) are used widely in the sections results and discussion as well as in the appendix. As mentioned above, we have removed the 1 km RACMO data, and we hope that this will further increase clarity.

We agree that the statement on RACMO2.3p2 output having insufficient detail was unclear. We have removed the statement on "insufficient level of detail for this study" because this is now better explained on line 73 (see our answer below).

Also, what does it mean (line 73) that *"parameters are unavailable"*?

Reply: This means that these parameters are not written to output by RACMO2.3p2 and are unavailable for that reason. We have clarified the statement accordingly.

2) The issue of the firn temperature difference between RACMO and MAR is inadequately examined. As a central focus of the paper is on the difference in runoff between the models, the temperature of the modeled firn is a very important contributor to the amount of refreezing that occurs, and therefore warrants thorough examination. Figure 6 shows a substantial difference in firn temperature (RACMO much colder) which would indicate that RACMO has substantially more cold content available to refreeze meltwater.

Reply: The differences in firn temperatures are now discussed in more detail, further highlighting the results by Vandecrux et al. (2024). However, the exact cause of the warmer MAR firn temperatures has not yet been identified (see our replies further below) and can thus not be described in the present study.

Line 398 says, *"An alternative explanation for the colder IMAU- FDM firn temperatures would be that the Figures 6, C3 and C5 give a wrong impression because latent heat in IMAU-FDM is released at depths greater than the max. 20 m shown in the figures."* This does not make sense to me – doesn't percolating meltwater in the bucket scheme first warm the firn in a given layer, refreezing as much water as needed to bring the temperature to 0, before percolating to the next layer below? So yes, latent heat can be released at deeper temperatures, but it does not bypass releasing latent heat in shallower firn as it percolates to those greater depths.

Reply: In a bucket scheme meltwater can bypass layers if there is insufficient pore space. This leads to the situation where water can bypass thick and even cold ice layers without any interaction. This is a general limitation of bucket schemes.

Likewise, the abstract and conclusion state that the implementation of the bucket scheme is the cause of the disparity in modeled runoff limits, but the results and discussion section feature no discussion of the bucket scheme – if this is indeed a main conclusion of the paper, I would expect the analysis and discussion to make the scientific argument that this is the case. As it is, it comes across as rather speculative. I do think that it is a bit more nuanced than just the bucket scheme, as temperature seems like it should play a role? Or, is heat transfer considered part of the bucket scheme? In that case, a more granular conclusion (process based vs. simply using bucket scheme as a catch-all) would be appropriate. The conclusion gets a bit at that granularity (470-474), but categorizes those under "bucket scheme", while I would contend that "(iv) the firn layer in MAR is warmer" is not necessarily a bucket scheme issue.

Reply: We agree that the bucket scheme is not the sole reason behind the disparities. It is also correct that the term "bucket scheme" did not appear in the discussion. The latter, however, referred in detail to various key parameters of bucket schemes such as (i) the choice of irreducible water content, (ii) the chosen depth of the firn layer, (iii) the choice whether water is considered runoff only at the bottom of the firn pack or also somewhere halfway.

We have now changed the wording in the discussion to state clearly which parameters represent "design choices" in the bucket schemes. Furthermore, we have reworded the abstract and the conclusions to avoid the impression that the bucket scheme is the sole reason for (i) the disparity in runoff limits and (ii) the warmer firn.

3) Qualitative results: In sections 4.2.1 and 4.2.2, the results are entirely qualitative. While qualitative results are not necessarily bad, they should be accompanied by quantitative results. Phrases like "very similar" and "generally good" do not provide adequate description in a results section.

Reply: The two sections are deliberately written in a qualitative style because these two sections mainly describe the figures which quantify the differences. We have added mean values as well as standard deviations to the section for improved quantification.

4) Clarity issues. I have tried to note some of these below in line-by-line comments. There are numerous places that meanings are obscured by overly verbose explanations. This may sound contradictory, but there are also numerous instances of statements being made without explanation. But, in both cases the paper loses clarity. I suggest a thorough read- through to add clarity to descriptions and to remove statements that are not germane to the topic of a particular paragraph. This is not an issue that should prevent publication but would improve the paper greatly.

Reply: Thank you for these suggestions on which we comment below. In addition, we have once again streamlined the text and removed unnecessary phrases.

Line by line comments:

Line 72/Table 1: There seems to be a discrepancy between this list and what is analyzed in the paper. For example, the list includes fac_10m and lwc_1m, but as far as I can see those are not actually included anywhere in the analyses.

Reply: These parameters were part of the analysis, but their analysis has not yielded added insights. We have removed them from the list.

84: Why are you considering this “relatively coarse” when it is significantly finer than you RCM/FDM grid?

Reply: We have removed the words “relatively coarse”. The next sentence states that the satellite sensor resolution is coarse compared to e.g. Landsat.

109: “(the maximum density of pure snow)” – this claim seems unfounded. Does snow need to have impurities to have higher density? Or are you suggesting that at that density it is no longer snow? In either case this seems to be a dubious claim; for example, snow at the base of an Alaskan snowpack will regularly exceed that value.

Reply: This statement was unclear and has been replaced by “Where surface density exceeds 450 kg m^-3, the minimum value of albedo declines between the minimum snow albedo (0.7) and clean ice albedo (0.55) as a linear function of increasing density.” We have removed the word “pure” which was confusing.

114: I realize that this is not the appropriate venue for a complete description of the changs in MAR3.14, but if you are going to list them it would be appropriate to list how those changes are germane to the present work, e.g. how do the bug fixes in the cloud scheme and the rain/snow partitioning changes affect the runoff limit and/or meltwater production?

Reply: The recent improvements in MARv3.14 do not impact the present work. The problem in firn temperatures found in Vandecrux et al., (2024) using MARv3.12 is still present in MARv3.14.1 used here. To artificiality reduce the problem illustrated in Fig. 7 of Vandecrux et al., (2024), the new version of MAR in development (MARv3.14.3), applies a weak spatial smoothing between pixels of the temperature of the 1st snow layer at 20 m below the surface.

195: "search for parameters that show peculiar or unexpected values in the broader elevation range": related to clarity comment above – what constitutes peculiar or unexpected? How do you search for parameters?

Reply: This was mainly done visually, by first plotting the parameters as shown in Figs. 5, 6, 7, C3 and C4. We have reworded the sentence for improved clarity.

218: Specify which Appendix A (1-4)

Reply: done

Figure 2: I noted in my previous review as well – in Figure 2 the names of the transects that the panels are illustrating are buried in the caption. It would be much easier if the panels had the transect name printed on the panel itself – e.g. next to or in place of the coordinates listed. If you are keeping the coordinates, than I recommending adding the degree symbol and direction, i.e. 79.2oN, -26.8 oE. I realize the present formatting may be obvious to some readers but I suspect that is not the case for all.

Reply: We have implemented the suggested modifications.

238: Consider adding note in text stating the SD for RACMO 1k is not shown; I read this paragraph and went to figure 3 specifically to look at 1km SD.

Reply: 1 km RACMO has been removed from the manuscript.

309: "this implies that the difference in simulated runoff between MAR and IMAU-FDM increases in high-melt seasons" – can't you calculate this directly by summing that total runoff from MAR and IMAU-FDM for each year, and comparing the difference to the total runoff? That would make a much stronger argument here, especially for extending the argument beyond the K-transect.

Reply: We do not fully understand the comment. This is shown and quantified in Fig. 8 already? Or does the reviewer ask to do such an analysis Greenland-wide? We do not think it would add to clarity, rather to confusion as the manuscript explains the differences with a focus on a clearly defined study area (the K-transect).

Paragraph 315-320: I've read and reread this paragraph several times, and the reality is that the second half of it is not clearly written. What is the core takeaway meant to be? I think that it is that in high-melt years the additional area of the runoff zone in MAR creates a disproportionate increase in total runoff relative to RACMO (i.e., MAR usually has more runoff than RACMO, but in high-melt years it has a lot more runoff due to the increased size of the runoff area). I don't dispute the claim, but please work on the language to make your point clearer. I flagged this issue in my previous review, but a look at the changes document shows no meaningful change. Describing a fraction of a percentage as done presently obscures the meaning here. I disagree with the assertion in the response that "Adding volumes can lead to more confusion as the analysis focuses on a transect where runoff has, in a strict sense, the unit m2". Why is it confusing to add actual values that readers can reference, even if the units are a bit odd? Regarding the unit issue – you could simply assume a unit width of the transect. Or, you could change to something like, "In 2012, total MAR runoff along the K-transect exceeds IMAU-FDM by 29 %. However, the partitioning of that difference is disproportional: in the common runoff area, MAR runoff exceeds IMAU- FDM runoff by X%, meaning that most (75%) of the additional runoff in MAR is generated in the zone above maxY$%&'()*%.

Reply: We now assume unit width and state this in the text. The units are now m^3. Furthermore, we also use m^3 in the text and we hope that the previously confusing percentage values are now clearer. Furthermore, we have changed some of the wording in the paragraph.

Paragraph 315-320: Introducing 2019 as an example here is a bit confusing. I get that it serves as another example, but the results section mostly focuses on contrasting 2012 and 2017. I fear that a reader who is reading quickly will just assume this is another 2012/2017 comparison. If you want to keep the 2019 information, I suggest being more explicit that you are pivoting and discussing another high-melt year, e.g., first discuss 2012, then say something like, "to examine the robustness of this finding, we also examined the runoff in 2019, which was another above average melt year. Consistent with the 2012 results, for 2019 MAR predicts total runoff that is X%".

Reply: Agreed, done.

416: It may be appropriate to cite Van As, 2017 here:

van As, D., Bech Mikkelsen, A., Holtegaard Nielsen, M., Box, J. E., Claesson Liljedahl, L., Lindbäck, K., Pitcher, L., and Hasholt, B.: Hypsometric amplification and routing moderation of Greenland ice sheet meltwater release, The Cryosphere, 11, 1371–1386, https://doi.org/10.5194/tc-11-1371-2017, 2017.

Reply: Done.

482: "This means the situation where the two models diverge the most will become more frequent, simulated runoff will further diverge and uncertainty grow." I don't think this statement is entirely backed up by the analyses done in the paper – this is a claim about the transient response of the models to warming, but the manuscript does not include a rigorous analysis of the temporal trends in the modelled maximum runoff elevation (rather, it provides snapshots in time). While the statement is likely true, the paper does not provide evidence that the simulated runoff will further diverge. I suggest a simple change to acknowledge the speculative nature, such as, "This means the situation where we observe the two models diverge the most will become more frequent. We hypothesize that as a result, simulated runoff will further diverge and uncertainty will grow."

Reply: Done.

Data availability: I recognize that the RCM data are too big for a Zenodo repository, but there are other options for making data available. From the Copernicus Publications Data Policy: "The best way to provide access to data is by depositing them (as well as related metadata) in FAIR-aligned reliable public data repositories, assigning digital object identifiers, and properly citing data sets as individual contributions" and "In rare cases where the data cannot be deposited publicly (e.g., because of commercial constraints), a detailed explanation of why this is the case is required." In the modern era of open science, "data can be obtained directly from the authors" does not meet the standard of publicly accessible, nor does this statement comprise a detailed explanation.

Reply: For the final version of the manuscript, we will filter the RACMO data to only include the parameters we have used. This will reduce the amount of data sufficiently so that they can be uploaded to Zenodo.

Appendix B, Line 587: "in IMAU-FDM, the runoff limit is typically located where summer melt exceeds annual accumulation ($C_{RACMO} - M_{RACMO} = -0.19 \pm 0.25$ m w.e.); in MAR melt and accumulation at $maxY_{MAR}$ are similar ($C - M = 0.03 \pm 0.14$ m w.e.)."

I don't get this – isn't this saying that the runoff limit is below the ELA in RACMO, and at the ELA in MAR? If accumulation is less than melt, you are in the ablation zone by definition?

Reply: Unfortunately, we do not fully understand the question. It is correct that the ELA must be located below the runoff limit. The ELA is defined as the elevation where the climatic mass balance is zero (Cogley et al., 2011). We now mention this definition much earlier in the manuscript. When there is substantial refreezing in the near surface, as it is the case in MAR, RACMO (and in reality near ELA and runoff limit of the Greenland ice sheet, see *Tedstone et al.*, 2025), then the ELA is located at altitudes where melt exceeds accumulation.

References

Cogley, J. G., Hock, R., Rasmussen, L. A., Arendt, A. A., Bauder, A., Braithwaite, R. J., et al. Glossary of glacier mass balance and related terms, Pub. L. No. 2 (2011). Paris: IHP-VII Technical Documents in Hydrology. Retrieved from http://unesdoc.unesco.org/images/0019/001925/192525e.pdf

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