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).

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 main 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.

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.

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 revised the text to more clearly argue that the percolation schemes are the main reasons for the observed differences. At the same time we also now provide a more detailed analysis of the influence of differences in melt on simulated runoff limits (see also our answers to reviewer #2).

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 prefer to 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. Nevertheless, differences in simulated melt also have an influence which we show now in an extended analysis.

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.

Agreed, will be done

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 hope this question is now clarified by the direct comparison of the two approaches.

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.

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 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