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

By Horst Machguth et al.

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.

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.

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

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.

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.

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.

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.

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.

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.

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

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.

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?

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?

218: Specify which Appendix A (1-4)

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. 79.2°N, -26.8°E. I realize the present formatting may be obvious to some readers but I suspect that is not the case for all.

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.

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.

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 max$Y_{rcm}^{IMAU-FDM}$.

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%".

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.

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."

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.

Appendix B, Line 587: “in IMAU-FDM, the runoff limit is typically located where summer melt exceeds annual accumulation (CRACMO − MRACMO = −0.19 ± 0.25 m w.e.); in MAR melt and accumulation at maxYMAR are similar (C − M = 0.03 ± 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?