Reply to Anonymous Referee #1 comments on

"Bias in modeled Greenland ice sheet melt revealed by ASCAT"

by Anna Puggaard, Nicolaj Hansen, Ruth Mottram, Thomas Nagler, Stefan Scheiblauer, Sebastian B. Simonsen, Louise S. Sørensen, Jan Wuite, and Anne M. Solgaard

Dear Editor and Anonymous Referee #1,

We would like to thank you for your insightful comments on our manuscript. In the following, we try to follow and implement your suggestions to the best of our ability. Below is a point-by-point response. To ease following the reply, we have your comments in **black** and our responses highlighted in **Blue**, and we suggest changes to the manuscript in **Red.**

 # General comments:

Two major comments remain not addressed properly. First, the confusion between the detection of water at the surface or in the snowpack and the actual melting process is still pervasive throughout the manuscript. ASCAT does not observe melt directly; it observes the presence of water, regardless of whether this water originates from melting, rain, or melt ponds, for instance. This is fundamentally different from the melting process itself. Throughout the manuscript, you seem to use the shorthand: liquid water at the surface = melt. However, this equivalence is overly simplistic and could confuse readers who are not fully aware of the differences or of what the sensor is actually measuring. This confusion is further reinforced by the use of terms such as "melt maps" to describe the ASCAT product (see specific comment).

We agree that there's a difference between the detection of liquid water and melt and further, that ASCAT only indirectly detects melt water. However, we apply a more sophisticated detection/classification method for liquid water than traditional methods (simple thresholding), and thus, we can distinguish between active melting (decrease and saturation of the backscatter signal) and refreezing (increase of the backscatter signal). See Figure 4 for an illustration of the backscatter signal and different classification classes. This is described in section 2.3 (L.171-173). Thus, we can more accurately than previously assume liquid water at the surface = melt. We have added a paragraph to the introduction detailing the "ASCAT melt maps" compared to other products; see the response to the second specific comment.

In order to clear up some of the confusion for the reader, we suggest changing some of the naming of the concept related to ASCAT detection and extent. We suggest making the following changes to mainly section 3.2 and the discussion:

- Melt classification -> liquid water classification in section (ASCAT melt maps)
- Melt extent -> liquid water extent in section 2.3 (ASCAT melt maps) and the discussion.
- We will also change the title of section 5.2 (L. 361) from Limitations of ASCAT melt observations to Limitations of ASCAT liquid water observations.

- Melt detection -> liquid water detection in section 5.2 (Limitations of ASCAT liquid water observations.

The use of the name “ASCAT melt maps” could be considered misleading. However, the goal of this MS is not to present the data product itself (This is done by Nagler et al., 2024) but rather to present the opportunities for using the “ASCAT melt maps” as validation of RCM melt output. We use “ASCAT melt maps” since it’s the name used in other literature (Nagler et al., 2024 & Boxall et al, 2024). We choose to retain this terminology to avoid any inconsistencies.

This brings us to the second point. I suggest that you also compare the extent of liquid water content (LWC) from the RCMs using the same method you applied to evaluate melt. The LWC should be as accessible as the melt variable in the outputs of the RCMs. Since liquid water first appears in the upper layer(s) of the snowpack before being accounted for in the melt variable, it is possible that LWC corresponds more closely to the ASCAT observations. Given that ASCAT detects the presence of liquid water, I would not be surprised if the snowpack's LWC provides a better comparison than melt, which appears later. This approach could also help explain the delays you have observed. A potential issue with LWC is that the snowpack layers are not equivalent to the three RCMs. However, since you are not examining the volume of liquid water, I believe there is still a feasible way to address this. This addition could significantly enhance your study by demonstrating whether it is more accurate to compare melt or liquid water content from an RCM when compared to ASCAT or other observational products to assess RCM performance.

We agree that it would be interesting to compare melt and liquid water content from the RCMs. However, we see several issues with the proposal.

Firstly, the LWC field is not saved consistently in time and space across all models to save disk space. Therefore, if we want to make the evaluation with the LWC, we will have to rerun several of the models to make such a comparison. Rerunning the RCMs to output LWC consistently through time and space is beyond the scope of this paper.

Secondly, as the reviewer mentioned, the RCMs do not have consistent layer schemes between models and in time. Thus, we cannot make a fair evaluation between models and against ASCAT. Moreover, the penetration depth of the ASCAT (radar) signal varies in time, but when liquid water is present on the surface, the signal cannot penetrate into the snowpack. Thus, it would not be a fair comparison between the models and ASCAT during the melt season, when ASCAT has no or little penetration into the snowpack.

Thirdly, since we are only evaluating the RCMs against ASCAT above the snowline, rain does not have a large effect compared to melt. Therefore, this cannot explain why we consistently see a delay in melt/liquid water compared to ASCAT.

Finally, the ASCAT product accounts for remaining liquid water in the snowpack from previous seasons by an annual recalibration of the winter signal pixel-by-pixel (as we write in L. 159). We propose to add the following sentence to the MS to make this statement more clear to the reader:

L170-171: “To account for the possibility of remnant changes in the snowpack from the previous melt season, the winter signal is recalibrated annually pixel-by-pixel.”

 # Specific comments:

Intro:

- L36: "However, these models are influenced by the chosen modeling approach, and substantial disparities persist among models (Rae et al., 2012; Vernon et al., 2013; Fettweis et al., 2020)." You should add now Glaude et al (2024).
  Done. The sentence now reads:
  L36: "However, these models are influenced by the chosen modeling approach, and substantial disparities persist among models (Rae et al., 2012; Vernon et al., 2013; Fettweis et al., 2020; Glaude et al., 2024)."

- If you want users of RCMs to refer to ASCAT as an independent product, it needs to be compared to other remotely sensed products that estimate the presence of meltwater. You already mention various studies and products in your introduction. A more detailed comparison of these different products with ASCAT is necessary here, even if they are not entirely independent. To use ASCAT as a reference, it is important to position it in relation to other similar products, highlighting its strengths and weaknesses.
  We agree that it's important to position the ASCAT melt maps to other products. Thus, we have added the following paragraph:
  L73-79: "Instead of using a simple threshold method, ASCAT melt maps utilize an algorithm that incorporates the temporal behavior of the backscattered signal. With this method, the ASCAT melt maps can not only detect the presence of liquid water on the surface but also distinguish between melting and subsequent refreezing of the surface meltwater. This makes ASCAT melt maps a unique product as they allow for a more fair comparison between the observed liquid melt water extent and the surface melt extent simulated by RCMs. Furthermore, by applying an annual recalibration of the winter signal, the product accounts for the formation of subsurface features from the previous melt season (Nagler et al., 2024). Again, this ensures a better classification of melt signals compared to previous products (Ashcraft and Long, 2006; Husman et al., 2023; Nagler et al., 2024)."

- As explained in the second major comment, I think than "ASCAT melt map" could be misleading. Even if you already improve the phrasing in the text by changing "ASCAT observes melt" into "ASCAT observes the presence of liquid water" , the term "melt map" is still being used. Why not use "wet-snow mask/map"? This distinction is well-established in the literature and has been discussed by some of the authors of the present manuscript.
  See the response to the first general comment.

Data:

- As you mention in the caption of the first Figure, could you also add in the text that you use AWSs included in the ASCAT melt detection domain? Also, I assume that the ASCAT melt detection domain is not the same than the mask of the maximum elevation of the snowline between 2007-2020 used to mask out part of the domain for the comparison. Could you specify whether all AWSs used to determine the melt rate threshold are within the comparison mask? I was not entirely convince by your explanation regarding the inclusion of AWS in the ASCAT mask (the maximum elevation of the snowline one). It's true that you want to evaluate RCMs' performance against observation over the entire ice sheet, so it's justified to consider all AWSs in this comparison. But you use all the AWSs as well (including the ones outside the ASCAT mask) to determine melting thresholds, which are considered to determine the starting point of your melt event in the RCMs. I understand that it's challenging to have observations that cover the entire domain and represent each of the different area of the ice sheet adequately and equivalently, but this limitation should be explicitly discussed, especially if a significant number of AWSs are excluded. You should also verify whether including only the AWSs within the mask affects the determination of the thresholds. If it does have an impact, this should be mentioned. If it does not change the thresholds (and therefore likely does not alter the conclusions of your comparisons) you could simply state that you checked this and that it does not influence your results.

You are correct that the ASCAT melt detection domain is not the same as the snowline domain. By using the ASCAT melt detection domain, we simply remove stations that are not located on the ice sheet, i.e. stations located on peripheral glaciers and tundra. We further only included temperature observations from 2007-2020 and excluded any stations that did not have any temperature observations within this period. Thus, 34 stations were included in the study even though PROMICE-GCnet included 54 stations. In the data section of the MS we have made the following edit:

L99-100: "Here, we include the 34 stations on the Greenland Ice Sheet and have measurements of air temperature between 2007 and 2020."

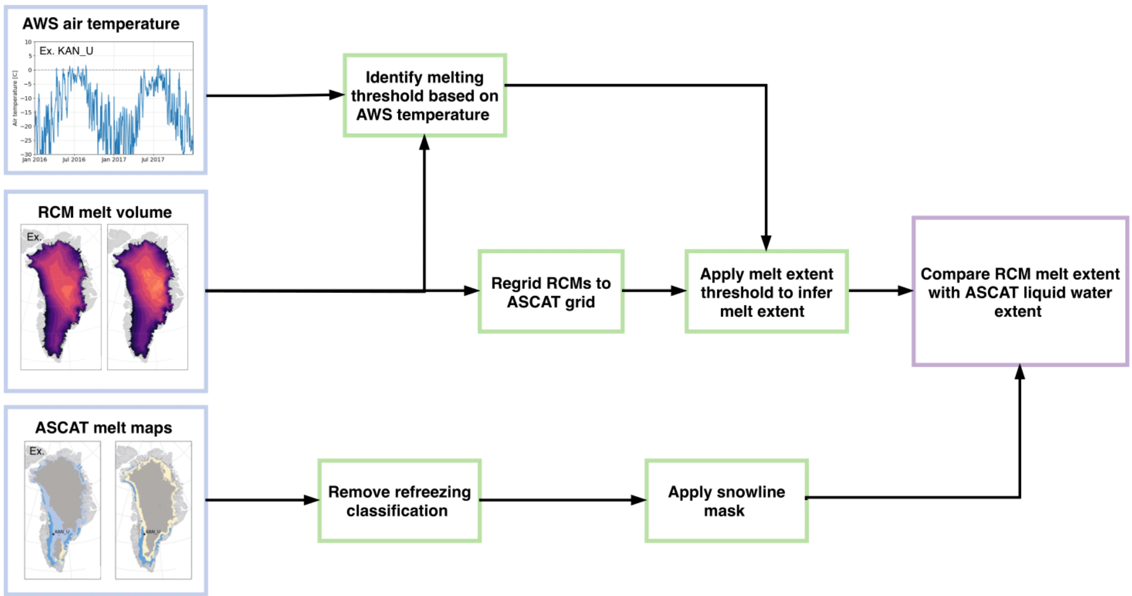
The PROMICE-GCnet AWSs are not equally distributed over the accumulation and ablation zones. Thus, only looking at the AWS inside the snowline mask, we only include 34 stations, . Moreover, as we describe in L215-218 in the method section, the dataset exhibits an imbalance since there are, in general, fewer melt days than days without melt. Excluding stations below the maximum snowline extent will further amplify this data imbalance since melt occurs very infrequently at the majority of these stations. We refer to the temperature observations at KAN_U in the newly added diagram in the method section to showcase the low number of days above 0^{o}C. Compared to the other stations above the snowline mask, KAN_U experiences more days at t>= 0^{o}C due to the relatively low elevation as well as the southern location. This will, of course, affect the identification of a melt threshold. However, the difference will not only be because of a different physical signal but mostly due to a change in data distribution and imbalance.

Furthermore, as pointed out in line L198-200, this method of aligning the RCMs with AWS only ensures that the modeled melt aligns at the specific stations due to local weather conditions. Thus, we are interested in including a larger number of AWS in this alignment to maximize the location that RCMs realistically model melt – even if it is outside the domain that we compare with ASCAT. To make this point, we have made the following edit to the MS L198-202: "It is important to note that the AWS measurements are significantly affected by local scale weather conditions, so this approach only ensures that the modeled melt by the RCMs aligns primarily at these specific locations. Therefore, we include the maximum number of AWS to align with, rather than being limited to those above the snowline's maximum extent, which is where ASCAT and the inferred RCM melt extents are compared."

Method:

- Your method section is still not entirely clear to me. I suggest including a sketch or diagram to illustrate each step of your methodology, specifying what data are used at each step, and so on. This would be particularly useful if you want your method to be replicable for future studies. For example, it was unclear to me how you link the temperature at AWSs with the melt threshold. It seems that a few connections between the different parts of your method are missing. A diagram summarizing your methodology would greatly enhance clarity.

  We have added this diagram (now Figure 5) at the beginning of the method section to make the evaluation prior to comparing ASCAT and RCMs more clear:



  Furthermore, we have edited the first paragraph of the method section based on this comment and your previous comment about AWS data:

  L191-204: "Modeled surface melt in RCMs is subject to large variability among models as seen in Fig. 2 and discussed in e.g Fettweis et al. (2020). The Greenland Ice Sheet SMB model intercomparison project (GrSMBMIP) suggested that discrepancies between RCMs are not systematic (Fettweis et al., 2020), thus there is a need for individual evaluation of each modeled melt volume product before we can compare the extent observed by ASCAT. We refer to Figure 5 for an overview of the evaluation taken prior to comparing RCMs with ASCAT. To establish a threshold (in mm w.e. day−1) to infer the melt extent from the simulated RCM melt volume, we compare it to the PROMICE GC-net AWS. With this comparison we identify how much meltwater must be in the models before we can also observe it at the AWS. We compare each AWS to the RCM grid cell, which has the closest center point to the AWS location; see Figure 1 for AWS locations. It is important to note that the AWS measurements are significantly affected by local scale weather conditions, so this approach only ensures that the modeled melt by the RCMs aligns primarily at these specific locations. Since melt is not directly measured at the AWS stations, we use 2m air temperature as a proxy for melt conditions, as near-surface air temperature is closely linked to melt processes. This approach allows us to identify and quantify temperature biases in each of the RCMs and assess how well the models simulate melt compared to in situ observations."

- Why did you use different methods to interpolate the 3 model results onto the ASCAT grid? For consistency, it would be preferable to use the same method for all three models. This choice should be justified in the text. Line 246. As you mentioned it in the text, “different method of interpolation could bring again more uncertainties” , you state that you account for this bias in your results. Could you explain more explicitly how you address the resulting bias?

  For RACMO2.3p (1km) and HIRHAM5 (5km) we upscale data to get on the common grid of ASCAT (5.5km), whereas for MAR3.12 (10km) we downscale data to get on the common grid. Therefore, we use a different interpolation method. We have added this to the methods section:

  L244-245: “To ensure consistency across datasets, all RCMs are regridded to a common grid, in this case, the ASCAT grid of 5.5 km resolution. For HIRHAM5 and RACMO2.3p2 we upscale data and apply the nearest neighbor interpolation method. For MARv3.12 downscale data and use a cubic interpolation method”

  Concerning the introduction of systematic biases. This was mostly a concern for choosing the baseline threshold, where we want to apply a threshold to all RCMs to assess the simulated melt in the RCMs without accounting for the warm/cold biases. In theory, a melt day in the RCMs should be defined as > 0 mm w.e./day, so this baseline threshold should be the smallest possible. However, we saw that the regridding introduced a systematic positive melt volume to the model output. Thus, the aim is to choose the lowest threshold, which accounts for the possibility of small biases in the models. This ended being a threshold of > 0.1 mm w.e./day.

  To make this more clear, we suggest adding:

  L250-252: “The potential implication of the regridding biases is considered when choosing a baseline threshold. Here, the aim is to apply a baseline threshold to all RCMs independent of warm/cold bias within the RCMs. Therefore, the baseline threshold was set to the smallest value possible without allowing regridding biases to impact the number of melt days.”

Results:

- Table 1 caption: “The corresponding mean air temperature for July and August are also showcased for each RCM. Figure 3a-d illustrates the mean JJA air temperature for each RCM. ”You need to clarify in the caption which numbers are for August and for July?

  We compute the mean of July and August Air temperature and not a mean for July and August separately. To clear up this confusion, we suggest changing the Table caption to the following:

  Table 1: ”Melting thresholds for the different RCMs based on in situ PROMICE AWS observations of air temperature. The table also gives examples of the mean July and August air temperatures at six selected AWS and the mean across all stations. The locations of all stations included in the study are shown in Figure 1. The corresponding mean July and August air temperature are also showcased for each RCM.”

- L245-246. I assume the melt days are considered into a common mask. The

ASCAT one? Could you precise? And could you also precise if it is considered before or after interpolation on the common grid? By the way, these details seem more appropriate for the methods section than the results section, so I suggest moving them to the methods section to complete it.

We agree that this information is not explicitly given in the current MS and that this information belongs in the method section. Thus, we suggest adding the following to the MS in the method section:

L251: "When comparing the RCM melt extent and ASCAT liquid water extent, we apply a similar snowline mask to the RCMs as ASCAT.

Discussion

- Paragraph 2 of ASCAT limitations: If you use LWC to complete your study, you'll also be able to better explain RCMs detection delay compared to ASCAT.
  See reply to second general comments

Typs

L176: "a effect" -> an effects

Done.

L229: "the PR-curve the suggest": remove the extra "the".

Done.

Figure 3: subplot indexes (a-p) are missing on the subplots.

Done.

References:

Boxall, K., Christie, F. D. W., Willis, I. C., Wuite, J., Nagler, T., & Scheiblauer, S. (2024). Drivers of seasonal land-ice-flow variability in the Antarctic Peninsula. Journal of Geophysical Research: Earth Surface, 129, e2023JF007378. https://doi.org/10.1029/2023JF007378s