Reply to Anonymous Referee #2 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 Anonymous Referee #2,

We first and foremost 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, and we sincerely believe that your review/comments have improved the manuscript. 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 suggest changes to the manuscript in **Red.** Moreover, line numbers in our replies to comment refer to the updated MS.

General remarks:

1. Essential pieces of information are missing in the method part (in some cases these are implicitly given in the text): which period is taken into account, what frequency is analysed (hourly, daily, other?), how is different spatial and temporal resolution in the different data sets treated? How are datasets regridded? There are regridding biases mentioned- these could be illustrated or estimated.

We agree that this information is either missing or unclear. Thus, we have revised the beginning of the method section (L184): "To evaluate temperature biases in the RCMs, we compare the modeled output with daily observations from 2007 to 2020 by the PROMICE GC-net AWS shown in Figure 1."

To clarify the process of regridding we have added to the end of the method section (L230-234):

"To ensure consistency across datasets, all RCMs are regridded to a common grid, in this case the ASCAT grid. For HIRHAM5 and RACMO2.3p2, we apply the nearest neighbor interpolation method, while for MARv3.12 we use a cubic interpolation method. It is important to note that regridding can potentially introduce a bias within the RCM output, meaning systematic errors not associated with internal parameterization choices within the RCMs. These potential regridding biases are taken into account when comparing to the ASCAT melt extent"

2. In my view the introduction could point out more clearly the potential benefit of the ASCAT data set: Surface mass balance estimates have been mostly evaluated and also tuned with respect to the (basin wide) mass balance of the GrIS. Potential biases in melt rates above the snow line might be overlooked like that, as these are not necessarily resulting in mass changes. However these regions, where melt occurs only sporadically today, might turn into regions which contribute to sea-level rise in the near future.

This is very true so thank you for your insights. We have included the following in L72-75: "By using ASCAT melt maps we aim to establish a framework for evaluating the performance of RCMs in simulating the temporal variability of present-day melt extent. As RCMs are often calibrated with respect to basin-wide surface mass balance, incorporating an independent satellite dataset like ASCAT melt maps enables a more comprehensive assessment of model performance."

3. It seems overambitious to try to investigate why different RCMs represent melt rates differently without a much deeper dive into characteristics of snow properties and climate forcing. On the other hand, to show that the different models simulate different melt extents does not require a satellite data set.

The aim of our paper is not to explore why RCMs simulate melt rates differently. Instead, we want to show how we can use satellite-derived melt extent (e.g., ASCAT) to evaluate the performance of RCMs in simulating melt extent. This approach allows us to assess how well the models reflect the present-day melt extent, providing a critical observational benchmark. As you mentioned, RCMs are often tuned with respect to basin-wide or point-wise surface mass balance. By incorporating an independent satellite dataset, such as ASCAT, we can make a more comprehensive assessment of model performance. This provides a broader and more comprehensive assessment of how accurately the models simulate melt patterns across the ice sheet.

Based on your comment and the 3rd general comment of reviewer #1, we suggest an addition to both the introduction and conclusion, see reply to reviewer #1, to clarify the scope of MS

Instead I would propose to emphasize and focus on questions like:

Is the onset of melt detected too late systematically? How many days?

Are there differences between regions which experience surface melt every year and regions where melt occurs only in extreme melt years like 2012.

Is the length of melt periods overestimated by ASCAT due to the residual meltwater in the snow pack? If so, is this bias particularly pronounced for long, intense or short periods? Where melt periods occur intermittently within one year, are later melt periods represented differently than early melt periods (idea: the albedo might not recover fully after a melt event and the snow surface might be more vulnerable- which is potentially not represented in the RCMs)

These aspects are already present in the manuscript here and there but not really given full attention.

Yes, these are interesting features to discuss. Table 2 and the Results section address many of these questions on a basin scale and a discussion of the observed differences is included in Section 5.1. However, we include more details on the difference in onset of melt in Section 4 (L.275): "At the beginning of the melt season, ASCAT detects the increase in the extent of liquid water 10-15 days earlier compared to when the RCMs simulates an increase in the melt extent."

In section 5.2 we have added the following:
"However, based on the Figure 7 we asses that on average ASCAT detects the liquid water more than 4 days before the RCMs simulates melt, meaning that the processing averaging cannot fully explain the differences between ASCAT and RCMs."
And further in section 5.2:
"On average, the magnitude of the melt seasonal cycle of melt extent agrees well with RCMs, suggesting that luquid water is observed earlier but at similar extent."
We lastly we include some of these observations in the first paragraph of the Conclusions section:
"For all RCMs included in this study, the onset of melt occurs more than a week later than what is observed from ASCAT even when accounting for the averaging of satellite data. Further we see that the regions with the largest differences in total number of melt days across all RCMs are the SW and SE basins of the ice sheet."

4. I have struggled to understand the goal of ensuring "that the RCM-modeled melt aligns with in-situ observations". Is the motivation here to suppress/separate differences in melt rates which are due to temperature biases? I see the danger that by applying such a first order bias correction, you might blur important spatial patterns (I find Fig. A1 quite informative). Another strategy might have been to diagnose the AWS temperatures for which melt typically is detected by ASCAT, (instead of the individual RCMs) and diagnose which melt rates are typically produced if this temperature is simulated- please motivate your choice. Also, are the melt thresholds a good choice for all regions, altitudes and seasons? Maybe put some additional figures in supplement to illuminate this.
To derive the melt extent from RCMs we need to used a melting threshold. By using the uniform 0.1 mmwe/day we favor models that simulates low melt. Instead we perform an indepent comparison using in situ observation to infer a melting threshold. To make this more clear we have added the following to the beginning of section 5.1 (L308): "To get the most valid comparison between each RCM and ASCAT, we utilize in situ observations to assess biases and to determine an appropriate threshold for the melt extent in RCMs. By fitting each RCM to in situ observations we minimize the differences that are introduced due to model set-ups like resolution, parameterization etc. Thus, we reduce overall inter-model discrepancies as well as differences in melt extent compared to that observed by ASCAT. Despite applying the in situ informed thresholds, persistent patterns between RCMs and ASCAT remain."

5. Figures could be improved by introducing color scales with discrete colors
We have revised accordingly. All colorbars are now segmented colorbars. Further, we have chosen another colorbar without yellow as a pivotal color for Figure 3, 6 and A1, as suggested by referee #1. We refer to the updated MS for the updated figures.

Specific comments:
Abstract:
it should be included which years are covered in this investigation
We have revised accordingly. L4-6: "Here, we explore novel processing of data from the Advanced SCATterometer (ASCAT) instrument onboard the EUMETSAT Metop satellites, which provides estimates of the spatiotemporal variability of melt extent over the Greenland Ice Sheet between 2007 and 2020"

Introduction

l. 23: More precisely: Since 1992→Between 1992 and 2020

We have revised accordingly.

l. 24: I think *Otosaka et al. (2023)* don't provide an estimate for the contribution of the SMB component. Please provide the reference (60% due to enhanced melt according to *van den Broeke et al., (2016)*?).

We have revised the references accordingly: L22-24 "Between 1992 and 2020, satellite observations have shown that the Greenland Ice Sheet has lost 4892 ± 457 Gt of ice or 13.6 ± 1.3 mm sea level equivalent (Otosaka et al., 2023) with half of the mass loss attributed to a decrease in the surface mass balance (SMB, van den Broeke et al., 2016)"

l. 33: "the only approach"→that is not true, maybe rather: the most comprehensive, or: RCM simulations agree best with observations (see Fettweis et al. 2020)

We have revised accordingly. L34: "At present, regional climate models (RCMs) provide the most comprehensive approach for obtaining ice-sheet-wide estimates of meltwater volumes and runoff, with simulations showing the best agreement with observations (Fettweis et al., 2020)."

l.38: Fettweis et al. (2020) do not analyze future scenarios

We have revised accordingly and found a more suiting reference. L40: "… representing differences in parametrizations, have large effects on projections of melt, runoff and surface mass balance when run into the future, giving greater uncertainty on sea level rise estimates than desirable for climate adaptation purposes (Goelzer et al., 2020)"

l.42: the AWS network does not directly measure melt intensity

We have corrected so the statement makes it clear that melt intensity has be derived from AWS stations, but not observed directly:

L42-44: "While melt intensity can be derived from in-situ observations at automatic weather stations (AWS), the sparse distribution of these stations across the ice sheet limits the evaluation of melt estimates beyond local scales (Fausto et al., 2018)."

l.53: I understand that ASCAT can detect the onset of melt, but can it also detect the cessation or interruptions of melt? Is the decreased backscatter signal solely due to the presence of near surface water or would it remain low after a melt event due to changed chrystaline structure?

This is described in full detail in the ASCAT melt maps section 2.3

l.56: maybe: properties of the snow *pack*

We have revised accordingly.

l.57: delete: "*Refrezing of* " or rephrase

We have revised accordingly (L: 63): "Refrozen meltwater from the previous melt season can percolate into the firn"

l.60: do you mean: can be weakened by moist subsurface layers?

To convey this more clear in the MS we have edited the sentence (L65): "Further, meltwater in the subsurface can still be detected after refreezing of the surface layer as the low-frequency signals can still penetrate into the refrozen surface layer"

l.62: is this really subsurface melt or rather meltwater in the subsurface?
See above for the suggested edit.

Sect.: 2.1: I am missing the information which stations and how many measurements are included, do you use hourly or daily data- maybe include a table with station, location, elevation and number of temperature measurements going into this study.
Agree. We have revised the beginning of the method sections: "To evaluate temperature biases in the RCMs, we compare the modeled output with daily observations from 2007 to 2020 by the PROMICE GC-net AWS shown in Figure 1."

l. 89: I think here you want to point out that the same climate forcing may result in different melt products? Maybe elaborate and discuss whether it is possible to distinguish differences due to atmospheric differences in the RCMs and differences due to different representation of the snow pack.
This is a good point, that even though the three RCMs are forced with the same reanalysis data the results can differ. This is due to model setups and parameterizations, some of these differences are described in the three subsections "HIRHAM5", RACMO2.3p2", and "MARv3.12" in the manuscript, some of these subsections have been updated in this round of review. Regarding the distinction between different atmospheric models and different snowpack schemes; it is not really possible to say where the differences arise from when we do not have a combination of the different RCMs with the different snowpack schemes. However, as stated earlier the aim of this study is not to explore why RCMs simulate melt rates differently. Instead, we want to show how we can use ASCAT melt extent. We have re-written the sentence (L.99): "However, different model setups such as horizontal and vertical resolutions, choices of parameters like surface albedo and subsurface schemes impact the surface energy balance simulated"

Page 5: the different albedo schemes in HIRHAM5-ERAI and HIRHAM5-ERA5 will influence the melt production considerably – I recommend to acknowledge this also by some different naming to avoid misunderstandings. Maybe also include figures illustrating the relation between temperature, albedo and melt (e.g. as scatter plots)
It's very true that the different albedo schemes will likely influence the melt production, which is also discussed in detail in the discussion section of the MS. We have chosen the naming conventions as concise as possible to maintain clarity throughout the paper.

l.109: Typo (R)ACMO
We have corrected accordingly.

l. 1l1: delete once*: "On the lateral boundary"*
We have corrected accordingly.

l. 113: which process/forcings influence snow grain size and impurity concentration?

Snow grain size can change due to multiple reasons like, the thaw/freezing cycle, compaction/settling. The impurities can come from dust.

l. 126: please provide some more information on the albedo scheme
Based on this comment and a similar comment from referee #1 we have added the following to the description of MARv3.12 (L.135-141): "MARv3.12 includes the snow model Crocus (Brun et al., 1992), that simulates a number of layers of snow, ice, or firn of variable thickness and energy- and mass-transports between each layer. The snow model also provides snow grain properties, which are used in combination with density, age, and type to simulate snow albedo (Brun et al., 1992; Fettweis et al., 2017; Antwerpen et al., 2022), MARv3.12 also have an albedo range for bare ice between 0.4 and 0.55 depending on the cleanliness of the ice (Fettweis et al., 2017). While both RACMO2.3p2 and HIRHAM5-ERA5 incorporate MODIS observations into the albedo computation, the surface albedo in MARv3.12 is only based on the internally computed broadband albedo (Brun et al., 1992)."

l. 128: Is the full 2007-2024 period included here?
No, while the satellite is operational since 2007 to present, the SIR data product only runs from 2007 to 2020. To make this more clear we have revised L158: "The ASCAT SIR product is available from 2007 to 2020 and is used to identify four different melt stages by applying a hierarchical decision tree using dynamic thresholds based on the previous winter reference month…"

l.145: "*first and second*" redundant?
We have corrected accordingly.

l. 150: please avoid jargon- what is a fully saturated signal?
We don't believe that a 'fully saturated signal' is jargon. Further, the meaning is explained in the same sentence: "an increase in the melt intensity does not lead to further lowering of the backscatter signal"

l. 159: confusing statement, maybe you want to state, that, against expectation, no melt is detected near the margin even though melt is detected at higher elevation?
We have revised accordingly to make it more clear in the MS (L160): "Further, against expectation refreezing or no melting is observed in pixels close to the margin when liquid water is detected at higher elevations on the ice sheet."

l.162: maybe: associated with bare ice outcropping?
We are not sure we can follow the question stated here, in the sentence we are argue for a removal (outcropping) of bare ice. Please elaborate this point.

Methods:
generally: how do you deal with the spatial and temporal resolution of RCMs, ASCAT and AWS data?
Se answer to 1st general comment, where we address this.

l.172: how do you diagnose temperature bias by comparing melt (flux?) with 2m-temperature? And why would you? Why don't you simply compare simulated to observed temperature? I think there is an implicit intention here, which should be spelled out.

The aim is to quantify a threshold for melt extent in the RCMs. Instead of choosing an arbitrary melting threshold we used in situ observations. However, PROMICE GC-net Aws do not observe melt directly. Therefore, we use temperature as a indicator of melt. Based on your comment and a similar comment by referee #1, we have edited the methods to: (L 185) "To evaluate temperature biases in the RCMs, we compare the modeled output with observations from PROMICE GC-net AWS. 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 simulates melt compared to in-situ observations."

l. 172ff: A lot can be said here: 1) Please distinguish clearly between observing melt (surface temperature at melting point) and using a threshold *air* temperature as an indicator of melt (especially if mean temperatures are used). Specifically here but also anywhere else. Do you use daily maxima from hourly temperatures (I would recommend to do so...)?

See above reply to comment and what we suggest changing the beginning of the method section of the MS.

2.) If we define a threshold in air temperature which marks the transition from no melt to melt- how much does it dependent on location, season, elevation? It could be instructive to diagnose threshold temperatures seasonally and locally in a similar fashion from simulated temperature and simulated melt.

While it's definitely interesting to diagnose the temperature thresholds seasonally and locally we consider that beyond the aim of the MS. In order to derive the melt extent, we need to apply a threshold. The goal of using the in situ temperature measurements are to get an indication about what threshold to apply while we also evaluate the temperature biases in the RCMs ice sheet wide. With the suggested changes to the MS we believe this becomes more clear.

3) and finally (-: ... there is no secret connection between the modeled and the observed world- so observations can show anything independent of what the model simulates...

We agree, however, we can't find any place in the MS where this is suggested to be otherwise, could you please expand on this statement.

l. 176: here the authors are risking that readers are equating surface temperatures and air temperatures (more accurate: near-surface air temperature or 2m temperature). Also, surface temperature of a melting snow surface cannot be above 0°C.

We suggest making the following change to the MS (L194-198): "Air temperature is strongly correlated with melt since melt is a response to a positive surface energy balance, which occurs when the surface temperature is reaches than 0°C (Cuffey and Paterson, 2010). However, it's important to note that air temperature and surface temperature are not the same; while air temperature influences surface conditions, surface temperature depends on a combination of energy exchanges at the surface. Additionally, the local properties of the snowpack can also affect when melt occurs, and melt can occur in the snowpack when air temperatures are below 0°C."

l. 177: melt can also occur also at the surface when T2m < 0°C (low albedo, intense radiation...).
See response to above comment, where the edits also takes this comment into account.

l. 182: the melting point is defined as the temperature (not the air temperature) at which snow/ice melts, please avoid to use this word in the context of air temperature.
We have corrected accordingly (L.202). "We explore various thresholds for temperature observations to account for other factors in the snowpack that influence when surface melt occurs"

l. 193: Isn't it: FPR=1-TN/(TN+FP) ?
yes, we have corrected accordingly.

l. 206 + l. 207: in my understanding statements are contradicting here.
They are not contradicting. In 206 we talk about the temperatures from the ROC-curve but in 207 we talk about the PR-curve. However, we do agree that this should be stated more clear (L.228): "For MARv3.12 and RACMO2.3p2 the ROC-curve suggest using temperature thresholds between 0.5 to 1 C to find the best RCM melting threshold, but the PR-curve the suggest a lower temperature threshold between -1.0 to -2.0 C yields better results."

l. 210: This statement should be supported by some statistics rather than selected measurements
We have put this statement in the results since we believe it fit better. Further we have revised the statement (L238): "However, the mean air temperature at selected stations suggest that the temperature bias is not systematic across the ice sheet."

Results:
l. 214: correct: each drainage basin
we have corrected accordingly.

l. 216: "*regridding biases*" should be introduced somewhere beforehand, in general iformation on how data are regridded
See reply to first general comment.

l. 226: Fig 4c→Fig. 4d?
we have corrected accordingly.

l. 241: average maximum is here the multiyear mean of yearly maximal melt extent? Maybe rephrase.
We have rephrased the sentence: (L.275): "The maximum melt extent is on average approximately 30 % of the ice sheet, except for HIRHAM5-ERA5 with >35 %."

l. 247: Provide the date of the melt event and also show only few days or weeks before and after the melt event in Fig 7b.
We have revised accordingly. L 281: "On 12. of July 2012, an extreme melting event was observed across almost the entire ice sheet (Nghiem et al., 2012)."

l. 262: maybe: indicate that HIRHAM-ERA5 overestimates melt
We have revised accordingly. L 295: "The melt threshold in HIRHAM5-ERA5 is considerably higher than the remaining melt estimates, indicating that HIRHAM-ERA5 overestimates melt."

l. 262-263: It is also possible that the bias of HIRHAM-ERA5 is related to the albedo scheme
Yes, it is very possible that it is the albedo, as HIRHAM5-ERAint uses albedo observation where HIRHAM5-ERA5 uses internally computed albedo.

l. 285-293: HIRHAM-ERA5/I are forced by reanalysis only at the lateral boundaries of the Greenland Ice Sheet and still seem to express similar differences in the interior domain as ERA5 and ERAI. This might indicate that the observed differences between ERA5 and ERAI originate from the farfield outside of Greenland- however Delhasse et al. () don't find corresponding differences in MAR-ERA5 and MAR-ERAI.
An important factor to consider is that regional climate models are often recalibrated to fit present-day ice sheet observations when forced with different reanalysis datasets. This recalibration process helps account for biases in the reanalysis. While Delhasse et al (2020). may not have explicitly described such recalibration in their study, it's likely that it was performed, which could explain why MAR doesn't show the same sensitivity to ERA5 and ERAI as HIRHAM5 does. The recalibration would mitigate the impact of any farfield differences in the reanalysis data on the MAR model. Another thing to remember, as written above, the two HIRHAM5 simulations, does not only have different forcings at the bounderies, but also different albedo schemes, one based on observations and the other one based on internally modeled albedo. Yes, it is very possible that it is the albedo, as HIRHAM5-ERAint uses albedo observation where HIRHAM5-ERA5 uses internally computed albedo. As the SMB model in HIRHAM5 is run offline, the albedo difference does of cause not explain the temperature difference, but it is likely to be the driving difference for the melt.

l. 305: "model parameter" would be understood as some internal parameter, which would change model behaviour; maybe: model parameters→simulated melt rates
We have revised accordingly (L344-345): "By aligning simulated melt rates more closely with observational data, we can improve the model estimates of meltwater production and ultimately runoff."

l. 316: I don't understand retieved in this context
Using the hierarchical decision tree approach allows us to classify refreezing in the ASCAT signal. However, we compare this to melt in the RCMs we do not include periods identified as refreezing by ASCAT. Based on a first general comment by referee #1 we have made several alterations to the description of ASCAT melt maps. Further to clear up any confusion we have edited the following: (L345): Since the refreezing periods identified from ASCAT data are not included in the melt season analysis…"

l. 320: it would help to know how many days earlier melt is detected. On could also produce melt datasets from RCM output which are smoothed by a 4-day moving mean, to test if the temporal averaging explains discrepancies.

We have included an estimate of the number of days ASCAT detects liquid water earlier compared to RCMs simulates melt. Since it's more than 4-days we do not compute the 4-day averages for the RCMs since this processing averaging cannot fully explain the difference. We have added the following to the result section (L276): "At the beginning of the melt season, ASCAT detects the increase in the extent of liquid water 10-15 days earlier compared to when the RCMs simulates an increase in the melt extent."

Fig 4d: Are there regions where no melt is detected? Please mask these out similar to figure 6.

We have now revised accordingly so that figure 4d have a segmented colorbar and zero melt days is shown as white. See MS for reference.

Fig 5: I think it would be more helpful to colorcode the melting threshold and to plot lines for fewer temperature thresholds which could be labeled with numbers.

We have updated the figure, so it includes fewer temperature lines and we have labeled the chosen melt threshold so that it is included in the figure now. We refer to the revised MS for the updated figure.

Fig 6: Please also provide differences relative to the number of melt days

We considered including a figure showing differences relative to the number of melt days; however, to keep the manuscript concise and focused, we limit the figure to number of melt days and differences in number of days. That way we want to keep a focus on where there is a high difference in number of melt days.

References (which are not also given in the manuscript):

Delhasse, A., Kittel, C., Amory, C., Hofer, S., van As, D., S. Fausto, R., and Fettweis, X.: Brief communication: Evaluation of the near-surface climate in ERA5 over the Greenland Ice Sheet, The Cryosphere, 14, 957–965, https://doi.org/10.5194/tc-14-957-2020, 2020.