A Review of: "*Bias in modeled Greenland ice sheet melt revealed by ASCAT*" Puggaard *et al*. 2024, The Cryosphere Discussion.

Purggaard and co-authors compare here extent of meltwater production over Greenland as represented by 3 different RCMs to a remote sensed product from ASCAT instrument. Their aim was to identify differences between these 3 RCMs. To realize this comparison, they needed to determine a threshold of melt in the model which append before melt events are actually detected at the surface which infer the melt extent in the RCMs. To do so, they compare melt rate from the different models to in-situ observation of the near-surface temperature from AWS (PROMICE and GC-Net networks). As these thresholds are adapted to each model, discrepancies between ASCAT product and RCMs, as well as between models, are reduced (compared to a fixed threshold common to all models).

Main conclusions are

- biases between the RCMs are coming from: different albedo schemes, representation of the snowfall, and temperature as well as differences in radiation schemes;
- The ASCAT product presents some limitations (mismatch in the ablation area of the ice sheet, bias in the melt season temporality);
- RCMs present later melt season, and earlier end of this season than ASCAT;
- Slightly smaller melt area in RCMs than in ASCAT.

General comments

The manuscript is generally well written and easy to read. The originality and interest of this study come from the comparison to several modeled melt products. I will have several major comments to improve the quality and the consistence of the manuscript.

In the method, different parts are poorly described which not deserve the study. More the way you build your datasets, and your tools are described, better is the clarity of the method, specifically if external people to remote sensing community is reading your work. These parts are highlighted with some specific comments in the next section and should not be difficult to address.

It seems to me to be rather poorly considered that the data derived from ASCAT don't give an indication of meltwater in particular, but of all the water present on the surface and in the first layers of the snowpack. This is well presented in the introduction, but in some other parts of your text, this is a bit more confusing. Does it make sense, then, to compare this extent with melt extent in the models? And not the liquid water content in the first snow layers (see for instance Dethinne *et al*. 2023 and Picard *et al.*, 2022)? Are we sure that the variables selected in the models are indeed representative of what the satellite observes? From another point of view, if we look at the problem the other way round, we

could ask ourselves whether an additional ASCAT data processing step could be applied to translate them more specifically into meltwater.

The comparison between RCMs is original for this type of comparison (not included in a "MIP" exercise) with remote sensed data but is also audacious and complicated. This kind of work require deep knowledge of how processes inside the model are modelized/parametrized to be able to compare it with other products. I recommend to deeply double check the way that components of meltwater production or water in the snowpack are determine in the 3 different models and be sure that everything is comparable. In that aspect, not including MAR and RACMO people is a risky move, or at least limit the study to HIRHAM alone. The idea is not to transform this study into an intercomparison of these 3 models, but to better understand it to highlight why the biases you mention are present.

Specific comments

Data

L119: For MAR add the same detail-level than the 2 other models (at least how albedo is retrieved).

L136-137: Could you a bit more detail the SIR algorithm? As you need more measurements, is the reconstruction constant in time and in space? Are there any supplementary uncertainties bring with this method?

Methods

- L170-171 : "We compare the RCM output of surface melt with observed 2 m temperature data", even if it could be obvious, if think this kind of sentence could be not so easy to understand at first read. I suggest switching some parts of the second paragraph of your method with the first one to be more readable.

- Could you more detail how you construct the ROC- and PR-curves. Specifically, how do you determine if it is a true or false melt-day compared to AWS if you already calculate it for different temperature rate. If I understand well, you don't have directly melt rate observed at the AWS, but a guessing relates to the temperature measured, transform to melt through a lapse rate correction (which should be explained here). I think something is unclear for me here because of a lack of details.

- Please, explain here your method to choose the grid cell(s) corresponding to your AWSs. Is it the nearest neighbor, or the 4-nearest, …?

- L195: You should reefer earlier to Fig. 5 to illustrate how understand the ROC- and PR-curves.

- L207: You're talking about a threshold of -2°C (also in Table 1), but in your Figure 5, your curves are only from -1 to 1 °C. How did you determine this threshold?

- Table 1: Why only one melt threshold and one per AWS for temperature? Is it an average from all melt rate retrieved at each AWS? Your method needs to be better described for that too. Moreover, are you only using these few AWSs as presented in the Table 1? Why only these ones? In the AWS description section (2.1), you mentioned that you will use all the AWS, at least much more than presented in Table 1.

- Still concerning observation from AWS, how do you manage the fact that most of the AWSs are in ablation area, and probably in the area where ASCAT cannot correctly detect the presence of liquid water, that you even mask outed? Or do you only consider AWSs inside the ASCAT mask? Otherwise, you decide to correct RCMs' melt with a threshold determine with comparison outside your area of melt comparison, which is not 100% valuable. Also, could please consider adding this mask in your first figure to well situated it compared to AWSs' localization as well as a more detailed comment on how you retrieved it.

Results

- L242-243: "ASCAT detects the increase in melt extent earlier compared to RCMs" Isn't it due to the detection of water by satellite and not directly melt (cf. 2nd general comment)?

- L249: can we talk about 'prediction' here as RACMO is prescribed by reanalyses at its lateral boundaries? Please rephrase with another verb.

- Table 2: Could you add the mean number of melt day for both observation and RCMs to compare your RMSE. You should also compare the difference between your two methods (uniform or in situ informed threshold) to determine if the gain with one or the other is significant (or real statistical test, it's even better).

Discussion

- L258-259: "Tab. 2 shows that by ensuring that the RCMs align with in situ measurements at specific locations." You cannot claim that RCMs align with measurement only by considering the RMSE. You need deeper statistical analyze to claim this.

- L274-276: Apply different melt threshold (based on the same in situ observation) for the different RCMs is also revealing a certain kind of bias in the model. Could you discuss that too in your Discussion?

- L297-298: First you say that RACMO present the lowest albedo, then you also explain that MAR and HIRAM have a lower albedo. Could you rephrase to better emphasize what are the differences and key features for each model/group of models?

- L300: Could you investigate a bit more why you have such differences in RACMO and HIRHAM MODIS-based albedo in the ablation area? If it's possible, it could be nice too also compare the different albedo of the models to the MODIS albedo.

- You do not talk about the differences on how the models represent the firn layer, whereas in you introduction you mention that "*The magnitude of the decrease in backscatter varies due to factors such as the snow water content and the specific properties of the snow, such as grain size and the presence of ice layers and lenses, which influence the dielectric properties and roughness geometries (Wismann, 2000; Long, 2017).* " I heard there that the signal to detect (melt)water at the surface is dependent of the snowpack conditions which are not represented/modelled in the same way in the 3 models. It should be a supplementary discussion point as melt event, and presence of water, could be delay, or more or less important, due these different way to model/parametrize the firn layers, then lead to difference when compared to ASCAT.

- Figure 3 a-d: center your color bar to 0, it's misleading as it is. And please use only 2 varying colors, one for positive and the other for negative values. Also, please avoid yellow at pivotal value.

- Figure 3, RCMs' Albedo: concerning the MAR model, you plot albedo for entire land areas and not only what looks like an ice mask in the 2 other models. Are you sure you plot the albedo used in the melt calculation, meaning the one for the ice grid points? Concerning the albedo from RACMO, considering the intercomparison and preliminary feedbacks from the PROTECT project, the albedo from RACMO presented here suggests high values, hinting a potential error when choosing which albedo plot.

Conclusions

- L340: "[…] *can lead to more accurate simulations of surface energy balance."* Could you rephrase, as you don't actually look at the entire surface energy balance, but only some components.

Appendix

Figure A1 is exactly the same than Figure 6. Is it necessary as the appendix are in the continuity of the text and not in another document as Supplements?

Technical corrections

- L109 ACMO2.3p → RACMO2.3p2;
- L111 2x in a row "On the lateral boundary,";
- L115 2x "." in a row;
- L144-145 : 2 times in a row : "the first and second";
- - Caption of Table 1: There is something wrong in this sentence: "Melting thresholds for the different RCMs based on in situ PROMICE AWS observations of 2m temperature and mean air temperature for August and July simulated by the RCMs at AWS stations and observed by the AWS stations using a lapse rate correction." I think you need to remove 'and observed by the AWS stations'.
- L232: close the bracket here: "(Fig. 6.";

- L357: HIMHAM5 data → HIRHAM5 data.

References

Dethinne, T., Glaude, Q., Picard, G., Kittel, C., Alexander, P., Orban, A., & Fettweis, X. (2023). Sensitivity of the MAR regional climate model snowpack to the parameterization of the assimilation of satellite-derived wet-snow masks on the Antarctic Peninsula. The Cryosphere, 17(10), 4267-4288.

Picard, G., Leduc-Leballeur, M., Banwell, A. F., Brucker, L., & Macelloni, G. (2022). The sensitivity of satellite microwave observations to liquid water in the Antarctic snowpack. The Cryosphere, 16(12), 5061-5083.