Review of "**Bias in modeled Greenland ice sheet melt revealed by ASCAT**" by Anna Puggaard and co-authors

This manuscript combines a novel processing of scatterometer data (ASCAT) with regional climate model (RCM) data to investigate the melt extent above the snowline of the Greenland Ice Sheet (GrIS). ASCAT data are here used as a reference to evaluate differences between four different RCM derived melt products. Additionally datasets are "aligned" to observations using near-surface air temperature observations from automatic weather stations (AWS) on the GrIS. The authors find distinct differences between the different RCMs with respect to the melt extent and the number of melt days. This study is of great interest to the community as it demonstrates that satellite data might provide a valuable reference to evaluate melt products for regions with little to moderate melt. With rising temperatures these regions might become increasingly important in the near future. The manuscript is generally well written but sometimes it lacks attention to detail. The objectives of the study could be accentuated more. Also the chosen strategy how to "align" observations and model data is poorly motivated.
In summary I recommend major revisions before publication.

General remarks:

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.

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.

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.
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.

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.

Figures could be improved by introducing color scales with discrete colors

Specific comments:

Abstract:
it should be included which years are covered in this investigation

Introduction

l. 23: More precisely: Since 1992 → Between 1992 and 2020
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)*?).
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)
l.38: Fettweis et al. (2020) do not analyze future scenarios
l.42: the AWS network does not directly measure melt intensity
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?
l.56: maybe: properties of the snow *pack*
l.57: delete: "*Refrezing of*" or rephrase
l.60: do you mean: can be weakened by moist subsurface layers?
l.62: is this really subsurface melt or rather meltwater in the subsurface?
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
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.
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)
l.109: Typo (R)ACMO
l. 111: delete once*: "On the lateral boundary"*

l. 113: which process/forcings influence snow grain size and impurity concentration?
l. 126: please provide some more information on the albedo scheme
l. 128: Is the full 2007-2024 period included here?
l.145: "*first and second*" redundant?
l. 150: please avoid jargon- what is a fully saturated 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?
l.162: maybe: associated with bare ice outcropping?
Methods: generally: how do you deal with the spatial and temporal resolution of RCMs, ASCAT and AWS data?
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.
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…)? 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.
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…
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.
l. 177: melt can also occur also at the surface when T2m < 0°C (low albedo, intense radiation…).
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.
l. 193: Isn't it: FPR=1-TN/(TN+FP) ?
l. 206 + l. 207: in my understanding statements are contradicting here.
l. 210: This statement should be supported by some statistics rather than selected measurements

Results:

l. 214: correct: each drainage basin
l. 216: "*regridding biases*" should be introduced somewhere beforehand, in general iformation on how data are regridded
l. 226: Fig 4c → Fig. 4d?
l. 241: average maximum is here the multiyear mean of yearly maximal melt extent? Maybe rephrase.
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.
l. 262: maybe: indicate that HIRHAM-ERA5 overestimates melt
l. 262-263: It is also possible that the bias of HIRHAM-ERA5 is related to the albedo scheme
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.
l. 305: "model parameter" would be understood as some internal parameter, which would change model behaviour; maybe: model parameters → simulated melt rates
l. 316: I don't understand retieved in this context
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.

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