1 General comments

This sensitivity study investigates possible assimilation parameterizations to integrate satellite-derived surface melt extent into regional climate model simulations. Such an assimilation is a promising approach with the potential to improve surface mass balance estimates for Antarctica. Melt extent assimilation could allow to translate satellite-detected melt water into quantitative melt rates and might reduce the uncertainties in model based estimates of surface melt and as such is a much needed development.

The authors use a generally similar assimilation strategy as in Kittel et al. (2022) by nudging the snow temperatures to improve the agreement between simulated and a satellite-derived melt mask. Here, instead of using just one satellite product, the authors include various combinations of data sets which differ in temporal and spatial resolution and penetration depth. Additional parameters tested are the depth down to which modelled snow temperatures are nudged and the minimum in simulated water content which is sufficient to count as an agreement with liquid water detected by satellites.

The paper has a few major issues but also numerous minor, rather technical shortcomings which in my view could have been straightened out beforehand. I recommend major revisions with an more active contribution from experienced co-authors. In some places I have tried to improve the wording, but this is only intended to give an inspiration and makes no claim to be perfect or a complete list.

2 Major and general comments

The paper generally identifies different sensitivities for the different assimilation parameters under consideration, however it is not attempted to identify a recommended assimilation algorithm. A basis for such an assessment could be to evaluate which experiment yields good agreement with the binary melt masks with minimal nudging. To this end it could be interesting to compare the values in Table 5 and Figure 11 to respective values diagnosed from the satellite-derived melt masks. Also Figures 12 and 13 might be extended and discussed in greater depth. And it would be interesting to map the total energy added and subtracted (separately) throughout the experiment within each grid box and to provide a budget for the whole domain.

The experiments which use only one data set should also be part of section 2 and should be discussed more systematically and in more detail. Due to the great number of experiments it is difficult to get an overview. It would be helpful to have one table with all experiments (table 2 does not include the one sensor experiments) and another one with some simple metrics such as the number of melt days and the total meltwater production over the whole melt season for each experiment.

Liquid water in the snow pack is not necessarily indicating ongoing melting- it can also indicate past melt events with incomplete refreezing at night (e.g. for cloudy conditions or at greater depth)- this should be distinguished and also discussed with respect to the different duration of the melt season for the different satellite data sets. Also I would not use the term of "binary melt masks" - but something like "wet snow masks". Furthermore I think that possibly valuable information is discarded when shallow penetrating data sets indicate no wet snow while the deep penetrating data set indicates wet snow. This is not necessarily a conflict but could occur after a melt period has ended but percolated water may remain liquid in deeper layers.

As the satellites only detect presence of liquid water and not melt, I am also suprised that sensor penetration depth and assimilation depth are so closely linked here: I would not expect melt at depths of 1 m or more. I would rather have limited

the temperature nudging to a much shallower surface layer. However one could still compare water content (here I would use absolute and not relative values) down to the respective penetration depths of the available measurments and then trigger melt only in the surface layer. I wonder if there are reasons against such a strategy.

Sometimes the word *sensitivity* seems to be used ambiguously. Most of the time it is used as in "simulated melt (whether more or less) depends strongly on the parameter choice" but for instance in l.14-15 it seems to rather mean "more melt is detected for a certain parameter choice"

The words used for the assimilation parameters are unnecessary diverse and confusing. I would recommend to consistently use something like melt water threshold and assimilation depth. The latter should not be named threshold in my view and it also should not be called penetration depth as this can be easily confused with the penetration depth of the individual sensors.

The introduction is too unstructured, I give some specific comments below, but these should be only considered after sorting the different aspects in a linear fashion.

Also the method part is hard to read and should be thoroughly revised.

Maybe it is not a problem for people from the remote sensing community- but the paper is not easily readable for the wider community. For instance, datasets are sometimes referred to by their mission (Sentinel), the general measurement (radar, radiometer, scatterometer), some general classifications (active or passive sensors) or their instrument name (ASCAT)- this is unnecessarily confusing.

3 Some specific comments

Title: it is the MAR snowpack which is sensitive to the assimilation, not the satellite-derived surface melt. Maybe: Assimilation of satellite-derived surface melt into the regional climate model MAR: sensitivity of the snowpack on the Antarctic peninsula to assimilation parameters

Abstract:
l. 1: please reword the whole sentence and possibly add 1 or 2 sentences. Here it would be good to introduce the problem (e.g. surface melt, runoff and accumulations cannot be directly observed on larger scales and models have uncertainties, remote sensing can only provide melt extent)
l. 3: rather use "reduce uncertainties"
l. 18: maybe: second parameter mostly influences the duration of the melt period but it has only limited effect on the absolute melt water production.

Introduction:
l.29: maybe: even moderate surface melt is thought to weaken ... leading to substantial mass loss.
ll. 31-34 too long, muddled, partly redundant.
l. 35 climate models do not monitor (wrong verb), they do not comprise the ice body and only few include the snow pack.
l. 47: correct: induce -> induced
l. 49: delete: In addition
ll. 54-61: this is better placed earlier in the section and merged with the earlier sentence about ice shelves
l. 62: be specific, here: melt -> melt areas
l. 62: include Kittel et al. (2022) here and generally explain that the strategy is to warm or cool the snow pack in order to better match satellite derived melt maps.

ll. 66-72: this is too general and does not get to the point. I think you wanted to say that the different available products yield either poor spatial or temporal resolution and in contrast to Kittel et al. (2022) you test combining several products.
l. 68: active / passive sensors should be explained in this journal before using these terms

Section 2.1: this section could be better structured. Consider implementing subsections for each satellite/sensor type and providing a table with technical specifications (e.g. mission, sensor, resolution, revisit time, reference). It would be good to have a table with unique names for the four data sets and their technical specifications, and then to only use the data set names.
l. 78: four data sets from three sources?
l. 78: the fact that the strategy is to produce binary melt masks from satellite data and assimilate these should already be spelled out in the introduction and abstract.
l. 90: this is confusing as here only three data sets are mentioned
l. 94: "level 3 products" seems to be an unnecessary detail.
l. 96: ascending and descending paths should be explained.
eq. 10: how is TP measured?
l. 104: the reference to Fig. 3 is confusing here; I propose to refer to Fig. 3 in l.99 and refer to Fig. 2 at the end of l. 105.
l. 116: maybe better: sensors will indicate the presence of water .. by changes in the backscattering.
l. 125: this is not coming to the point: the -2.66 dB threshold is used in this study?
p. 7, Fig. 3: use coast line contours also in the upper panels
l. 140: explain or avoid the word scene in this context
l. 145: it is unclear to which part the word "else" is related to.
p. 8, Fig 4: it could be interesting to see panel B after normalization.
l. 146: A figure for the ASCAT data could be included, similar to Figs. 3 and 4. Also a reference for this data set is missing.

Section 2.2:
l. 157: "transfer between atmospheric part ... and the atmosphere" is this right?
l. 159: What is the typical vertical resolution in the upper 1.5m?
Also please cover the percolation algorithm which seems to be crucial to understand the response in subsurface liquid water content

Section 2.3:
Table 2 should be introduced in this section. Also it should include the single-data set experiments and I find experiment name $MAR_{a01-ku-02-c10}$ unfortunate as it does not indicate that here a different input is used and it does not indicate the assimilation depth of the third data set. Generally the experiment names are not very handy. I would suggest something like $AsSd_l$ for AMSR with shallow assimizlation depth+S1 with deep assimizlation depth and low water content threshold.
l. 173-174: check grammar
l. 175: correct: As up to three...
l. 183: is it possible to heat beyond 0^oC?
l. 187: either percolate into the ice or accumulate in the ice
l. 190: better: discarded -> ignored
l. 192: shorter: if the two masks agree, the two observations...
l. 199: more precise: at the same time -> within the same 3-hour time window
l. 206-207: maybe put this first
l. 213: check unit

Section 3:
Since the assimilation and the analysis are dealing with the snowpack it would be helpful to also evaluate precipitation and melt. Maybe compare to Wang et al. (2021).
l. 239: correlation **(r)**
l. 249: better: a weak correlation and/or a strong negative bias

l. 250: actually the bias is also strongest in summer (winter insolation should be weak anyway) and biases in net longwave radiation and net shortwave radiation almost cancel out and indicate underestimated cloud cover.
l. 255: "Combined with ...": unclear

Section 4:
$Assim_{mean}$ is an unfortunate name, as it sugests an experimemt of its own right- I would propose $\overline{Assim}$ or mean(Assim). Also it should be stated here that three experiments were discarded in $Assim_{mean}$
l. 258: correct is->are
l.: 266: correct model-> simulation
l. 269: gives different results from -> differs from
Table 3: Evolution is not (but should be) mentioned in the caption- I understand that evolution is relative change due to assimilation in $Assim_{mean}$, the name evolution is maybe misleading. It is not clear whether LWC and ρ are mean state or final states at the end of the period. Another column for LWC_5m or some other deeper layer would be interesting. Also: replace *mean value of the assimilations* with *mean value of the 16 assimilations selected for* $Assim_{mean}$
Figure 7: runoff could be shown in the same figure with a different y-axis on the right, maybe also highlight $Assim_{ref}$
l. 271: rephrase without the first part
l. 275: please check: 63.8 is the value for runoff according to table 3.
l. 279: the same -> almost the same
l. 280 ff: clumsy, please rephrase.
l. 282: correct: depending on the energy balance
l. 284: densify -> densifies
l. 287: SMB is either $snowfall + winddrift + refreeze - melt - sublimation$
or $precipitation + winddrift - runoff - sublimation$
also please clarify whether snow drift is represented in MAR.
l. 292: please specify "deeper"
l. 293: not all ice shelves exhibit lower liquid water quantity
l. 295ff: Table 4 should be discussed in more detail: there is no explanation given, why LWC for Larsen C is increasing. Also for individual ice shelves it is not true that relative changes in melt and runoff are of similar size. It is particularly not true for Wilkins where additonal melt almost entirely becomes additional runoff. For a better understanding it could help to map the degree of saturation in the upper snow pack and to look at different stages of the melt season. A deeper interpretation of Table 3 and 4 is also difficult due to ambiguous variable definitions (see below)
ll. 299-302: this is completely unclear to me.
l. 303: It is not really suprising that the mean of the assimilation experiments is close to the central reference experiment. However without evaluation this is not necessarily meaning that this is more realistic than other members.
Table 4: The caption to this table is sloppily formulated. As in table 3 not all lines and columns are well defined or self explanatory and there is room for guesswork but no sound basis for interpretation. Also I wonder why LWC is consistently one order of magnitude smaller than in Table 3.
l. 310: refreezing is indeed releasing energy and heating the ambient snow! Please revise ll. 310-314.
l. 310: not sure what prevails means here. Maybe prevail->prevent?
l. 317: please specify what exactly qualifies results to be inprobable. Such a exclusion criterion should be defined beforehand. And the exclusion of the members should be stressed when introducing table 2 in the method section.
l. 333: lesser -> less or smaller
Table 5: Which are the experiments considered here? Are these numbers the same for all experiments with $\alpha = 0.1$ and $\alpha = 0.2$? Also maybe noteworthy: number of melt days larger for MAR_{ref} than one of the assimilations on Wilkins.
l. 343: is this referring to the whole 20m snow pack?
l. 347: I don't find these numbers in Table 5- I stop reading this paragraph here.

l. 364: this cycle -> daily melt - refreezing cycle?
l. 373: is other frequency here higher frequency?

section 5: this probably needs to be rewritten anyway after the other parts of the manuscript have been revised.
l. 394: effect of assimilation was not studied here.

References

Kittel, C., Fettweis, X., Picard, G., and Gourmelen, N.: Asimilation of satellite-derived melt extent increases melt simulated by MAR over the Amundsen sector (West Antarctica), Bulletin de la Société Géographique de Liège, 78, 87–99, 2022.

Wang, Y., Ding, M., Reijmer, C. H., Smeets, P. C. J. P., Hou, S., and Xiao, C.: The AntSMB dataset: a comprehensive compilation of surface mass balance field observations over the Antarctic Ice Sheet, Earth System Science Data, 13, 3057–3074, https://doi.org/10.5194/essd-13-3057-2021, https://essd.copernicus.org/articles/13/3057/2021/, 2021.