The authors would like to thank the reviewer for their comments and feedback. Our responses are presented in blue.

Reviewer #2

1. Description

In the article ”The sensitivity of satellite microwave observations to liquid water in the Antarctic snowpack”, the authors discuss and provide new insights into the theoretical basis of passive microwave liquid water detection algorithms over shelf ice.

First, the authors compare different permittivity models for wet snow and discuss their applicability for microwave emission modeling. Further, based on a Monte-Carlo approach, the authors retrieve the dry snowpack properties at 8 different sites on the Antarctic shelf ice and then simulate the sensitivity of different microwave channels to the appearance of liquid water in snow. The results are compared to satellite observations.

Several experiments are performed analyzing e.g., the sensitivity of microwaves to liquid water layers in snow at different depths and with different magnitudes.

In the last part of the manuscript, the authors give recommendations how their results could be used to improve existing or develop more advanced passive microwave melt detection algorithms and for an improved evaluation of firn models.

In general, this article provides interesting and new results which are of high scientific relevance and can help to improve existing or develop more advanced passive microwave melt detection algorithms. The article is clearly structured and the individual sections and experiments are well motivated. However, some of the figures need to be improved. A few statements are not clear or hard to reproduce from the figures. Additional clarification and information is sometimes required. Therefore, I recommend the article to be published after minor revisions.

2. General Comments

I have two general comments. 1) the manuscript would benefit from improving some of the figures. Some of the lines in the figures are hardly visible or distinguishable from other lines. I recommend that the authors check all figures again and increase the line width when appropriate or choose different colors for different lines. Further details are provided in the specific comments.

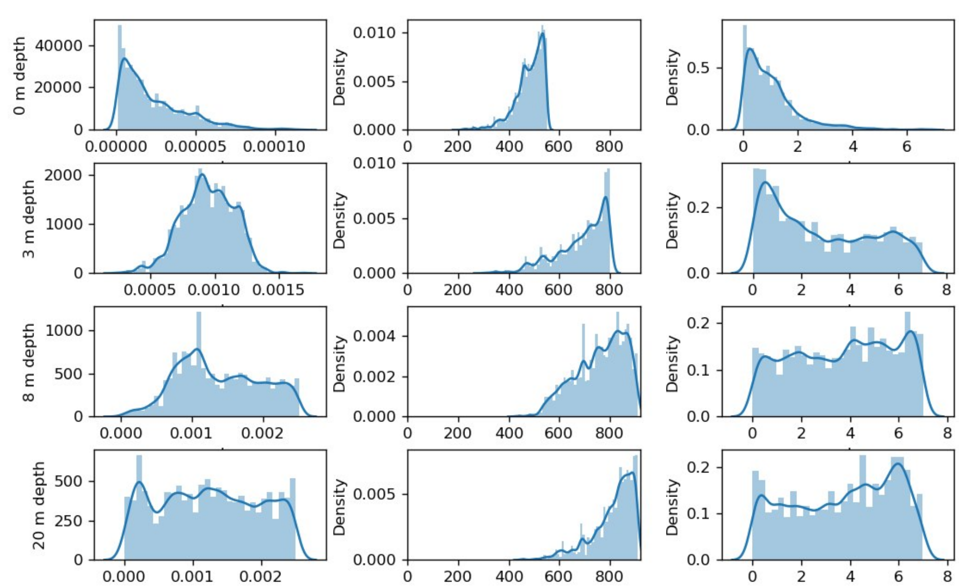
Regarding Figures 5 and 7, we acknowledge that some curves overlap but this serves our intent to communicate the range or the variations rather than each individual curve. Figure 5 is designed to show the ensemble and the range spanned by this ensemble, not the members. Figure 7 shows the variations for many sites. Our goal here is precisely to highlight these inter-site variations to support the text in the Section describing this figure. The reader interested by a specific site can still zoom in (in pdf) but this is not the purpose of this figure.

Regarding Fig 6, Fig 9 and Fig 15, the readability is affected by a limited number of colors. In fact, it is due to our adoption of a common color scheme for plotting brightness temperature throughout all our recent papers since ~2020. Each AMSR channel has a specific color. This color scheme is color-blind optimized with large contrast. However, as a consequence, figures showing one or two frequencies only have low contrast. Despite this disadvantage we think that the figures are still visible.

Regarding Fig 1, it will be improved as suggested.

2) Overall, the modeling setup is well described with the right amount of detail. This method is a very nice approach highlighting the potential strength of snow emission models like smrt. I think it could be very useful for other scientists who are interested in using the model or the method if the authors would provide a more detailed description of the experiment setup in the supplement (e.g., adding the initial range of all input parameters, differences of the different sites, similarity of the ≈100 output profiles,...)

We will add the prior ranges of all input parameters in the text, and all the results for the different sites are in Figure 3 (the mean of the properties). Only the variability of the properties is not shown in the paper. It has not been studied in detail in fact but it is overall large (see Figure below of the probability distribution of all the parameters at a selected location). This is by design of our method (e.g. using less observations than unknowns) that aims at reproducing the microwave signature during winter time, not at retrieving the properties accurately. As a consequence we don’t think we can learn a lot about the real snowpacks from the diversity of the profiles in the context of the present paper. This diversity is mainly a consequence of our choices rather than something real, different choices in the number of tie-points and type of properties would have led to different diversity. Using this method for another goal is certainly possible and interesting – e.g. to retrieve the snow properties to be used in comparison with in-situ or firn model outputs --, but it implies to reconsider these choices – e.g. take a simpler representation of the profile to let less free parameters and try to reduce parameters with correlated or anti-correlated impact on the microwave signal, in order to reduce the equifinality and increase the interpretability of the retrieved properties.



Distribution of the retrieval grain size (left column), density (central column) and ice layer number density (right column) for the four snow depths at Shackleton.

3. specific comments

L10: and on the site → Here I would specify what controls the layer depth (e.g., density, grain-shape, terrain properties)

We will reformulate as suggested:

“ii) the detection of a buried wet layer is possible up to a maximum 1 to 6 m depth depending on the frequency (6–37 GHz) and on the snow properties (grain size, density) at each site”

L18: climatic indicator → climate indicator

will be done

L47 when it reaches → It reads like you mean a threshold for the surface but instead it is a threshold for the brightness temperature

will be done

L55-56: What about snow redistribution? I believe this could be more important than snowfall events and also much harder to predict/simulate

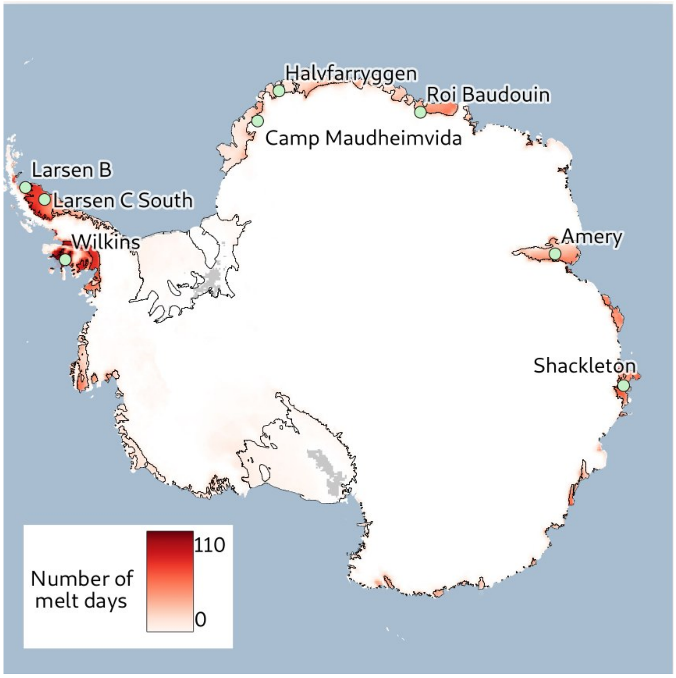
We will reformulate the examples, adding “blowing snow”.

L92: ...presence of aquifer Montgomery... → presence of aquifer, Montgomery

will be done

L95: it would be helpful if the authors would provide a map with the different study sites marked

We will insert the suggested figure (Figure 1 in the paper)..



L100: The Soil Moisture... → Observations from the Soil Moisture…

will be done

L106: Do you average the observations from the incident angle range?

The data are already binned in this range.

Table 1 Caption: AWS names are from (Jakobs et al., 2020) → AWS names are from Jakobs et al. (2020)

will be done

Table 1 Caption: You need to specify which temperature is used (i.e., 2-m or skin-temperature?)

will be done

Table 1 Caption: melt days are from AMSR2 19 GHz H-pol channel. → I assume a published algorithm is used? (Then the reference is missing) Or is it based on the 20K threshold? (Then why did you chose this channel?)

We have used a published algorithm here and will add the reference for this algorithm (Torinesi et al. 2003). Torinesi et al. is in fact often close to the 20K to our experience.

L128-130: Earlier, you wrote that you use SMOS data from 50 to 55 ◦ . I was wondering, do you then also simulate the SMRT output for this incident angle range?

We will change the text to indicate that 55° is used for both AMSR2 and SMOS in the simulations.

L134: microwave → microwaves

L175ff: Based on Figure 1, I cannot reproduce the authors argumentation that the MEMLS V3 model behaves close to the coated spheres model in the high water regime. Aren't the differences between these two model much larger than e.g., to the Colbeck 1980 (Pendular) model?

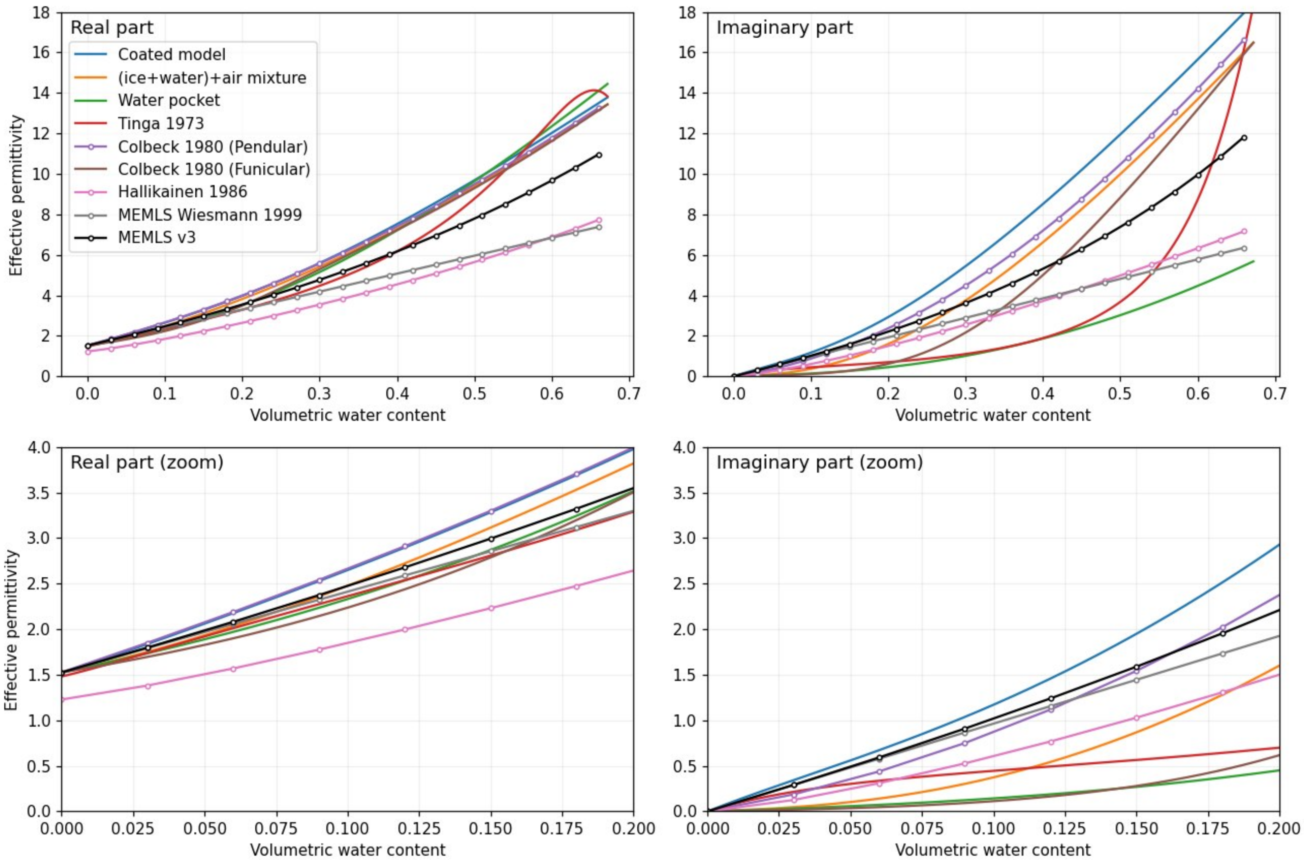
We implicitly meant relative to Hallikainen, 1986, but it was not clear. We will change this paragraph.

"In this study, we selected the MEMLS v3 formulation for the reference simulations because it is based on actual measurements, and has an intermediate behavior"

In addition, it is really hard to distinguish the different models in the low water regime, which would be the most important part in case one wants to know the minimum water content detectable by microwave observations. I suggest adding a zoomed-in version of Figure 1.

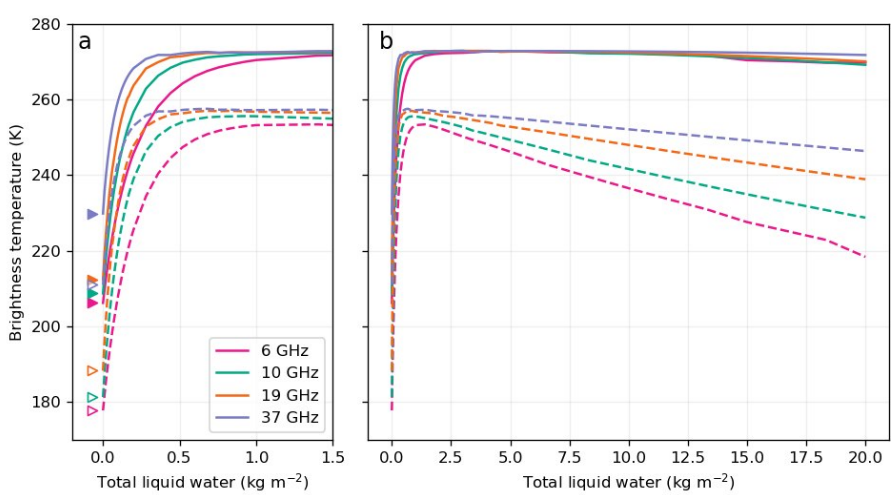
Here, we present the Fig. 1 with zoomed graphs. However, it seems not useful for the purpose of the paper w.r. to the extra space needed, because the differences between the formulations are not discussed in the paper (because we don't have clue on why these differences). The paper actually only uses MEMLS formulation.

The code to produce the figure will be made available and the interested user will be able to zoom in the graph very easily.



Also, would in not make sense to, in addition to the Hallikainen 1986, use a model which predicts higher changes in permittivity (e.g., the Colbeck 1980 (Pendular) model), to have an idea of the possible ranges of the sensitivity to liquid water?

We have added the figure here and the notebook (that will be made available upon acceptance of the paper) has the option to generate this graph. The difference is small compared to MEMLS v3 even though it features as expected a higher sensitivity, particularly at low frequencies (i.e. 6 GHz).



Same as Figure 4 but for the coated sphere permittivity formulation.

L192: Sensitivity of what? The brightness temperatures? I would've assumed that (small-scale) surface roughness variations can have an impact at least at H-Pol. Maybe the authors could provide a rough number for the impact of surface roughness variations at the snowpack brightness temperatures

We'd also expected some effect but only got a strong effect for the active simulations (not used in this paper). The use of IEM in SMRT in passive mode has not been explored in detail to our knowledge in the literature; we are not able to provide a reliable value. We propose either to remove the sentence and completely overlook this aspect, or keep it as "after preliminary tests performed with SMRT and the IEM rough surface model" which indicates that this result is weak.

L201: I assume that the 2-m air temperatures is used?

Yes, it will be added in the text.

L204ff: I wonder if the authors would have included L-Band data, how would the optimal snowpack change? I guess the other frequencies would then look worse?

Yes, it is explained in L285. Adding L-band in the optimization without adapting the snowpack representation indeed fails, and negatively impacts the higher frequencies. The reason is probably because L band is sensitive to different (big) objects that are unrelated to the snow grains. Jezek et al. 2018 mention the role of ice pipes in Greenland, and we are convinced that this also applies to Antarctica, despite a weaker melt in general.

Adding such big objects in SMRT is feasible in principle but would imply to add even more free parameters in our optimization method, increasing the problem under-determination. As a tradeoff, we have decided to focus on the higher frequencies but still perform some L band simulations, because of the high interest, with the associated adequate warnings in the text.

L201: 400−910 → wrong symbol. Also later in the document, − and - are sometimes mixed up. I suggest the authors to carefully check throughout the document

L218: d/2–2d → This reads d/2 minus 2d but I guess you mean d/2 to 2d

The symbol will be corrected (double - in latex) but we will reformulate to avoid the symbol "recommended number between d/2 and 2d"

L221: The citation should be earlier

We will reformulate " the effective sample size as defined in Martin et al. (2022) is estimated ~100"

L230: I was wondering, since you compare point simulations with large-scale satellite observations, have you considered the effect of slopes in the footprint? I guess there might as well be slopes on the Antarctic shelf-ice within one satellite pixel. If they have an impact on the observations, your retrieval method would compensate with changed snow properties and thus the snowpack might not be representative for the specific satellite pixel.

The slope on the ice shelf is usually very small (<0.5°) and the expected effect on the passive measurements is probably of the same order as a change in the incidence angle. We believe that this is a very small effect but are not aware of studies in the literature about this problem. It is true that the optimization would compensate for that, but not for a good reason.

We will add "The terrain is assumed flat" in the description of the method, next to the "surface roughness".

L250: distinctively low SMB (Table 1)

will be added

L249-250: According figure 3, the snowpack at Amery has the highest correlation length (i.e., largest grains) of all sites at 8 m (which is around the e-folding depth at 6 GHz, I guess). This could (partly) explain the relative low brightness temperatures observed at this site.

Yes, but it is more correct to say that our Bayesian method estimates a high correlation length because the observed brightness temperature is low.

Here in L249-250 we only mention the particular SMB at Amery and address the mechanistic link between the brightness temperature and the correlation length mentioned by the reviewer in L 255.

L262-264: Here, you discuss the dependence of H-Pol on the ice layer density. However, I miss the relation/implication of this for the brightness temperatures at the different site. Since you use the ice layer density as an variable in you model input, it would be nice to also discuss how (qualitatively) this variable is different at the different sites and how this relates to the H-pol observations.

We propose to reformulate the beginning of the paragraph, following the reformulation of the previous paragraph on V-pol to better explain how in principle the ice layers contributes to H-pol:

"In general, the H-pol brightness temperature is more complex because it is in part controlled by snow scattering and snow temperature (exactly as V-pol) and in addition, it is sensitive to the surface density and the vertical density fluctuations in the snowpack (layering). The ice layers decrease the brightness temperature at H-pol due to the reflections on the high dielectric contrast between snow and ice in the upper part of the firn \citep{montpetit_2013}. The variations in V-pol and H-pol are correlated and of similar amplitude only if the ice layer effect is negligible. Here we find that at the highest frequency (37\,\unit{GHz}), the H-pol variations are close to that at V-pol. The reason is that the microwave e-folding depth is about one meter (e.g. 1.3\,\unit{m} for Halvfarryggen, and 0.75\,\unit{m} for Roi Baudouin) and only a limited number of layers are crossed by the upwelling radiation over such a small depth."

The site to site differences in the retrieved ice layer number are addressed in the paragraph related to Figure 3 a bit further. These changes are certainly not sufficient to give a clear view of how H-pol and the number of ice layers vary from site to site, but this seems too difficult to achieve. In fact, we don't understand all these subtle variations.

L272: For clarification, is the RMSE calculated from the set of 100 snowpacks for each site?

Only with the best parameters. This is indicated in the sentence “The simulations with the optimal parameters”.

L275-282: I was wondering, did you use the same scaling factor for all sites? From figure 2, it looks like you sometimes over- and sometimes underestimate the observations at L-band, so wouldn’t a variable scaling factor make more sense?

We use a constant scaling factor. We will change the text to make this point clearer:

“(we found that an **overall** factor of ~2.8 is necessary)”

“The grain size is multiplied by 2.8, **for all the sites** and only for simulating the L-band brightness temperature”

Using a variable scaling factor would certainly improve the simulations in dry conditions, but we believe it would not make “more sense”. We don’t have a fully satisfying solution to improve the L band in a consistent way, but still believe the results in wet conditions are useful.

L295-299: For Amery, larsen C and Larsen B, the correlation length at 8 m is much higher than at 20m. I have difficulties to find a physical explanation for that. Could this is an artifact of the choice of the depth of the tie points. Since the e-folding depth at 6 GHz is well below 20 m depth (based on the examples you provided for Baudouin), the contribution of this tie point to the model result is very low. If that’s the case, this should be shortly discussed in this paragraph.

The profiles of the properties are linear between the tie-points, so the 20 meter tie-point does have some effect up to the upper tie-point (8 m). But we agree that with an e-folding of 5.2m at Beaudoin Ice Shelf, the contribution of depths >10m is certainly weak, and as a consequence the Bayesian method returns a relatively random value which turns out to be low here but could be large (Bayesian experts would not call this ‘artifact’ because this is an expected behavior of Bayesian approach when no information is available to constrain the value).

We propose to amend the paragraph on retrieved parameters:

“The mean retrieved parameters for all the sites are shown in Fig. \ref{fig_parameters_experiment0}. Some general observations can be made despite the risk of compensation between parameters (equifinality). **It is also worth noting that the properties at 20\,\unit{m} are not always constrained by the observations, as for instance when the e-folding depth is only 5.2\,\unit{m} at Roi Baudouin. In such a case, the method returns a virtually random value for this depth.**”

Figure 3: The color for 0 cm is hard to see. I would recommend to increase opaqueness here

will be done

Figure 3, caption: grain size → correlation length will be done

L310-311: here you write that the surface density is 220 Kgm^{-3} . However, according to figure 3, this value varies between 200 and 400.

This section is about Roi Baudouin, we will move this information before referencing the surface density (220 kg/m3).

L314: imaginary part of the water permittivity is extremely high (Fig. 1) → I'm not able to see from figure 1 due to the large range of the permittivity.

We will move the reference to fig 1 at the end of the sentence so it is now clear that the value is to be compared to 0.0017 which is extremely small compared to the range in Fig 1.

"In the first regime, the sudden apparition of water at the surface of the ice crystal sharply increases the snow absorption because the imaginary part of the water permittivity **is extremely high compared to that of ice 0.0017** at 19 GHz (Mätzler 2006) (Fig. \ref{fig_permittivity})"

L316-319: What does pixel.day mean? Which AMSR-2 observations are you using? Only for the Baudouin grid cell or for the whole Antarctic shelf region?

We will reformulate this sentence "In reality, such high brightness temperatures are rare, they can be found only 347 times in the full daily AMSR2 19\,\unit{GHz} records gridded at 12.5\,\unit{km} over 9 summers (2012–2021) in Antarctica. For comparison the number of times melt is detected on this same grid and for this same period is 2.5×10^6."

L320-324: I have difficulties reproducing the numbers given in this paragraph. Are you still describing the results shown in figure 4 at 19 GHz? E.g., none of the H-pol values in figure 4 (a-c) reach 260 K and ΔTb at 19 GHz, H-pol seems to be less than 60 K

We have corrected to "240-250 K" and "60K"

Figure 5: This figure needs to be revised since different lines are very hard to distinguish. One option could be to show less frequencies and then show an "errorbar" plot with the σ or 2σ spread as shaded contours.

As mentioned above, the purpose of this graph is to show the overall ranges, not to distinguish each line. The message we want to transmit is that the general shape of the curve is the same as in Fig 4, that the variability at V-pol is much smaller than at H-pol in general, and it is even possible to see that the variability for each frequency is almost the same.

Presenting aggregated stats (mean and sigma) rather than the raw data (ensemble member) has also some disadvantages. This is mainly a matter of personal preference and interest, it seems difficult to satisfy every reader (and reviewer) here.

Figure 5: There seems to be a step at several frequencies e.g., at 19 Ghz H-pol between 12.5 Kgm^{-2} and 15 Kgm^{-2} total liquid water. I was wondering what is the possible reason for that?

This is due to the numerical method DORT. We will add a note to point to the problem addressed in the main SMRT paper:

“Note that the small slope changes (e.g. around 17.5\,\unit{kg\,m^{-2}} at 19\,\unit{GHz}) is a numerical artifact due to this increasing strong permittivity and how SMRT deals with the refraction in the DORT method \citep[details in][]{picard_2018}.”

L345-350: Another conclusion would be use V-pol over H-pol since its much less affected by surface processes while having sensitivity to liquid water content

We don’t understand how this differs from our last sentence L349.

Figure 6: Same as figure 5, the contrast of the curve for the highest standard deviation is too low

See our response 1 to the general comment of the reviewer.

L373-374: Mention that the figure only covers the 1st regime

The sentence will be changed: “Fig. \ref{fig_sites_experiment1} shows the brightness temperatures as a function of the total liquid water conten**t (for smaller amounts than in the previous figures)** for a homogeneous pixel for the best snowpack at every site at 19 and 6\,\unit{GHz}”

L376-377: I don’t see how the Zwally algorithm mitigates the problem of high (winter) brightness temperatures, since they are using a fixed threshold (30 K) which would not be reached for e.g., Halvfarryggen

The Zwally algorithm is adaptive as it takes into account the inter-pixel variations of winter brightness temperature across the continent, which is a first step but as noted by the reviewer this does not solve all the problems. Torinesi algorithm developed a few years later is more adaptive, not only because it adapts the 30K but also (and more importantly we believe) it calculates the mean winter temperature every year.

Our sentence is to briefly illustrate the adaptability, it seems correct.

Section 4.2.4: While it is interesting to discuss the shortcoming of the H86 formulation, it would be also of value to compare the results of the (reasonable) selection of models shown in figure 1 to get an idea of the spread of the solutions.

We have shown the possible shortcomings of the H86 formulation because it seems important however, addressing the similarities of the other formulations becomes a detailed intercomparison experiment and seems overdetailed for the purpose of this section, especially because we have no objective argument to choose one or another. This would lead to a complex but inconclusive analysis.

Since we will provide the code to run the simulations and make the figures, it will be easy for the interested user to investigate this issue.

Figure 8: decrease the Y-extend

it is on purpose the same as in Figures 4, 5 and 6. We only made an exception for Fig 7.

Figure 9: The lines for 1 and 2 cm are hardly visible. Why is in figure 9b only H-Pol shown?

This figure is not perfect but the choice of color, intensity and line type is a compromise between the consistency of the color scheme throughout the paper and the specific purpose of this figure. The main purpose of this figure is to demonstrate that the total liquid water is a better variable than the liquid water content because in the second case, the results depend a lot on the layer thickness. The important point is that most curves overlap in Fig 9a but do not overlap in Fig 9b. The visibility is not great but this is also what helps to distinguish these curves from the others.

V-pol is not shown in Fig 9b because it would add even more clutter. We could remove V-pol from Fig 9a to avoid this dissymmetry, but in the end decided that this was not a serious problem.

We will add a remark in the caption: “V-pol is not on graph b for the sake of visibility.”

L405-406: We find that varying the thickness of the wet snow layer has little influence on the results if the total amount of liquid water is fixed when varying the thickness

We will change to “We find that varying the thickness of the wet snow layer has little influence on the results if the total amount of liquid water is fixed” but disagree with adding “varying the thickness” in the same sentence as suggested.

L409: for small amounts of water

will be added

L411: Given the statement ”many authors” I at least expect some references

will be added two examples: Tedesco et al. 2006 and Naderpour, R., & Schwank, M. (2018)

Figure 10: It would be nice to add the dry snowpack brightness temperatures (e.g., as triangles at the right end of the different lines)

will be added, on the left for the consistency with the other figures.

L403: (decreasing in exponential shape)

will be corrected

L434-435: especially at 37 GHz, due to the accumulation of fresh, fine-grained, snow over the summer, metamorphosed, coarse-grained snow. I have problems understanding the last part of this sentence.

We will change to “due to the accumulation of fresh snow with fine grains over the summer layer made of metamorphosed snow with coarse grains.”

Figure 11: (Caption) Time-series of AMSR2 brightness temperature at Wilkins at V-pol. The figure would benefit from adding the ERA5 2m air temperature so one can easily assess when melting starts/ends

We will adde ERA5 T2m and highlighted the temperatures above freezing.

L455: with the depth of the wet snow layer

we will add “layer”

L455: and weaker than at the

will add “weaker”

Figure 12: Add polarization shown here.

Will Add “ The polarization is horizontal. “

L456: because the dry snowpack is already close to a black body even when dry

We will move “dry” to the beginning of the sentence as suggested.

L459: features → feature

We will reformulate “The brightness temperature at 6 and 10\,\unit{GHz} features “, to solve another problem, the number 6 at the beginning of the sentence was not well separated from the 13 of the figure number in the previous sentence.

L461-463: To me, it looks like at 6 and 10 GHz, the results at V-Pol are well comparable and only at the higher frequencies, they strongly differ

They are overestimated in all the cases, but we agree that this is quite subjective. We will remove this remark.

L475: brightness temperatures

will be done

L487: characterized by a very low

we will remove “a”.

L497:500: This sentence is quite lengthy and written in a somewhat colloquial language. I would recommend to rephrase this paragraph and maybe split it into several sentences

We will reformulate the paragraph:

“Searching for a significant decrease in brightness temperature in the AMSR2 dataset have been unsuccessful even on shelves subject to ponding, as for instance on the north George VI Ice Shelf where a maximum of 15\% lake coverage was observed \citep{banwell_2021}, or in on the eastern Roi Baudouin Ice Shelf.”

L504: triggered → triggering

it seems that triggered is correct.

Figure 15: Use solid lines for V-Pol threshold

solid lines are already used for the brightness temperature variations. We need a different line style, at least to make it clear what these horizontal lines are in the caption.

L510: overall lower sensitivity of L-band compared “at L-band” will be added

L511-512: I’m not sure I understand what you are referring to here. From Figure 16, it does not look like the signal at V-pol becomes saturated at 30 kgm −2 , also I do not see a maximum at H-pol at 14 but rather at 8 kgm −2 . Please explain what exactly you are describing here

For the H-pol, there was an error in the text, the value is 6.5 kg m^{-2}. We have rerun the simulations for Figure 17 for 6.5 kg m^{-2} instead of 14 kg m^{-2}.

For the V-pol as well the limit is smaller than indicated in the text and it is better to refer to a maximum rather than a saturation, because the signal slightly decreases for larger liquid water contents. This will be corrected.

L521-522: Earlier you stated that a threshold of 10 K brightness temperature differences would still be way above the noise level. Refining this statement (so what would be the minimum acceptable threshold) would help the interpretation of the modeling results

We will add the values to make the small difference more explicit. “This value is relatively close to the dry brightness temperature (211 K vs 207 K),”.

At such a low level of differences, we can not precisely elaborate on the acceptable threshold. 10 K is certainly acceptable to our opinion, but 4 K, we don’t know. It depends on where and on the precise way the detection algorithm is working.

L527: (213 K at H-pol)

We will add H-pol earlier, after brightness temperature.

Figure 17: The different behavior of V-pol and H-pol needs to be addressed in the text. Why is there a maximum at H-pol when the wet layer is buried just below a snow layer?

This is addressed from line 526, but it was not clear because “H-pol” was not mentioned (see previous comment). This will be made clearer.

L563-563: Since the snowpack evolution during the wet season is not (or only partly in experiment 3) covered by your simulation setup, this statement is not really a finding of this study but more a general problem.

Experiment 3 shows that the maximum detection depth is slightly lower in autumn than in winter in most cases (note that the caption of Fig 3 was wrong, we have inverted dark and light). We acknowledge that the results are not marked (not as marked as we would expect), but this statement is based on this finding.

L571: I don’t think, using V-pol would ”avoid” the problem of not detecting wet snow with high water content. It would rather mitigate the problem towards slightly higher water contents

We will change avoid → mitigate.

L584: particularly on the ice shelves

will be corrected.