Review of “Cloud liquid water path detectability and retrieval accuracy from airborne passive microwave observations over Arctic sea ice” by Nils Risse et al.

The authors present a retrieval algorithm for cloud liquid water path (CLWP) over sea ice using passive microwave observations between 22 and 183 GHz. The method, based on optimal estimation theory, attempts to account for the highly heterogeneous surface emissivity by loosely coupling the PAMTRA atmospheric radiative transfer model with the SMRT model, which simulates the microwave radiative transfer through surface snow and ice. They evaluate the performance of the retrieval and apply the method to aircraft observations obtained during the HALO-(AC)^3 field campaign.

This is a substantial piece of work and represents a significant attempt to progress our ability to exploit satellite observations in polar regions. The complex methodology is clearly described and the results are presented well. However, to me the study shows that it is, in fact, very difficult to retrieve meaningful CLWP over arctic sea ice in many circumstances because the signal due to the cloud cannot be separated from the signal due to the surface properties. This is discussed by the authors, but in my opinion it is not given sufficient prominence and the results are presented in an overly positive a manner. I therefore think that major revisions are required before the manuscript is suitable for publication.

Major comments

As discussed by the authors, for a successful retrieval of cloud liquid water over sea ice it is necessary for the parametrization of the surface within the model to be realistic enough such that, in the absence of cloud, the brightness temperature departures in the retrieved state are small compared to the cloud-induced signals. These departures are presented in Sec. 4.1.1, and the authors state that “the retrieval finds a state that closely matches the observations” (L319). However, Fig. 4 shows that this is not, in fact, the case. At many frequencies the R1 (clear sky) retrieval does not find a state that matches the observations to within the effective measurement uncertainty, particularly at 118 and 183 GHz, where differences of up to 10K are seen. This implies to me that the surface parametrization is not sufficiently representative, and these departures need further analysis. It may be the case that they represent the situations with significant young ice fractions, but this is not discussed in the text. As a minimum, perhaps Fig. 4 could be divided into “central arctic” and “all” regions as in the later analysis to explore this point. The clear-sky surface classification from VELOX could also be used to separate these results by the thin ice fraction. The source of any remaining discrepancies in the absence of young ice should be further discussed.

Even in the absence of significant young ice fraction, the synthetic results presented in Sec 4.3 demonstrate that the retrieval has significant problems reproducing the true CLWP. The synthetic retrieval assumes ideal circumstances where the model representation is fully accurate, but it is still not able to predict the correct CLWP. The RMSE is dominated by a significant bias, where CLWP is underestimated by almost 50% in all cases. This is not consistent with the statement that “Generally, the retrieval is able to reproduce the real CLWP” (L358). For me, this is a key finding of the study and deserves further detailed analysis. The

authors attribute the large absolute bias at higher CLWP values to the use of a cloud-free prior and that the larger CLWP values are multiples of the assumed standard deviation. This could be tested by performing synthetic retrievals with different prior values or larger standard deviations. I would also recommend testing a retrieval of log(CLWP) (as suggested in L224), as the current assumption of a strongly truncated zero-mean Gaussian distribution is not consistent with the mathematic formalism of the OEM, which assumes the retrieval parameters have a true Gaussian distribution. The ability, or otherwise, of the retrieval to distinguish between surface and cloud parameters is also important and deserves more attention and prominence. The substance of Appendix C, particularly the strong correlation between CLWP and wind slab correlation length, should be included in the main text of the paper. It would also be instructive to show the signature in brightness temperature space of each of the retrieval parameters, in a similar way to the plot in Fig. 6(d), i.e. show the change in simulated brightness temperature for a perturbation in each of the retrieval parameters by the assumed standard deviation. This will emphasize which parameters give similar brightness temperature sensitivity, and hence cannot be easily distinguished by the retrieval. The impact of these ambiguities should be discussed in detail.

The paper should also give a more balanced discussion of the retrieval performance as shown by the case studies, and the implications for our understanding of arctic clouds. For example, Case 1 appears to show that it has little value for observing low-level stratiform cloud since most of the CLWP is below the detectability threshold and there is a significant amount of false detection. It would be helpful to discuss the retrieval performance in the context of the expected climatology of liquid cloud in polar regions. The paper references existing studies using ship-based measurements (e.g. Walbröl et al., 2022) that could be used to provide this context.

Minor comments

L40: replace “largely” with “significantly”

L117: replace “in the lowest about 100m” with “in approximately the lowest 100m”

L119: replace “with about” with “at approximately”

L130: How could you identify scenes which might be affected by frozen hydrometeor scattering in the absence of co-located radar?

Sec 3.1: I would like a clearer explanation in this section of the difference between “state”, “model” and “fixed” parameters.

Sec 3.2: Include reference for OEM method, e.g. INVERSE METHODS FOR ATMOSPHERIC SOUNDING: THEORY AND PRACTICE, Rodgers (2000) p85-86

L222: What is N? I can’t see where it is defined.

L247: If I add together the prior values of wind slab and depth hoar thickness in Table 2 I get 28cm, not 38cm as stated here. Please explain this difference.

Fig 3: I found it slightly confusing that CLWP is listed as an input in R1 (clear-sky retrieval). Perhaps indicate on the figure that it is fixed at 0.

L230: Is it possible to identify the presence of wet snow from the passive microwave observations?

Table 2: Caption – Presumably the standard deviation is the square root of the diagonal of the covariance matrix, since the diagonal should be the variance.

Sec. 3.5 Synthetic retrieval setup: I would prefer to see this information in Sec 4.3 where the synthetic retrieval is discussed.

Fig 4: The legend is confusing, with the open rectangles representing the a-priori departures. Please replace these with simple lines to match what appears in the plots.

L351: I do not understand how Fig 6(c) shows a correlation between thin ice fraction and falsely detected CLWP. The figure appears to show falsely detected CLWP appearing at all values of thin ice fraction. Please explain further.

L359: A bias of 50% does not support the statement that the retrieval can reproduce the real CLWP.

L386: I cannot clearly see a cirrus layer in the radar or lidar plots in Fig. 8. Perhaps the colour scales need to be adjusted.

Fig 8: In panel (b1) the cross-track data does not add much information and makes it harder to see the magnitude of the IR temperature. Perhaps consider using a line plot of the KT-19 brightness temperature, rather than the imager data?

Fig 8 (and similar): What does the shaded orange region represent in the state space plots?

L428: Even though the air-snow interface temperature shows a similar magnitude to the KT-19 data in the clear sky region there is a different spatial trend, suggesting that the retrieval is not capturing the true behaviour.

L433: How do you know the liquid cloud signal is well represented? It's in significant disagreement with ERA5, and as mentioned is correlated with a strong decrease in wind-slab correlation length which causes strong ambiguity with the CLWP.

L480: Why not average the HAMP retrievals to a resolution of 31km to more fairly compare with ERA5?

L517-522: The study does not discuss using surface property information from a thermodynamic model, so this statement is highly speculative. The forward model used in the study does not include the effect of rain on snow, surface melt or refreezing so it is not clear how the microwave signals due to these changes could be used.