Review of ‘Improved water vapour retrieval from AMSU-B/MHS in polar regions’ by Triana-Gomez et al.

The authors present an update to a water vapour retrieval suitable for the polar regions, expanded to a more modern sensor (MHS) and now including a screening procedure intended to mitigate the deleterious effects of significant scattering from clouds on the retrieval. The paper presents itself as describing these two ‘advances’ to the previous retrieval, ‘intended as groundwork’ for a future planned combined product that would incorporate oceanic microwave imager retrievals as well, combining the two into a pan-Arctic product that could potentially cover a long time period. While the work is nicely presented and well written, I do not believe that this rises to the level of significance to the community that merits publication in this journal. For that reason, I recommend rejection of the manuscript as it stands, with encouraged resubmission if the authors follow through on the stated future work. More in-depth comments follow, broken up into a few major bullet points and minor comments.

1. The abstract lays out the paper as presenting two advances to an old and established retrieval of Arctic water vapour. The first is simply expanding the old retrieval to a new sensor, which is almost a trivial exercise since the channels are almost identical and the difference of absorption characteristics between the 150 vs. 157GHz channels is fairly trivial. It is nice to see that the retrieval works similarly for MHS as it does AMSU-B, and this is well laid out by the authors, but it is not surprising or noteworthy for the community that reads AMT -- perhaps a small technical challenge but not scientifically significant. The abstract’s second advance touted is a new screening for artefacts caused by convective clouds. While this is potentially quite interesting, it is essentially a footnote in Section 2 of the manuscript, and the description of this new filtering method is literally restricted to 5 lines of the total manuscript (P6 L23-27). Furthermore, I am not convinced that it is necessarily filtering out ‘high cloud ice content in convective clouds’ as the abstract states; rather the authors infer that the low retrieved TWV is indicative of a scattering signal from cloud ice, but this is not demonstrated in the paper and thus it appears that it is just an assumption. It could be justified as such if compared to other satellite imagery or an IWP product. The remainder of the paper holds no strong conclusions: ‘The improved retrieval performs better when compared to another satellite product and to in situ data’ (P9 L32) and ‘the results are satisfactory’ (P10 L5). It is unclear what exactly is demonstrably better as much of the discussion is qualitative, or even that the updated retrieval could outperform reanalysis datasets, which is an almost necessary test for retrievals to demonstrate.
2. The second major comment has more to do with the methodology upon which the study rests. This is a subjective opinion, but so-called ‘non-physical’ retrievals such as Miao et al. (2001) were quite important twenty years ago when radiative transfer codes were slower and less advanced, but are becoming less useful today. The results show the downsides to such a regression-based bin method, with big gaps visible in Fig. 1. Why would the community use such a product when reanalyses have no such gaps in coverage or artefacts between bins, not to mention blended TWV products that exist

too? Modeling of sea ice emissivity is of course still a big challenge, but physically-based retrievals from microwave radiometers already exist over sea ice and indeed all surfaces (for example, see the MIRS retrieval from NOAA: https://www.star.nesdis.noaa.gov/mirs/geonwp.php). If regression-based retrievals such as the one presented are to remain relevant, they need to demonstrate their worth relative to similar products, including reanalyses (see e.g. Duncan and Kummerow 2016). If this paper had presented the validation against in-situ sources alongside comparison with say ERA5 data and shown that it outperforms the reanalysis, then it is of much more interest to the community. Even the proposed combined TWV product of this retrieval with RSS data (P10 L13) would need to prove this, and it is as of yet far from certain; the bin-based artefacts are a major concern and merging with RSS would be difficult in itself due to their own biases and simplifying assumptions made for radiative transfer. If the methodology can be shown to outperform physically-based retrievals (with a full forward model) then it has interest for the community, but otherwise it strikes me as requiring corrections on top of corrections that do not lead to greater physical understanding, and could be perhaps be better accomplished by a neural net retrieval.

3. The radiative transfer equation upon which the methodology rests struck me as maybe being incorrect (Eq. 1). If we take the case of surface emissivity of 1, then TB is directly proportional to surface temperature; if we take a fully opaque atmosphere with negligible transmittance (tau>>1), then again the second term goes to zero and TB is again directly proportional to Ts; if surface emissivity were zero, then TB is essentially Ts minus an atmospheric contribution? I apologise if I am misinterpreting this, but it makes no sense to me when I consider these cases. However, it is indeed the exact same equation given in Miao et al. (2001) and originally in Guissard and Sobieski (1994), so I am perplexed. I did not have the time to follow the full derivation in the G&S 1994 paper, but it seems suspect to me. I would suggest examining this in detail to make sure this isn't a typo, because it appears like a form given in Grody (1976) but with Ts and To flipped. Again, apologies if I have misinterpreted this--it just struck me as odd.

Minor comments:
P1 L12: The title uses 'polar' but the paper almost exclusively uses 'Arctic' only. Unless there is some focus on the Southern Hemisphere too the title should be reconsidered.
P2 L1: Is 1m squared a typo?
P2 L10: Fix citation Bobylev and Mitnik
P2 L16: According to OSCAR SSM/T2 confusingly stands for Special Sensor Microwave Humidity (https://www.wmo-sat.info/oscar/instruments/view/535)
P2 L20-21: Is there proof of this statement? A citation or elaboration would be good here.
P3 L10: Is a table with launch dates necessary? It does not really impact the paper.
P3 L16: Typo in citation, Sobieski
P4 L15: What are the units on k? Since absorption coefficients for water vapour are very well known, the derived regression parameters C could be compared against values in the literature.

P5 L17: Perhaps I missed this, but does the manuscript state how the ‘surface types are obtained’? This is a key part of the algorithm and surely any future combined product. There is something at P7 L25, but it is unclear if this is how the algorithm functions or if that was just for that particular analysis.
Section 3.1: How are coincident points defined?
P7 L9: Is there any justification for saying that time differences are ‘likely’ the cause of differences, or is this speculation?
P7 L12: What was the ‘expected amount of data’? I found this confusing.
P7 L19 It would be interesting to investigate why there is this ‘low agreement in summer’ rather than just to ‘presume’ -- this could possibly be tested by contrasting open water with retrievals over ice.
P8 L5: I don’t understand this -- you eliminated the outliers from the analysis and then found that there was good agreement? What was the justification for eliminating the outliers?
P8 L30: The bias values should be smaller than RMSD by definition.
P13 L5: Typo ‘Anctarctica’
Fig. 10: I really like the colour scale used, but it seems insufficient for the July panels. Suggest using separate colour scales, one for each season so that patterns over sea ice can be seen in both seasons.
Fig. 12: Some discussion of the third row here seems necessary. Surely it’s not physical to expect TWV=14 or more in the southern Hudson bay with TWV<3 just to the south even after screening?