Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC1
https://doi.org/10.5194/amt-2022-2-RC1, 2022

Comment on amt-2022-2

Anonymous Referee #1

Referee comment on "Retrieval of ice water path from the FY-3B MWHS polarimetric measurements based on deep neural network" by Wenyu Wang et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-2-RC1, 2022

General comments

The core idea of the manuscript, to use polarized microwave observations fromthe MWHS sensor to retrieve ice water path, is certainly of scientific interest given that these observation have not been used for this purpose before. In its current form, however, the manuscript lacks novelty and scientific rigor.

My main criticism is that the authors do very little to tie their results to any reference data, which hampers the credibility of the presented retrieval results. Although they provide a comparison of the global distributions of monthly mean IWP to ERA5, MODIS and the 2C-ICE product, the latter of which is used as training data for the retrieval, I consider these results insufficient to conclude that the retrieval works reliably given that retrieval artifacts are clearly visible over the Tibetan plateau for the winter time retrievals.

While I consider the topic fit for publication, major revision will be required to improve quality and relevance of the presented results.

Specific comments

- Fig. 4 and 5: I would suggest analyzing only observations from the swath edge or to separate the analysis of observations from edge and center of the swath. This will make it easier to compare your results to observations from conical scanners. I also suspect the scatter plot is misleading here as many markers are likely lying on top of each

other. I suggest replacing the scatter plot with a density plot. Information on the hydrometeor content can be added by drawing a contour plot of the distribution of pixels with
$IWP > 10^3$ or 10^2 on top of the density plot (with a sufficiently different colormap of course).

- l. 177: You cannot reference a results that you haven't yet presented.
- Fig. 5: Please elaborate what causes the low TBs for clear sky observations.
- Sec. 4.2: The presentation of the evaluation needs to be improved. You need to be clearer about what data is used to calculate the error for the cases you present. In particular, you need to state whether for Cases 1 - 5 the changes to the input data are also applied to the validation data. If that is the case, these error metrics have little meaning as the distribution they are calculated over cannot be expected to represent that of real measurements.
- l. 236: Define relative error. I assume you are referring to the absolute percentage error here. Note that when the absolute percentage error is used to select between models it is biased towards models that underestimate the true value. This together with the fact that you are using MSE of log(IWP) to train your network, this will likely lead to non-negligible biases in your retrieval. You should therefore also add bias to Tab. 2. I would also suggest adding correlation as an additional metric.
- l. 236: It does not make sense to include the errors over land and ocean here as this is nothing that you can tune. Please move this analysis to the end of this subsection and perform it for the final retrieval configuration.
- l. 263: This is only true if your training data set is too small to include cases with PD from the surface. Otherwise the network can easily learn to handle ambiguous inputs given that it is trained properly.
- l. 268: While the scatter plot is useful here it is insufficient to fully characterize the retrieval. For this you should apply your full retrieval, i.e. the combination cloud classification and IWP calculation, to all pixels from January 2015. Please provide a table containing at least bias, MSE, correlation and potentially the relative error calculated for IWP > 100 g/m^2. Here you can then also assess performance over land and ocean.
- l. 270: This error propagation doesn't make sense. Even if your relative errors would follow a Gaussian distribution your mean relative error wouldn't be an estimator of its standard deviation. This is even less the case for the median absolute error.
- Sec. 4.3.1: Although retrievals of cyclones are certainly scientifically interesting, the retrieval results that you provide are not very meaningful as they can't be tied to any reference value. I suggest you try to find a co-located overpass of both CloudSat and MWHS. There's a large number of CloudSat Cyclone overpasses available from https://adelaide.cira.colostate.edu/tc/tcs-50km.txt, it should be possible to find one that coincides with an overpass from MWHS within 30 minutes or so. This would allow you to compare your retrieval results to both 2C-ICE as well as the MODIS retrievals andthus add more credibility to the results presented in Sec. 4.3.2.
- Sec. 4.3.2: Please add a figure with the distribution of the zonal mean IWP similar to Fig. 3 in Duncan and Eriksson, 2018. This will allow for a more quantitative evaluation of the retrieval results. I also think a logarithmic color scale (as in Duncan and Eriksson 2018) would be more suitable to display global distributions of IWP in Fig. 15 and Fig 16.

Technical corrections

- l. 4: Missing space after 'Information'

- l. 29 - 31: You cannot conclude that individual measurements can only sense certain properties of clouds only based on their sensitivity to microphysics.
- l. 59: ICI will only have channels up to 668 GHz
- l. 62 - 64: How has MWHS been 'proven to give information about IWP' if it was 'hardly analyzed information
  past studies'?
- l. 68: The name is Cloud ProfilING Radar
- Fig. 2 (a): Are there really gaps in the distribution or is that an artifact of the bin boundaries? If the former please explain what could
  cause them. In the case of the latter please select the bin boundaries to avoid them.
- l. 167: Please typeset unit according to manuscript preparation guidelines.
- l. 170: Figure should be abbreviated with Fig. except at the beginning of a sentence.
- l. 222: Please also provide false alarm rate and probability of detection since accuracy alone can be misleading for imbalanced datasets.
- l. 289: Specify channel in which the low TB are observed
- l. 325: showed -> shown
- l. 341: The different measurement resolution of Modis and MWHS cannot affect the retrieved mean on a 5x5 degree grid.
- l. 344: Given that there are obvious artifacts in the retrieval results, I don't think that this can be concluded.
- l. 365: This discussion of the limitations of the neural network retrieval is too superficial. First of all, once the code is written extracting more co-locations is extremely easy, so I don't think there is a valid excuse to use a training data set that oneself deems too small. Moreover, although there are uncertainties related to the co-locations of the CloudSat and MWHS observations, these uncertainties are represented in the training data and can thus be predicted using for example quantile regression neural networks. The real issue are the uncertainties in the 2C-ICE data as these are much harder to quantify and cannot be predicted.
- l. 385: You cannot conclude that performance is good for the Cyclone cases because you don't have any reference to compare to.