

# Response to comments from Anonymous Referee #1

July 17, 2024

We thank the reviewer for taking the time to read the manuscript and provide detailed and valuable feedback.

## Specific comments

Line 24: I would leave out "now" in this sentence, as the reference is 30 years old  
Fixed.

Line 44: the authors list the MWI frequencies later in the paper, but a range (19-183) would be useful here since it specifically mentions the frequency coverage.

To avoid repeating the range (since it is mentioned later), the phrase has been edited to now say: "extending the coverage down to 18.7 GHz".

Line 72: is the overpass time known? 1:00?

The local time of the descending node is 9:30, and this information has now been included when describing ICI.

Line 89: reword "precipitation and snowfall" as both are precipitation. Suggest "liquid and frozen precipitation" or "rainfall and snowfall"

The phrase has been updated to say "liquid and frozen precipitation".

Line 170: this is done in section 3.2, but I would suggest adding information right at the beginning of this discussion about the period and coverage used, as I was asking this while reading this section. (two years, global CloudSat overpasses) and ancillary model data used

Information about the coverage, period, and the use of ERA5 ancillary data has now been included as the second paragraph in Section 3.1, since the choices were motivated by the requirements on the database listed in the previous paragraph.

Figure 2: bottom panel - would prefer frequencies rather than channel#. Should the y axis be labeled Ta rather than Tb?

Yes, the y axis should say  $T_a$ . This is fixed, and the legend and figure caption have been updated to give frequencies rather than channel number.

Line 451: suggest add "conically scanning"  
Added.

Line 461-465: Just a note that this is almost certainly the case and emphasizes the need for better understanding of emissivity at these higher frequencies. This is a good point, and certainly a limitation of such simulations. The need for better understanding of emissivities is now explicitly stated in this discussion.

Figure 5-6: would prefer labeling with center frequency rather than channel number (lots of flipping between pages)

Both figures 5 and 6 have been updated to state the channel frequency and the channel polarisation (when necessary).

Line is the same 614: Not sure if I missed it but should have a brief introduction of DARDAR somewhere

A description of DARDAR and the satellites that the product uses is now given when DARDAR data is first discussed (at the end of Section 3.3).

Line 618: There are a couple of features in the zonal mean plot that I think are worth discussing. Retrieved IWP is lower than the database for the northern midlatitudes. Perhaps a land/vegetation issue? Retrievals are conversely high for the ITCZ. Thoughts on this?

This is a good point, and we have now looked into this further. Plotting a zonal mean of the test set, i.e. the 'true' IWP corresponding to each retrieved case, leads to equally high IWP around the ITCZ. This implies that there is not a problem with the retrievals in this region, since it is present in the data too.

Approximately 10,000 scenes were randomly selected to form the test set. This is around 20% of the total number of scenes. This is not enough to perfectly reproduce the zonal mean of the full database at every latitude. However, good agreement is seen overall.

In regards to the northern mid-latitudes, the conclusion was similar. The zonal mean of the test set shows lower IWP at such latitudes, agreeing more with the retrievals. Small discrepancies still remained, but they were not as significant as seen in the comparison of the retrievals and the full database. We believe that this is likely due to generally poorer retrieval performance at mid-latitudes, due to both lower altitude clouds and potentially to increasing uncertainty in surface emissivities as snow cover increases. Since this is important to note, the above findings are now included. The test set zonal mean has not been included in the original figure to avoid the plot becoming too confusing. Instead, the findings are described in the text. Discussion of the zonal mean has also been moved into a separate paragraph to improve the clarity.

Line 784: Somewhere here it seems worth mentioning the importance of rep-

representativeness in the database for retrieving the highest and lowest values, as demonstrated by the author's plots

This is true, particularly for high cases. The importance of including high values in the database to achieve retrievals of such cases is now emphasised in the conclusion. The lower cases are influenced by more than just their representation in the database. Although retrievals of low values appear somewhat uncertain, there are in fact many of these cases in the database. Instead, retrievals of the lower values are more influenced by the sensor sensitivity. Furthermore, particularly low cases (less than  $0.1\text{g/m}^2$ ) are not retrieved. Therefore, we avoid placing emphasis on only the importance of representation in this case.