Response to Reviewers

We would like to sincerely thank the reviewers for their professional comments and helpful suggestions. We believe they help us to improve the manuscript significantly and provide many useful ideas to our work. We have revised the manuscript according to the reviewer's comments and answered the reviewer's question point by point below.

Reviewer comments are in italic blue and our responses in black, the manuscript changes are in red.

Reviewer 1

Reply to comments

- *l. 11: . . . in its retrieval.*

Reply: Thanks, revised.

- *l. 19: . . . of inputs of auxiliary variables and . . .*

Reply: Thanks, revised.

- *l. 24: ... in the midlatitudes ...*

Reply: Thanks, revised.

- *l. 91: The scanning is across the orbit, isn't it?*

Reply: Yes, the scanning of MWHS is across the orbit.

- *l. 241 . . . the training data (or collocations) are well representative.*

Reply: Thanks, revised.

- *l. 290: "So is the ocean/land mask information.". I guess you want to express that the land/ocean mask helps the retrieval distinguish hydrometeor scattering from the effect of the surface. However, that's not how I understand this sentence. Please consider reformulating it.*

Reply: Thanks. The sentence has been revised.

- *l. 306: I do not see the connection to Fig. 7. Do you mean Fig. 10?*

Reply: Thanks. The sentence has been deleted.

- *This is really a detail, but please consider using vector graphics for all line plots.*

Reply: Thanks. The vector graphics will be used and submitted separately in the final stage.

- *Fig. 7: Please add y-axis labels to all plots or at least the first plot of every row.*

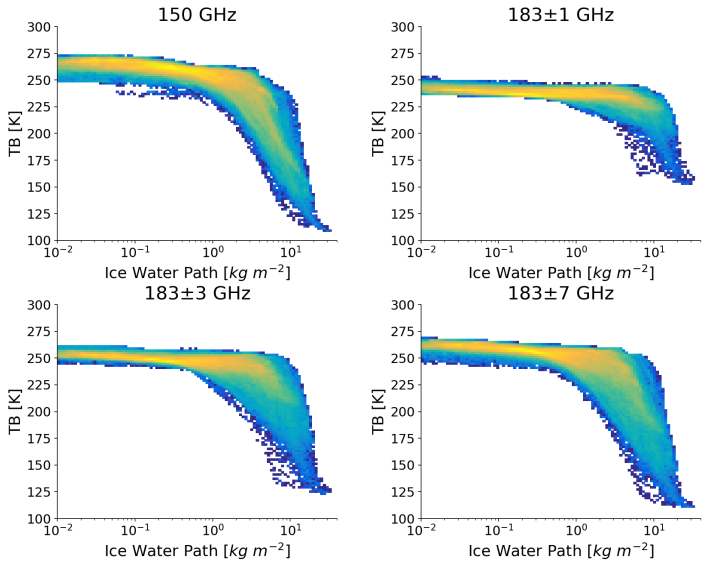
Reply: Thanks. Y-axis labels have been added.

Reviewer 2

Reply to comments

1. *The biggest issue could be found in the IWP-TB relationship in figure 4. Since the topic of this study is to perform the global IWP retrievals, the NN training database is required to cover the entire possibilities in the measurement space. I did quick forward model simulations using the mid-latitude atmosphere/cloud profiles, and the IWP-TB relationship I get is shown below. The Comparison of these two figures indicates that the collocation dataset only captures a very small fraction of the possible TB range, especially when the IWP is over 100 g/m2. Similar conclusions can be drawn by comparing figure 4 and the MWHS measurement of the tropical cyclone in figure 12. For instance, the lowest TB of 183+-1GHz channel in the center of the cyclone reaches 180K, but the smallest TB value of the same channel in the collocation database is around 230K. The NN is impossible to handle such a level of extrapolations. The global TB*

measurements used in section 4.3.2 are not given, and I believe there must be considerable amounts of TB measurements that are out of the coverage of the collocation database.



The IWP-TB relationship obtained from forward model simulations using mid-latitude atmosphere/cloud profiles.

Reply: We are very sorry that our plots caused confusion. In fact, the collocation database does cover the TB range mentioned by the reviewer. In the revised version, we have updated Figs. 4 and 5 (essentially the corresponding colorbar) for a better representation of all TB and IWP cases considered in the NN training. It can be seen that the grey area corresponds to the cases with a smaller number of samples.

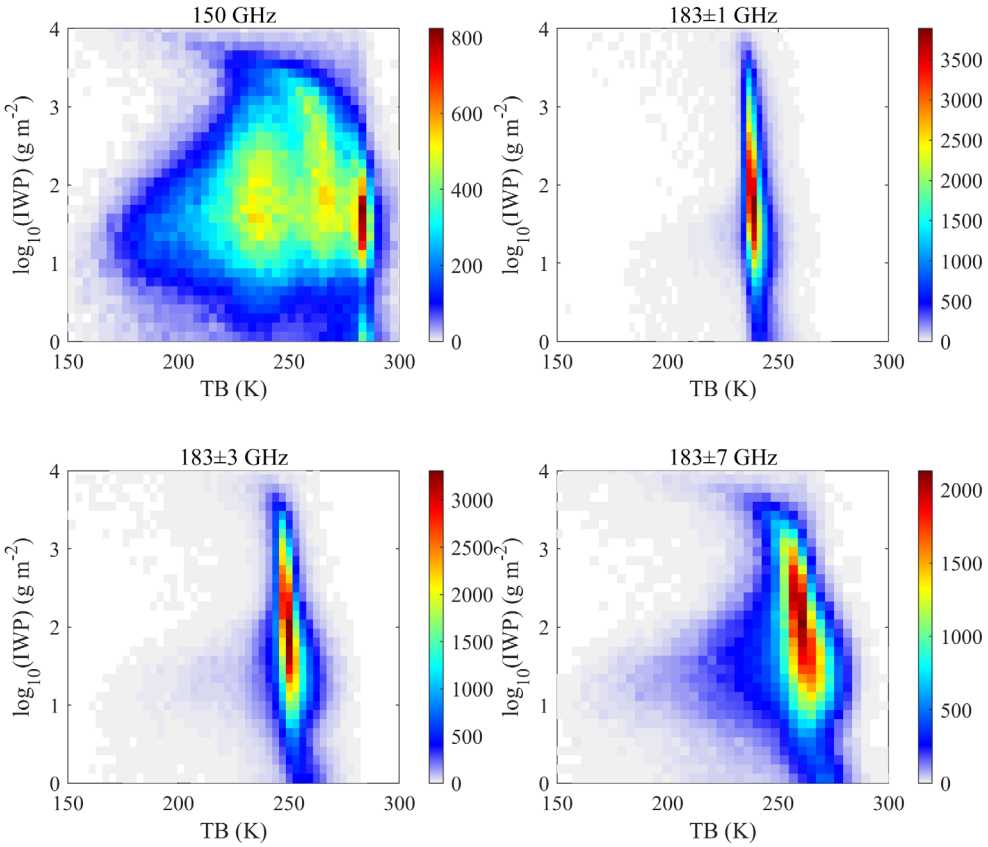


Figure 4. Statistical information of TB and IWP for different channels.

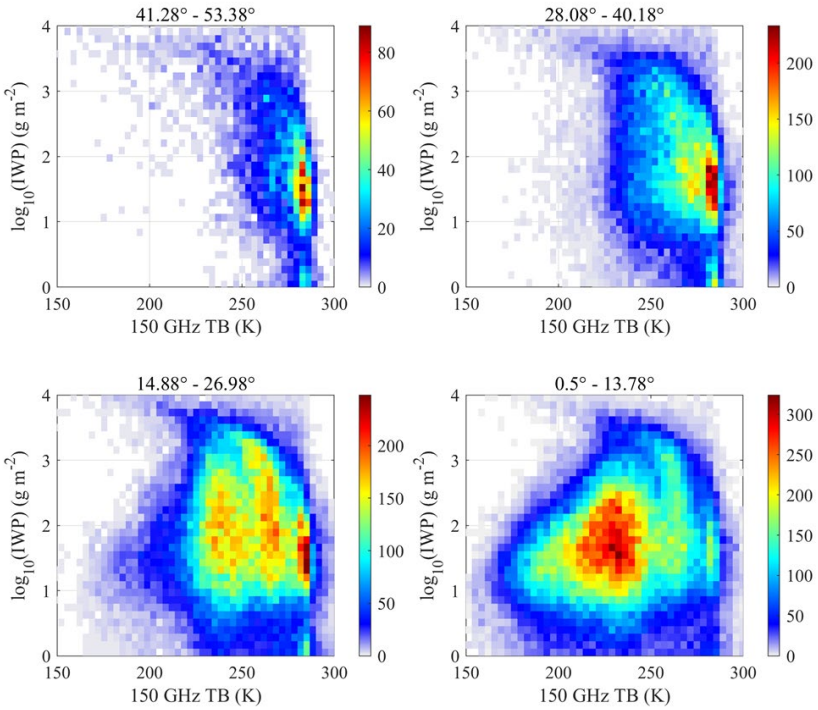


Figure 5. Statistical information of IWP and 150 GHz TB for different scan angles.

2. Figure 9 is not what I asked when I suggested investigating the sparsity of the measurement space in the last round of review. The training dataset and the validation dataset are both split from the collocation dataset, and there is no doubt they share the same statistics. What I intend to see is the comparison of TB in the training/validation database versus the cyclone and the global TB measurements applied in section 4.3. As discussed above, figure 4 shows that the collocation database is far from fully covering the TB space, and therefore we cannot expect the NN to produce sensible retrieval results.

Reply: Thanks for your suggestion. Figure 9 has been updated using the MWHS dataset in 2015 (blue) and the collocation dataset (red). Due to a large amount of MWHS data, 10^6 measurements were randomly selected for each month, i.e., a total of $1.2 \cdot 10^7$ measurements. We believe that the collocation dataset already covers the most range of the TB space and should be sufficiently representative in NN training. Nevertheless, there is no doubt that the collocation dataset is not possible to fully cover the measurement space due to the different orbit and scan methods.

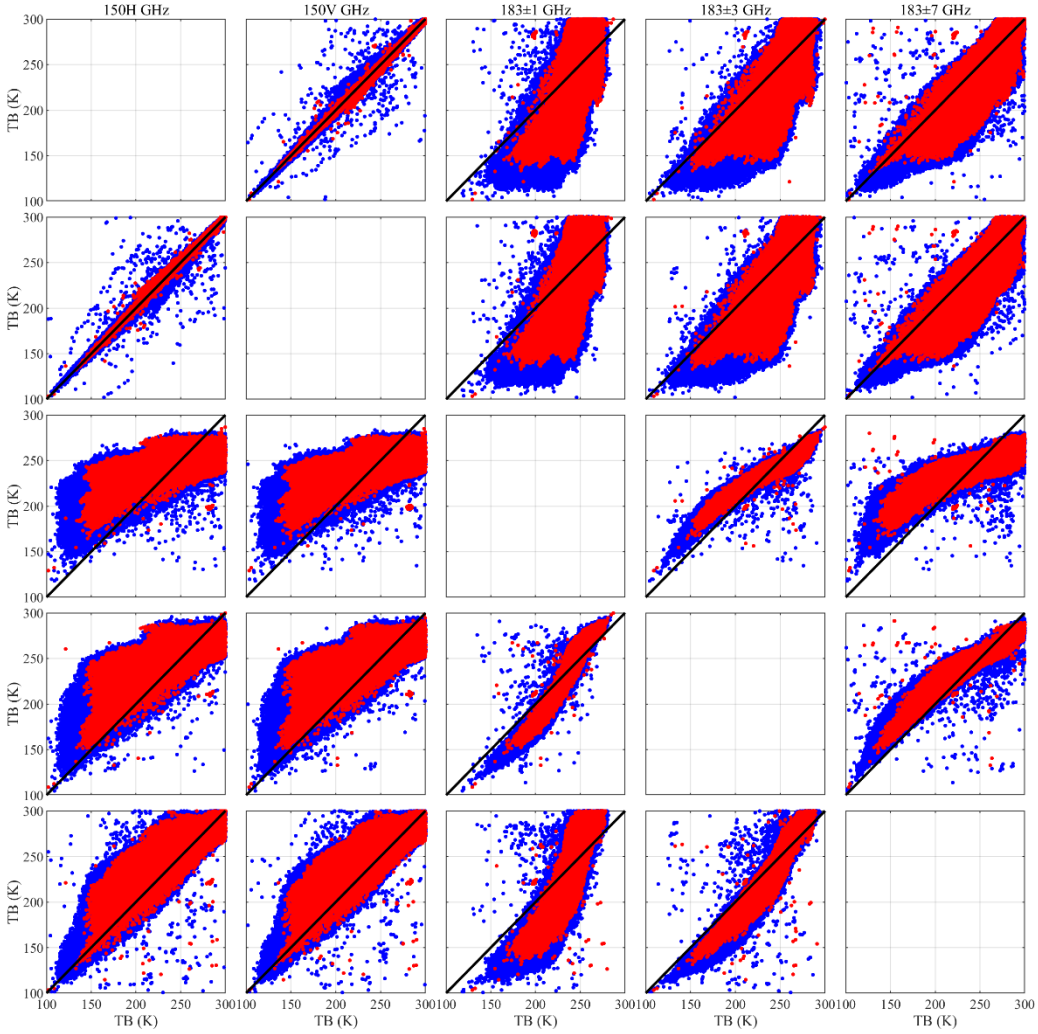
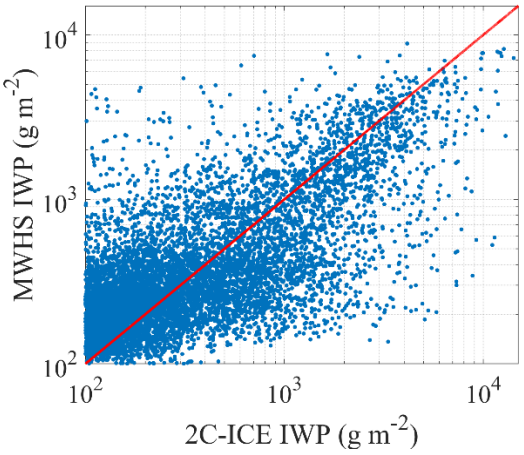


Figure 9. Measurement comparison from different channels of MWHS measurements in 2015 (blue) and collocation dataset discussed above (red).

3. Another critical issue is the retrieval experiment in figure 11, which tests the NN retrieval accuracies by comparing the retrieved parameters with the reference IWP using a testing database. The testing dataset is obtained from the same collocation finding procedure as the training/validation dataset but over a different time. Although the testing dataset is not used in the training, a well-established NN is capable to produce very accurate results since the testing and training/validation datasets have very similar statistics. However, figure 11 shows the correlations between the retrieved IWP and the reference are terrible when IWP is smaller than 1 kg/m2. The authors say this "may be due to the lack of sensitivity of the MWHS to thin ice clouds" (line

319), but even figure 4 shows that the MWHS channels are sensitive to the IWP when it is over 100 g/m2. Figure 8 in Holl et al., 2014 conducted an identical testing experiment, and the NN results they got are consistent with the 2C-ICE along the whole range, which is in line with expectations. The training/validation dataset is undoubtedly one contributing factor to the poor performance, but we cannot eliminate the possibility that the NN is not appropriately implemented.

Reply: Thanks for your comment. We believe that the result in Fig. 11 is normally interpretable and there is no problem with the collocation database or NN implementation. Although the MWHS channels are sensitive to the IWP when it is over 100 g/m^2, the ΔTB is considered to be small. Referring to Fig. 6 in Holl et al., 2014, IWP at 100 g/m^2 is where the microwave retrieval error is largest. Figure 8 in Holl et al., 2014 used the TIR channels to improve the performance in this IWP range, since the retrieval using the TIR channels performs better when IWP less than 1000 g/m^2 (see Fig. 5 in Holl et al., 2014). It should be noted that Fig. 8 in Holl et al., 2014 does not yet consider the cloud filter, whereas Figure 11 in our manuscript is the IWP retrieval after cloud filtering. Since the cloud filtering only using microwave channels does not perform satisfactorily compared to that using IR channels (used in Holl et al., 2014), the final retrieval result does not perform well at IWP of 100 g/m^2. However, the correlation of retrieval results seems better without considering the cloud filtering, see figure below.



Comparison between 2C-ICE and MWHS IWP without cloud filtering

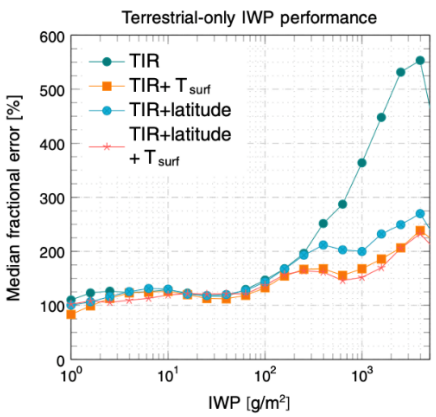


Figure 5 in Holl et al., 2014

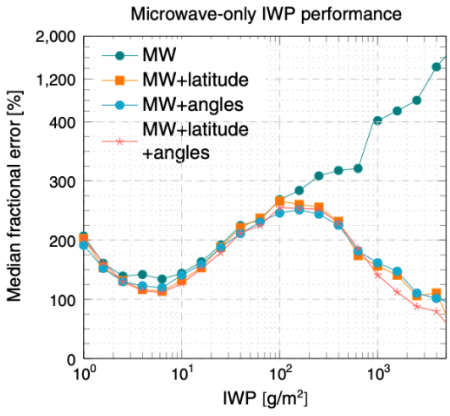


Figure 6 in Holl et al., 2014