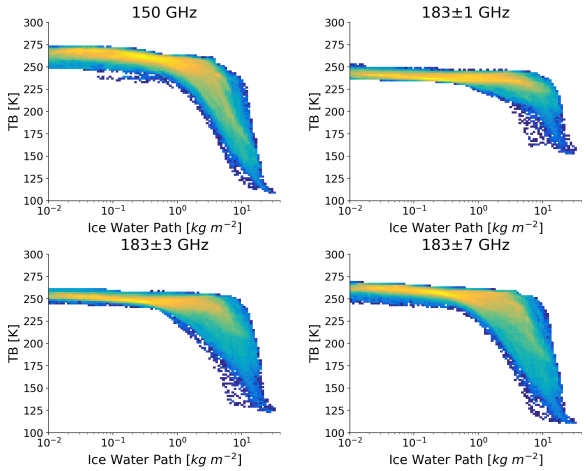
The manuscript has been revised based on reviewers' comments, and several new figures have been added. However, more devastating flaws have been exhibited through the new figures, indicating that the developed NN algorithm simply cannot do the global IWP retrievals as the authors claim.

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/m^2. 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.

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.

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/m^2. 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/m^2. 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.

In summary, I do not believe the developed NN algorithm has the capability to perform the global IWP retrievals and support the quantitative conclusions the authors claim. The idea of using collocations to train NN is great, but fundamental improvements and validations are required before this algorithm can be applied in practice.