General comments

This paper proposes an IWP retrieval method using machine learning. Combining a quantile regression neural network (QRNN) and a convolutional neural network (CNN), the measurements from Meteosat-9 SEVIRI IR-only channels can be used to retrieve IWP. The retrieval using QRNN and CNN is encouraging because the results show good agreement with the DARDAR dataset. The paper is clearly written and the study is well explained. I believe the manuscript should be published in AMT but I have a major concern and some minor comments.

My main concern is that I don't see any statistics and analysis results for the collocation database. More statistical plots of the retrieval database are suggested to provide. Since the collocation database essentially represents the prior knowledge about the ice cloud distribution, these plots could help to verify if the collocation database captures the right statistics.

Specific comments

Line 75: In general, the TIR band does not penetrate as well as microwave band and is only sensitive to signals from cloud tops, which means that for large IWPs, the IR-only measurement is likely to lack valid IWP information, especially when compared to the 94 GHz CPR of CloudSat. Therefore, the question is, if the range of IWP is not limited, are the IWPs

in the results that do not fall within the sensitive range of the IR band inferred from a priori information rather than from the measurement? Are such results reliable? Also, are the results better without constraining the range of IWP than constraining the range of IWP?

Line 130: How many DARDAR observations will there be in a SEVIRI cell in the best case? Figure 1 shows the DARDAR observations do not fill the SEVIRI cell due to the different resolution. How to illustrate that the DARDAR profile can represent the scene in the SEVIRI image?

Line 136: All data is randomly divided into training, validation and testing sets, which means that the features of the test set are also learned by the neural network. I think it is better to use a separate set of data such as data in 2012 for testing the generalization ability of the network.

Sec. 2.3: It is necessary to analyse the statistical characteristics of the distribution of the collocations with a plot. Although the relationships between IWP and visible and infrared (VISIR) radiances have been stated, it is not sufficiently visual. I suggest adding a plot here to illustrate the relationship between these parameters and IWP to show that the collocations are valid. What also needs to be illustrated is the coverage of the observations in the dataset and whether it is representative of the

majority scenarios. Also, what is the proportion of scenes with and without ice clouds in the dataset and is there a problem with uneven data distribution ?

Line 181: What is the purpose of random image mirroring and rotation?

Line 250: Does the spatial information refer to the correlation between pixels? Why it is useful for retrieval?

Line 281: In the comparison between the two products, is the instantaneous IWP retrieval using the test dataset? Why not use the data for 2012 as used in the monthly mean diurnal cycles?