

Response to Reviewers

Reviewer 1

We would like to sincerely thank the reviewer for his professional comments and helpful suggestions. We believe they help us to improve the manuscript significantly and give us 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.

Reply to General comments

The core idea of the manuscript, to use polarized microwave observations from the MWHS sensor to retrieve ice water path, is certainly of scientific interest given that these observations 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.

Reply: Thank you very much for the comments. We agree with the criticism and the manuscript has been revised and rewritten in response to reviewers' comments. The major changes in the revision are as follows:

(1) The collocation data set has been reanalyzed according to the suggestions from reviewer 1 and reviewer 2, the scatter plots were replaced by density plots; collocations in high latitudes were randomly removed to make the data set balance. (Sec. 2.2, Fig. 3-5)

(2) Most of the graphics have been revised.

(3) Network evaluation metrics were added, including FAR, POD, RMSE, MAPE, BIAS, and Pearson correlation coefficient. (Sec. 3)

(4) The networks have been re-trained using the new data set, and the classification network was redesigned using the focal loss function to reduce the effect of data imbalance. The artifacts over the Tibetan plateau have been reduced. However, the classification of ice clouds using MWHS channels is still limited, especially over the Tibetan plateau. (Sec. 4.1)

(5) Reanalysis of the new retrieval results, including the final configuration results. (Sec. 4.2)

(6) A case of tropical cyclone observed simultaneously by MWHS and CloudSat was added. (Sec. 4.3.1, Fig. 9-10)

(7) The annual mean IWP for 2015 was calculated and compared with the other three data sets. (Sec. 4.3.2, Fig. 11)

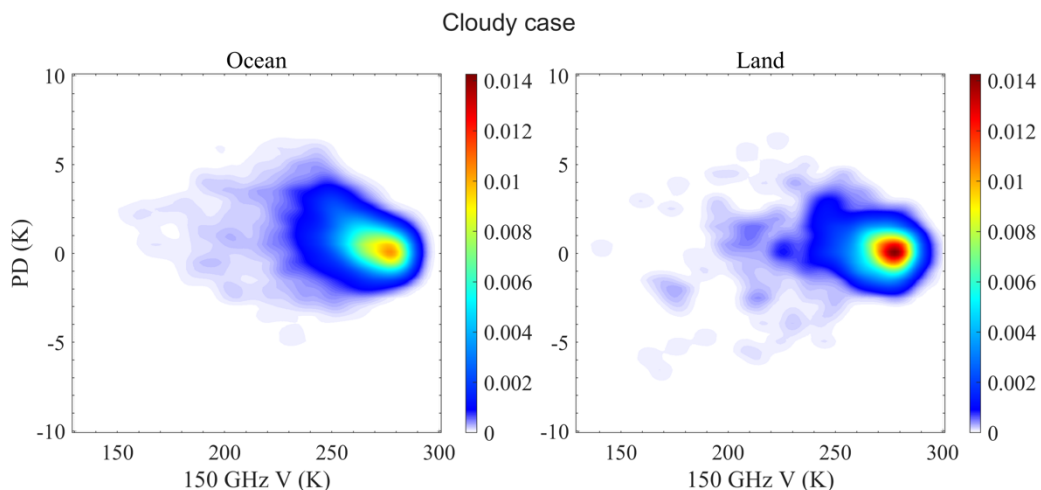
(8) Comparison of zonal means of IWP was added. (Sec. 4.3.2, Fig. 12)

Reply to 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).

Reply: Thanks for the good suggestion. The scatter plot here is indeed misleading, we have re-analysed the edge observations scan angles from ($\pm 40.15^\circ$ to $\pm 53.35^\circ$) and plotted the density. The specific changes are as follows:

(Sec. 2.2) “The density plots of the PD and TB at 150 GHz (clear-sky and cloudy scene) and the corresponding IWP from 2C-ICE over the ocean and land are depicted in Fig. 4 and 5. Scan angles from $\pm 40.15^\circ$ to $\pm 53.35^\circ$ are selected to compare the results with observations from conical scanners. In the cloudy case, the TBs are distributed between 150 K and 290 K, with the largest PD occurring at 230 K (corresponding to $IWP > 1000 \text{ g m}^{-2}$). This is similar to the result of Gong et al (2017, 2020). However, due to the cross-track scanning mode, the PD of MWHs is much lower than conical scanners. The lowest TB generally appears in the center of deep convection clouds, and the PD is small due to the randomly oriented ice particles; the largest PD due to the horizontally oriented particles generally appears in the warmer ice clouds. From Fig. 4, it can be seen that the lower the TB, the larger IWP, but the TB is also influenced by the local atmospheric temperature. Comparing Fig. 4 and Fig. 5, the TB of the clear sky is generally above 240 K. The PD from the ocean surface is relatively large, while the PD from land is small.



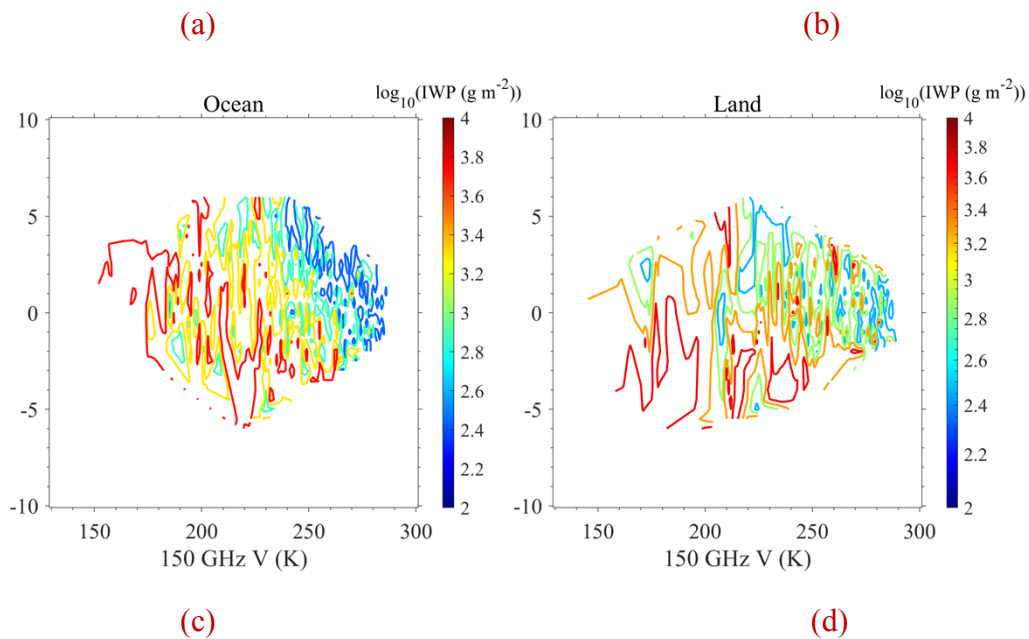


Figure 4. The PD-TB_{150V} density plots for the collocations in the cloudy scenes over the ocean (a) and land (b). The (c) and (d) show the corresponding IWP from 2C-ICE.

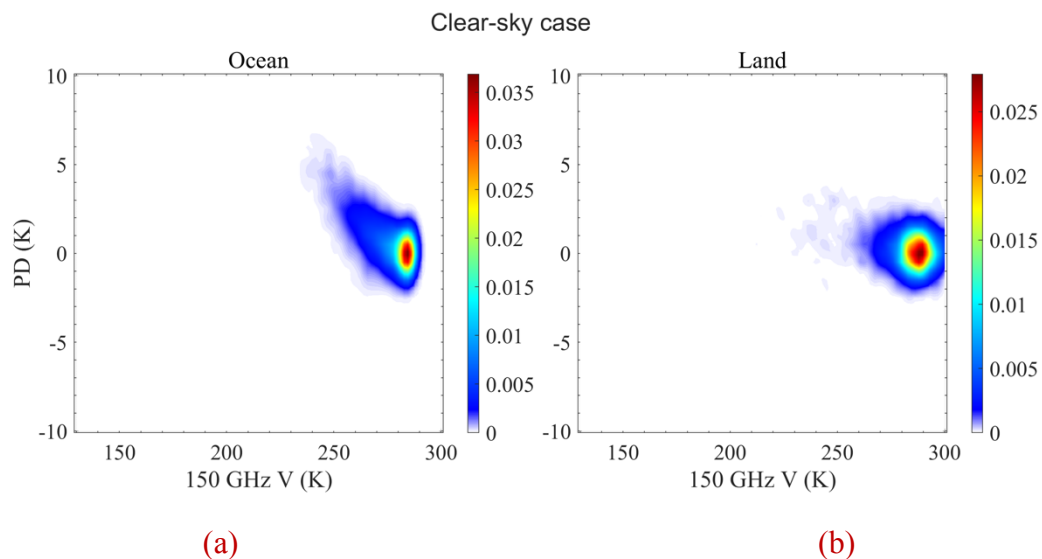


Figure 5. Same as Fig. 4 but for clear-sky scenes.”

l. 177: You cannot reference a result that you haven't yet presented.

Reply: Thank you for the comment. The sentence has been deleted.

Fig. 5: Please elaborate what causes the low TBs for clear sky observations.

Reply: Thanks for the comment. The low TBs in clear sky observations is mainly caused by low atmospheric temperature, occurring at high altitudes such as the Tibetan Plateau or in regions with atmospheric temperature anomalies. There is also a possibility that ice clouds exist outside the FOV of CloudSat (only one-tenth of the MWHS pixel) but in the FOV of MWHS.

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.

Reply: Thanks. We agree with the comment. Cases 1-5 have been deleted. The presentation is revised as follows:

“For the global IWP retrieval, clear-sky scenes were excluded from the training data. Different combinations of the network input are compared to find the best retrieval strategy. The auxiliary information cases and their retrieval errors are listed in Table 3. In these cases, five channels are all used. Additional information including latitude, scan angle and ocean/land mask and their combinations were added to train the networks.” (Sec. 4.2)

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.

Reply: Thanks for the good suggestion. The definition of errors is added to the end of Sec. 3. The network does tend to underestimate the true value. We have added RMSE, BIAS and correlation to the Tables. And we used MAE as the loss function in this version since its result is better than MSE. The changes are as follows:

(Sec. 3) “The performance metrics employed for the retrieval are defined in the following.

The commonly used binary classification metrics are chosen for the cloud filtering network. A confusion matrix M is defined as

$$M = \begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix} \quad (1)$$

TP and TN are the number of true positives (both MWHS and CloudSat find ice clouds) and negatives (both MWHS and CloudSat find no ice clouds), respectively. FP and FN are the number of false positives (MWHS finds ice clouds but CloudSat not) and negatives (CloudSat finds ice clouds but MWHS not), respectively

From the confusion matrix above, the accuracy (AC), False Alarm Ratio (FAR), Probability of Detection (POD), F1 score and Critical Success Index (CSI) can be derived as

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$FAR = \frac{FP}{TP + FP} \quad (3)$$

$$POD = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (5)$$

$$CSI = \frac{TP}{TP + FN + FP} \quad (6)$$

The performance evaluation for the IWP retrieval network is based on the RMSE, MAPE, BIAS and Pearson correlation coefficient (CC), defined as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{pred,i} - y_{valid,i})^2} \quad (7)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_{pred,i} - y_{valid,i}|}{y_{valid,i}} \times 100\% \quad (8)$$

$$BIAS = \frac{1}{N} \sum_{i=1}^N (y_{pred,i} - y_{valid,i}) \quad (9)$$

$$CC = \frac{\frac{1}{N} \sum_{i=1}^N (y_{pred,i} - \overline{y_{pred}})(y_{valid,i} - \overline{y_{valid}})}{\sigma_{pred} \sigma_{valid}} \quad (10)$$

(Sec. 4.2) For the global IWP retrieval, clear-sky scenes were excluded from the training data. Different combinations of the network input are compared to find the best retrieval strategy. The auxiliary information cases and their retrieval errors are listed in Table 3. In these cases, five channels are all used. Additional information including latitude, scan angle and ocean/land mask and their combinations were added to train the networks.

Concerning the errors shown in Table 3, a significant improvement in retrieval performance is achieved by adding latitude or ocean/land mask information while the contribution of just adding the scan angle to the retrieval is not significant. In MWS measurements, the signal from ice clouds is a reduction in TB by scattering effects. In the absence of latitude information, it is difficult to distinguish whether the decrease in TB is due to the ice particles or the low radiance from the surface or atmosphere. So is the ocean/land mask information. According to cases 1, 2, 4 in Table 3, the CC is improved from 0.50 to about 0.62, RMSE and MAPE are also improved significantly. However, MAPE and BIAS are in conflict, reducing MAPE will increase BIAS. Thus, the correlation is an important metric for evaluating the model. The combination of

auxiliaries can further improve the retrieval results, although the effect of using the scan angle alone is not obvious. Case 5 and 6 in Table 3 indicates that the scan angle combined with latitude and ocean/land mask can also further improve the retrieval capability. The retrieval MAPE of each IWP bin is shown in Fig. 7 (a). The MAPE in different IWP bins gives a more detailed comparison. Compared to no auxiliary model, adding auxiliaries can significantly reduce the retrieval errors, especially at $IWP < 200 \text{ g m}^{-2}$ and $IWP > 1000 \text{ g m}^{-2}$.

Table 3. Errors of IWP retrieval using different auxiliaries

	RMSE (g m^{-2})	MAPE (%)	BIAS (g m^{-2})	CC
1. No	1085.75	109.94	-91.09	0.50
2. Lat	943.68	84.53	-125.98	0.61
3. Ang	1020.52	106.43	-93.64	0.53
4. Mask	943.80	81.84	-126.03	0.62
5. Lat+Ang	908.59	79.88	-145.70	0.64
6. Lat+Mask	908.48	75.80	-141.02	0.64
7. Ang+Mask	895.98	78.60	-143.64	0.65
8. Lat+Ang+Mask	875.20	75.30	-117.05	0.67

The performance of the different channel combinations (all the auxiliary information is added) is presented in Table 4. Since the 183 GHz channels (CH. 3-5) of MHS have proved to have good sensitivity to CloudSat IWP, the influence of the 150 GHz channel and its PD is mainly focused here. The results of case 2 and 3 in Table 4 show that adding the 150 GHz window channel (CH. 2) give an improvement to all the metrics. Considering the contribution of PD in the retrieval, the results show that the addition of PD alone (case 4) contributes to the retrieval of IWP, while the combination including both H and V polarization channels has the best performance (case 1). Figure 7 (b) illustrate the MAPE of different channels. Comparing case 3 with case 4 in Table 4, the addition of PD gives an obvious improvement in the retrieval results at $IWP > 2000 \text{ g m}^{-2}$. This conclusion is close to the analysis in Figure 4. In general, all channels of MWHS contribute to ice cloud retrieval.”

Table 4. Errors of IWP retrieval using different channels

	RMSE (g m^{-2})	MAPE (%)	BIAS (g m^{-2})	CC
1. CH.1-5	875.20	75.30	-117.05	0.67
2. CH.2-5	901.84	76.75	-139.49	0.64
3. CH.3-5	932.29	79.34	-158.89	0.61
4. CH.3-5+PD	894.08	79.82	-134.88	0.65

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.

Reply: Thanks for the comment. It has been moved to the end of this subsection.

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.

Reply: Thanks for the comment. The sentence has been deleted.

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 $IWP > 100 \text{ g/m}^2$. Here you can then also assess performance over land and ocean.

Reply: Thanks for the good suggestion. We have added the table and analysed the full retrieval results. The performance over land and ocean is also discussed. The changes are as follows:

(Sec. 4.2) “The final retrieval models (case 1 in Table 2 and case 8 in Table 3) were selected according to the metrics. Combining the cloud filtering network and the IWP retrieval network with the test data, the final results are shown in Table 5. The performance over the ocean and land is also listed. After adding the cloud filtering network, the accuracy of IWP retrieval decreased, significantly for MAPE and BIAS, and slightly for CC and RMSE. The results are better over the ocean than over land, especially the correlation. Figure 8 shows the scatter plot between MWHS IWP and 2C-ICE IWP in January 2015. The result shows relative agreement, but MWHS IWP has significant dispersion at low IWP, which may be due to the lack of sensitivity of MWHS to thin ice clouds. The final model underestimates the true value overall but overestimates it when the IWP $< 300 \text{ g m}^{-2}$.

Table 5. Errors of the final selected models

	RMSE (g m^{-2})	MAPE (%)	BIAS (g m^{-2})	CC
Final model	916.76	92.90	-213.12	0.65
Land	942.81	92.56	-260.47	0.55
Ocean	908.20	92.76	-196.79	0.69

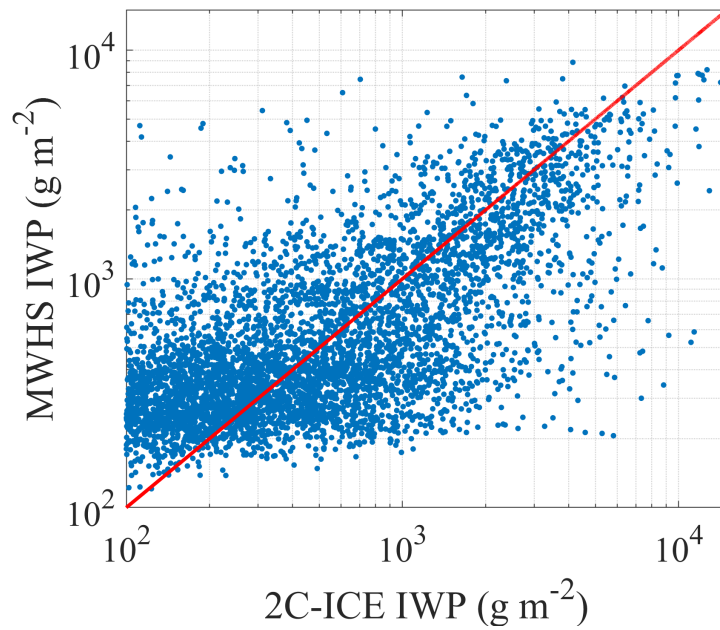


Figure 8. Comparison between 2C-ICE and MWHs IWP. The red line represents the diagonal 1:1 line. Clear-sky scenes are not shown.”

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.

Reply: Thank you for the comment. It has been deleted.

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 and thus add more credibility to the results presented in Sec. 4.3.2.

Reply: Thanks for the good suggestion. A tropical cyclone Bansi is found on January 12, 2015 with the collocation time of about 3 minutes. The changes are as follows:

(Sec. 4.3.1) “A tropical cyclone Bansi observed by MWHs and CloudSat simultaneously (the time difference is about 3 minutes) on 12 January 2015 is selected for the validation of the final networks. MWHs observed TBs of the cyclone are manifested in Fig. 9. Quite low TB (as low as 150 K) can be found at 150 GHz and 183-7 GHz channels in the regions of the eyewall (the eye is not seen) and spiral rain bands which are mainly caused by the scattering of ice particles in the clouds. The 183-1 GHz and 183-3 GHz channels are strongly influenced by water vapor, the shape of the cyclone is not observable, but clear low TBs can still be seen in the eyewall and rainband. The PDs at 150 GHz ($TB_V - TB_H$), their distribution characteristics are the same

as the low TBs. The PD reaches its maximum in the anvil precipitation regions (around 5 K, consistent with the result in Fig. 4) and decrease in the remote clear-sky or cirrus regions.

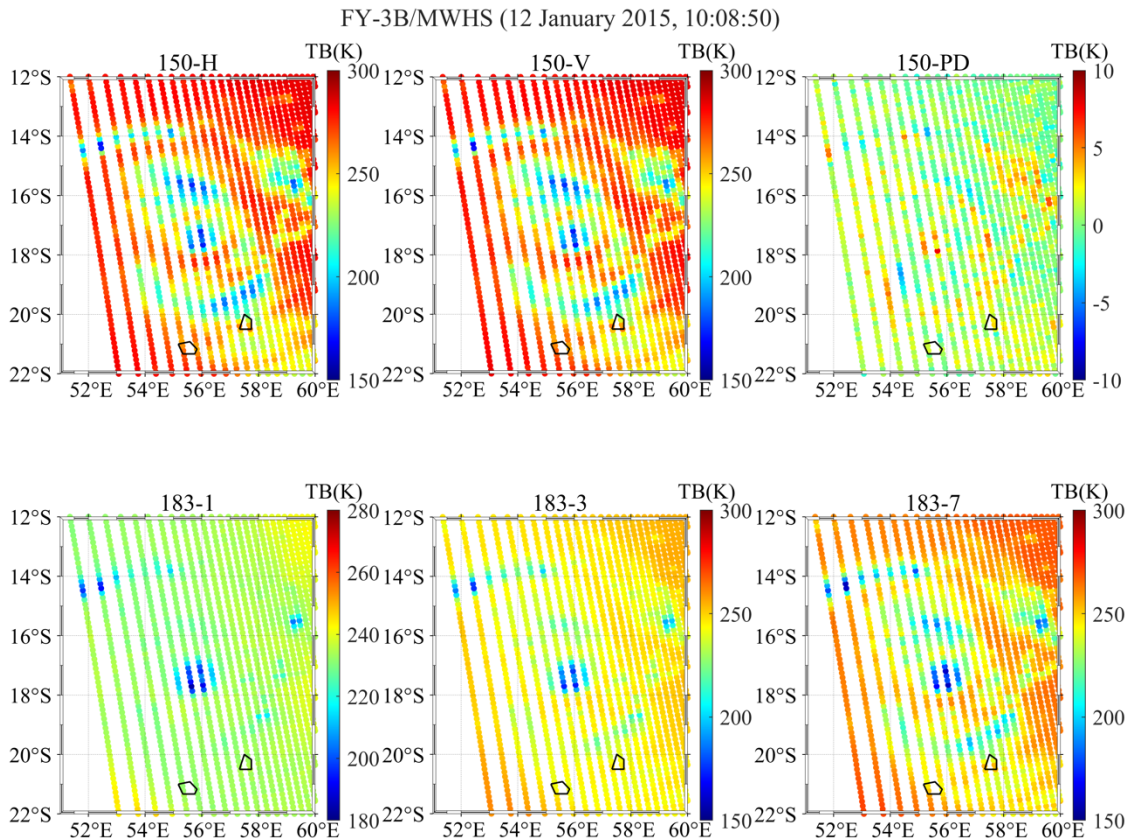


Figure 9. Tropical cyclone Bansi on 12 January 2015 as observed with FY-3B/MWHS channels.

Applying the two neural networks trained above to the tropical cyclone, the retrieval IWPs are shown in Fig. 10 in comparison with 2C-ICE, and the retrieval errors are listed in Table 6. Due to the narrow field of view of CloudSat, a total of 21 pixels of MWHS are collocated in the tropical cyclone region. The results show that MWHS IWP has a high correlation with 2C-ICE, the MAPE and BIAS are better than that in Table 5, although the RMSE is larger, it is reasonable in tropical cyclones.

Table 6. Errors of the tropical cyclone retrieval				
	RMSE (g m ⁻²)	MAPE (%)	BIAS (g m ⁻²)	CC
Bansi	1191.3	77.69	82.07	0.73

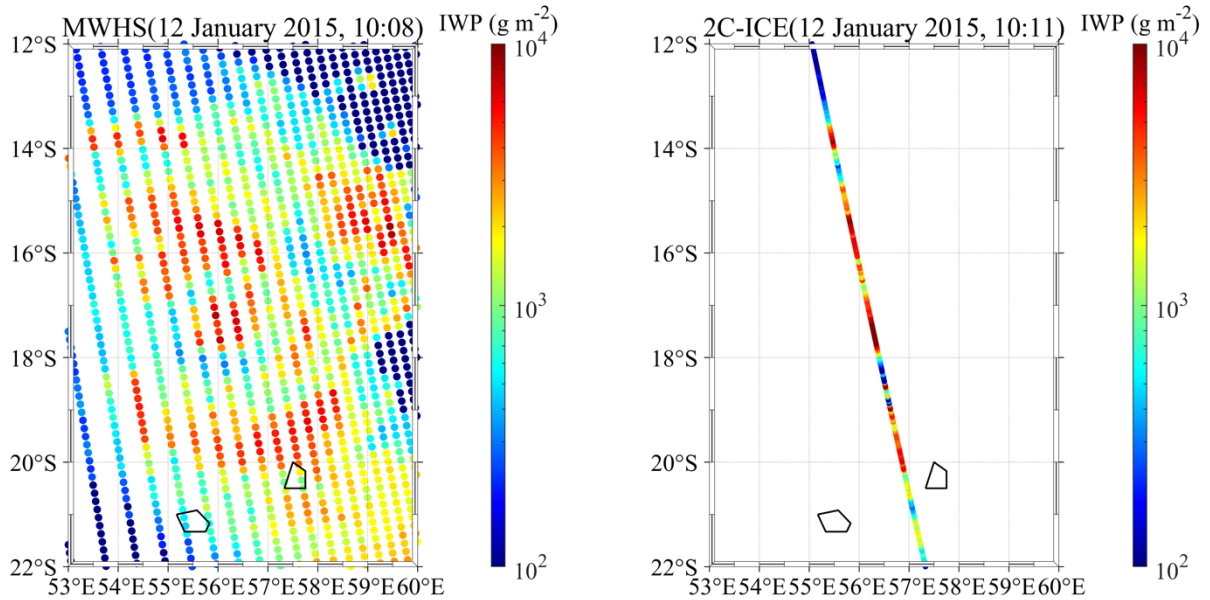


Figure 10. IWP comparison of MWHS and 2C-ICE at the tropical cyclone Bansi.”

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.

Reply: Thanks for the good suggestion. The changes are as follows:

(Sec. 4.3.2) “Figure 11 shows the global mean IWP for 2015 from Aqua/MODIS L3 product (MYD08_M3, C61, Platnick et al., 2017), CloudSat 2C-ICE, FY-3B/MWHS retrieval and ERA5 reanalysis data set. ERA5 IWP data shown here is combined of its total column snow water (CSW) and cloud ice water (CIW) data since it differentiates between precipitating and non-precipitating ice. The overall distribution of the annual mean IWP for the four data sets is similar. The MODIS product has a significantly higher IWP than the other three products, while the ERA5 has a lower IWP overall. IWP from 2C-ICE is the same as MODIS near the equator and between ERA5 and MODIS elsewhere. Since 2C-ICE is used to train the networks, MWHS IWP is certainly approaching the 2C-ICE. The zonal means of IWP for 2015 are given in Fig. 12. The overall shape of the IWP zonal averages is fairly consistent across data sets. However, there are large differences in the overall magnitude of the IWP. These differences are particularly pronounced at mid-latitudes, especially between the MODIS product and the other three products. Compared to the IWP maps in Duncan and Eriksson (2018), this version of MODIS IWP is more similar to 2C-ICE near the equator (10°S - 10°N), but with increasing latitude, the IWP is much larger than the other products. The MWHS IWP is very close to that of 2C-ICE but lower than 2C-ICE in the mid-latitudes of the southern hemisphere. This may be due to the lack of training data in the middle and high latitudes of the southern hemisphere.

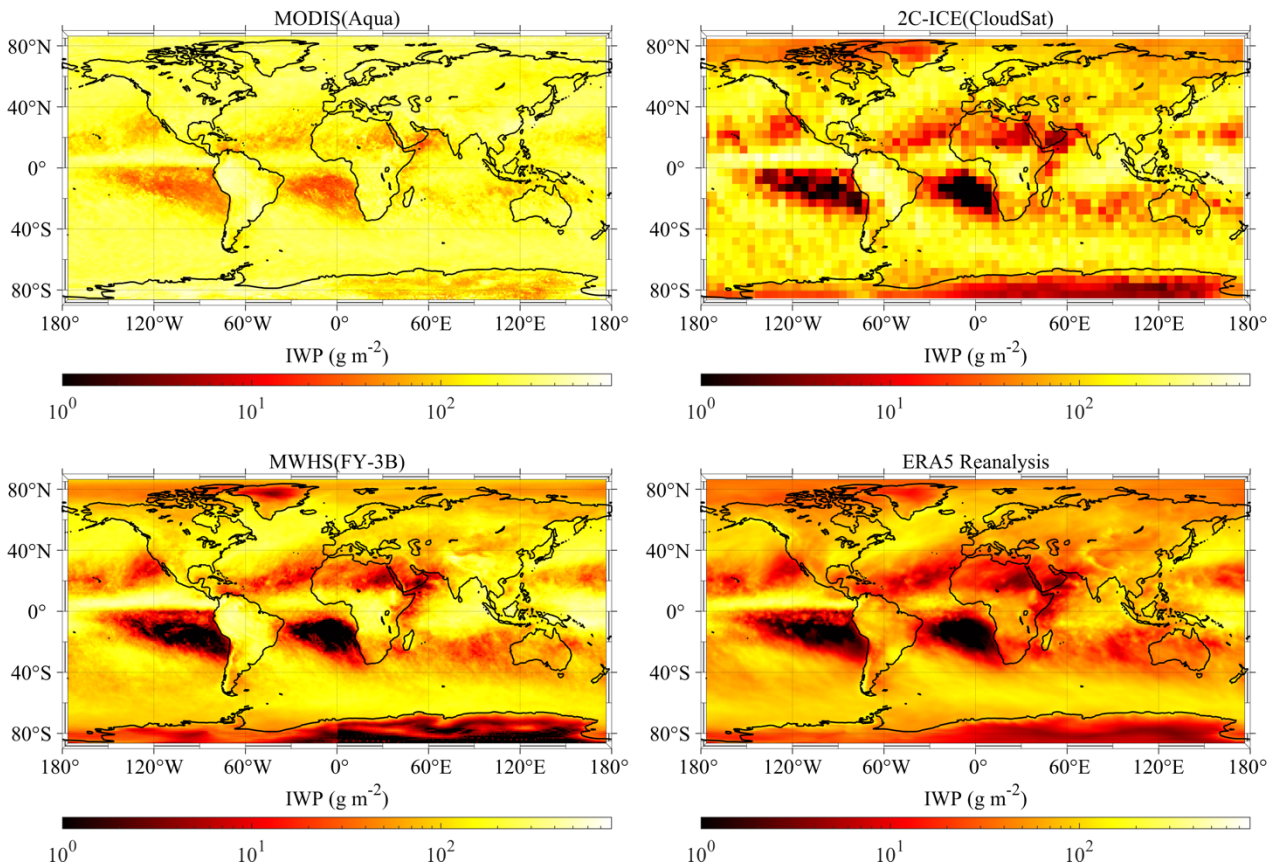


Figure 11. Global mean IWP maps for 2015 from MODIS, 2C-ICE, MWHS and ERA5. 2C-ICE is gridded on a 5° grid, while the other products are gridded on a 1° grid.

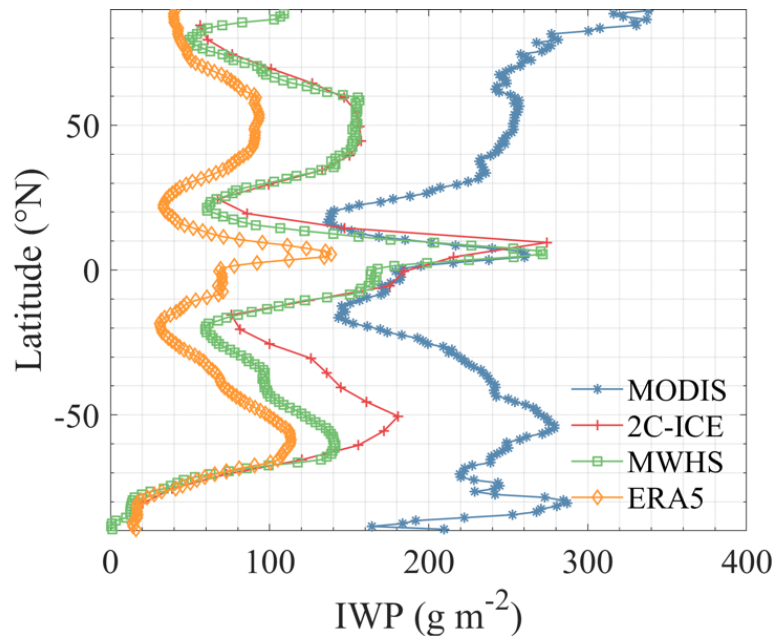


Figure 12. Zonal means of IWP for 2015.”

Technical corrections

l. 4: Missing space after 'Information'

Reply: Thanks for the comment. It has been corrected.

l. 29 - 31: You cannot conclude that individual measurements can only sense certain properties of clouds only based on their sensitivity to microphysics.

Reply: Thanks for the comment. The sentence has been deleted.

l. 59: ICI will only have channels up to 668 GHz

Reply: Thanks for the comment. It has been corrected.

l. 62 - 64: How has MWHS been 'proven to give information about IWP' if it was 'hardly analyzed information past studies'?

Reply: Thanks for the comment. The reference (He and Zhang, 2016) use the 89 and 150 GHz channels of MWHS-I and II onboard FY-3B/C to analyze the distribution of cirrus over the Tibetan plateau. However, the study did not conduct a more detailed analysis.

l. 68: The name is Cloud ProfilING Radar

Reply: Thanks for the comment. It has been corrected.

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.

Reply: Thanks for the comment. This is an artifact of the bin boundaries and the figure has been corrected.

l. 167: Please typeset unit according to manuscript preparation guidelines.

Reply: Thanks for the comment. The manuscript has been revised according to the guidelines.

l. 170: Figure should be abbreviated with Fig. except at the beginning of a sentence.

Reply: Thanks for the comment. It has been corrected.

l. 222: Please also provide false alarm rate and probability of detection since accuracy alone can be misleading for imbalanced datasets.

Reply: Thanks for the good suggestion. The FAR and POD have been provided in Sec. 4.1. The changes are as follows:

(Sec. 4.1) “The results show that all channels have cloud information, and CH. 4 (183-3 GHz) is the best for cloud detection. This channel is also used by the traditional method to distinguish cloudy from clear sky. However, the detection of ice clouds using MWHS channels is still limited. The FAR and POD of the best network are 0.26 and 0.63, respectively.”

Table 2. Errors of cloud filtering using different channels

	AC	FAR	POD	F1	CSI
1. CH.1-5	0.92	0.26	0.63	0.67	0.51
2. CH.2-5	0.92	0.27	0.61	0.66	0.49
3. CH.3-5	0.91	0.30	0.62	0.65	0.49
4. CH.4-5	0.91	0.29	0.59	0.64	0.48
5. CH.5	0.83	0.32	0.38	0.49	0.33

l. 289: Specify channel in which the low TB are observed

Reply: Thanks for the comment. The changes are as follows:

(Sec. 4.3.1) “Quite low TB (as low as 150 K) can be found at 150 GHz and 183-7 GHz channels in the regions of the eyewall (the eye is not seen) and spiral rain bands which are mainly caused by the scattering of ice particles in the clouds. The 183-1 GHz and 183-3 GHz channels are strongly influenced by water vapor, the shape of the cyclone is not observable, but clear low TBs can still be seen in the eyewall and rainband.”

l. 325: showed -> shown

Reply: Thanks for the comment. It has been corrected.

l. 341: The different measurement resolution of Modis and MWHS cannot affect the retrieved mean on a 5x5 degree grid.

Reply: Thanks for the comment. The previous version of the manuscript had an error in the geographical location of the MODIS map, which has been fixed.

l. 344: Given that there are obvious artifacts in the retrieval results, I don't think that this can be concluded.

Reply: Thanks for the comment. The sentence has been deleted.

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.

Reply: Thanks for the good suggestion. The changes are as follows:

(Sec. 4.4) “However, there are some limitations of using neural networks for IWP retrieval. Collocation is the first limitation since there are some uncertainties in the field of view of MWHS and CloudSat due to the large resolution difference. These uncertainties are represented in the training data and can be predicted using

for example quantile regression neural networks. The most important issue is the real sample (2C-ICE) used in training, which has uncertainties that are difficult to quantify. Therefore, it is also impossible to make accurate error estimates of the model results. In the absence of access to a large number of real samples, the use of neural networks can only converge to a certain product with the highest accuracy (such as 2C-ICE). An alternative approach is to use simulation results (typical profiles) of radiative transfer models, where the generalization ability of the network will strongly depend on the model itself and the input field. In addition, the microwave band below 200 GHz is sensitive only to large ice particles and thick clouds and is relatively less effective for cloud detection.”

l. 385: You cannot conclude that performance is good for the Cyclone cases because you don't have any reference to compare to.

Reply: Thanks for the comment. The comparison has been added.