Retrieval of ice water path from the FY-3B MWHS polarimetric measurements based on deep neural network

Wenyu Wang[1], Zhenzhan Wang[1], Qiurui He[1, 2], Lanjie Zhang[3]

[1]Key Laboratory of Microwave Remote Sensing, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China

[2]School of Information Technology, Luoyang Normal University, Luoyang 471934, China

[3]School of Information & Communication Engineering Beijing Information Science And Technology University, Beijing 100101, China

Correspondence to: Zhenzhan Wang (wangzhenzhan@mirslab.cn)

Abstract. Ice water path (IWP) is an important cloud parameter in atmospheric radiation, and there are still great difficulties in retrieval. Artificial neural network is a popular method in atmospheric remote sensing in recent years. This study presents a global IWP retrieval based on deep neural networks using the measurements from Microwave Humidity Sounder (MWHS) onboard the FengYun-3B (FY-3B) satellite. Since FY-3B/MWHS has quasi-polarization channels at 150 GHz, the effect of polarimetric radiance difference (PD) is also studied. A retrieval database is established using collocations between MWHS and CloudSat 2C-ICE. Then two types of networks are trained for cloud scene filtering and IWP retrieval, respectively. For the cloud filtering network, the microwave channels show a lack of capacity with a false alarm ratio (FAR) of 0.26 and a probability of detection (POD) of 0.63. For the IWP retrieval network, different combination inputs of auxiliaries and channels are compared. The results show that the five MWHS channels combined with scan angle, latitude, and ocean/land mask perform best. Applying the cloud filtering network and IWP retrieval network, the final root mean squared error (RMSE) is 916.76 g m^{-2}, the mean absolute percentage error (MAPE) is 92%, and the correlation coefficient is 0.65. Then a tropical cyclone case measured simultaneously by MWHS and CloudSat is chosen to test the performance of the networks, and the result shows a good correlation with 2C-ICE. Finally, the global annual mean IWP of MWHS is similar to that of MODIS, 2C-ICE and ERA5 in overall shape and very close to that of 2C-ICE in magnitude. ~~For the cloud filtering network, using IWP of 10 g/m² and 100 g/m² as the threshold show the filtering accuracy of 86.48% and 94.22% respectively. For the IWP retrieval network, different training input combinations of auxiliary information and channels are compared. The results show that the MWHS IWP retrieval performs well at IWP >100 g/m². The mean and median relative errors are 72.02% and 46.29% compared to the 2C-ICE IWP. PD shows an important impact when IWP is larger than 1000 g/m². At last, tropical cyclone case chosen to test the performance of the networks, the result show a good agreement with the characteristics of the brightness temperature observed by the satellite. The monthly MWHS IWP shows a good consistency compared to the ERA5 and 2C-ICE while it is lower than MODIS IWP.~~

1 Introduction

Ice clouds play an important role in the global climate (Liou, 1986), and their distribution strongly affects precipitation and the water cycle (Eliasson et al., 2011; Field and Heymsfield, 2015). Long time series and global observations of ice clouds are essential for understanding the Earth's climate system. Depending on the wavelength of observation, satellite remote sensing can measure different cloud microphysics. Microwave measurement can penetrate deeper into cloud layers to measure thick and dense ice clouds, while infrared and visible instruments are mainly used for thin clouds measurement around the cloud-top (Liu and Curry, 1998; Weng and Grody, 2000; Stubenrauch et al., 2013). ~~Since the measurements depend on microphysical properties of cloud particles (shape, size, concentrations), the individual instrument can only be sensitive to partly information of clouds.~~ Although the ice water path (IWP) obtained from different instruments show several folds of differences (Stephens and Kummerow, 2007; Wu et al., 2009), it is of great importance to use remote sensing to get microphysical of clouds. Active observations such as lidar and radar as well as passive measurements such as visible/infrared imaging spectrometers and microwave radiometers have been used to produce cloud products (King et al., 1998.; Austin et al., 2009; Delanoë and Hogan, 2010; Deng et al., 2010; Boukabara et al., 2011). Millimeter frequency radiometers are sensitive to larger precipitating hydrometeors while sub-millimeter frequencies are sensitive to smaller ice particles (Buehler et al., 2007). Cloud radar has the advantage of higher vertical resolution and sensitivity than passive radiometer and can determine the vertical structure of ice clouds. However, this usually comes at the cost of low spectral range and low spatial coverage of the observations (Pfreundschuh et al., 2020).

The brightness temperature (TB) depression caused by the scattering of ice particles is usually proportional to the IWP which simplifies the retrieval method from radiometric measurements (Liu and Curry, 2000). Researches on ice cloud retrieval using radiometers such as AMSU, SSMIS, MHS and MWHS, as well as limb sounders such as MLS, SMR, SMILES have been published for years (Zhao and Weng, 2002; Eriksson et al., 2007; Wu et al., 2008; Sun and Weng, 2012; Millán et al., 2013; Wang et al., 2014). However, these spaceborne radiometers lack the ability of polarization measurement while dual-polarization measurements above 100 GHz show obvious polarized scattering signals of ice clouds. The recent theoretical model research indicates that the non-spherical and oriented ice particles are the main reason for the polarization signal (Brath et al., 2020).

With the increasing frequency, polarimetric measurement will lead to a new understanding of clouds and their microphysical (Buehler et al., 2012; Eriksson et al., 2018; Coy et al., 2020; Fox, 2020). Most passive microwave sensors that have dual-polarization channels are limited to frequencies below 100 GHz. However, these sensors are strongly affected by surface contamination. Currently, only GMI and MADRAS have observed polarimetric signals from ice clouds above 100 GHz (Defer et al., 2014; Gong and Wu, 2017). By analyzing the polarization differences between the 89 GHz and 166 GHz channels of GMI, Gong and Wu (2017) found that large polarization occurs mainly near the convective outflow regions (anvil or stratified precipitation), while in the inner deep convective core and the distant cirrus regions, the polarization signal is smaller. It is roughly estimated that neglecting the polarimetric signal in the IWP retrieval will lead to errors of up

to 30% (Gong et al., 2018). Their further study showed that the main source of the 166 GHz high polarimetric radiance difference (PD) is horizontally oriented snow aggregates or large snow particles, while the low polarization signal could be small cloud ice, randomly oriented snow aggregates, foggy snow, or supercooled water (Gong et al., 2020). The Ice Cloud Imager (ICI) will provide a more comprehensive observation of ice clouds. By covering 176 GHz to 668 GHz~~GHz to 800 GHz~~, ICI has good sensitivity to both large and small ice particles, and its dual-polarization channels also allow observation of horizontal particles (Eriksson et al., 2020).

The Microwave Humidity Sounder (MWHS) onboard the Fengyun-3B (FY-3B) satellite has been proven to give information about IWP (He and Zhang, 2016). The MWHS has quasi-polarization channels at 150 GHz that can provide polarization information of cloud ice~~, and it was hardly analyzed in past studies~~. The neural network is an easy way to find the nonlinear relationships between TBs and IWP while the only problem is the lack of true IWP values. CloudSat is recognized as a relatively accurate instrument for cloud measurement, and its official Level-2C product is used in this paper. Numerous studies have been conducted to compare CloudSat products with in-situ measurements, the results show that the Level-2C product is quite reliable when using a combination of Cloud ProfilING Radar~~Cloud Profile Radar~~ (CPR) and Lidar. Its ice cloud water content (IWC) is fairly close to the in-situ observation (Deng et al., 2013; Heymsfield et al., 2017). Although CloudSat products still have considerable uncertainties (Duncan and Eriksson, 2018), they can give us a relatively accurate reference of IWP and IWC. Holl et al. (2010, 2014) present an IWP product (SPARE-ICE) that uses collocations between MHS, AVHRR, and CloudSat to train a pair of artificial neural networks. The 89 GHz and 150 GHz channels were excluded since they are surface sensitive. However, the 150 GHz channel shows good sensitivity to precipitation-sized ice particles (Bennartz and Bauer, 2003). Brath et al. (2018) retrieve IWP from airborne radiometers of ISMAR and MARSS using neural networks.

In this paper, we present an analysis of IWP retrieval from the FY-3B/MWHS observations based on the deep neural network. Both 150 GHz (QV and QH) channels and their PD are investigated. First, we collocate the MWHS measurements with the CloudSat/2C-ICE IWP according to the observation time and geolocation. Second, we train deep neural networks (DNNs) that are used to filter cloud scenes and retrieve the IWP. The effects of different channels (including PD) and auxiliary information on DNN retrieval are also discussed. Finally, the performance of ~~each~~ the final configuration network is evaluated~~ and the corresponding error is estimated~~. The trained neural networks are used for ~~the IWP retrieval of two~~ a tropical cyclone cases and the ~~monthly averaged~~global annual mean IWP map of MWHS. Zonal mean IWP of MWHS is also compared with Aqua/MODIS L3 product, 2C-ICE and ERA5 reanalysis data. The main aim of this study is to analyze the ability of the MWHS in IWP retrieval, especially the role played by the dual-polarization channels in IWP retrieval.

This paper is organized to describe the data analysis in Sect. 2, followed by the retrieval method in Sect. 3. The IWP retrieval results and analysis are discussed in the subsequent section, with conclusions in the end.

2 Satellite Observations

2.1 Instruments

2.2.1 FY-3B/MWHS

The FY-3B satellite was launched on November 5, 2010, and the MWHS was equipped as one of the main payloads. The MWHS performs the cross-track scanning along the orbit at an angle of ±53.35° from nadir to make 98 nominal measurements per scan line, which is corresponding to a scan swath of 2645 km in 2.667 s with a resolution of 15 km at nadir. It measures at frequencies from 150 GHz to 190 GHz (two window channels at 150 GHz and three channels near the water vapor absorption line at 183 GHz water vapor absorption line), these channels are labeled as CH.1 to CH.5 hereafter. The details of each channel are shown in Table 1 (Wang et al., 2013). Compared to its successors (i.e. MWHS-II) onboard the FY-3C/D/E satellite, the 150 GHz channels of MWHS have quasi-horizontal and quasi-vertical polarization that can include unique cloud information. These channels can provide information near the Earth's surface and lower atmosphere, and can also be used to measure atmospheric cloud parameters. For the 150 GHz channels, Zou et al. (2014) investigated the polarization information and concluded that the polarization signal is related to the scan angle and also to information such as surface wind speed, wind direction and salinity, especially in the clear-sky condition. Under all weather conditions except heavy precipitation, all five channels of MWHS can observe water vapor and ice in the atmosphere. In this study, the Level-1B brightness temperature data set of MWHS is used.

Table 1. Channel characteristics of MWHS

Channel	Central frequency (GHz)	Polarization	Bandwidth (MHz)	NEDT (K)
1	150	H	1000	0.8
2	150	V	1000	0.8
3	183.31±1	H	500	0.9
4	183.31±3	H	1000	0.5
5	183.31±7	H	2000	0.5

2.1.2 CloudSat/CALIPSO

CloudSat is a cloud observation satellite launched into the NASA A-Train in April 2006, with a 94 GHz cloud profiling radar providing continuous cloud profile information (Stephens et al., 2008). The footprint size of CPR observation is about 1.3 km × 1.7 km, with a vertical resolution of 240 m. The scan time for each profile is about 0.16 s, and its sensitivity is -30 dBZ. It has an orbital inclination of 98.26°, which is similar to the FY-3B satellite. The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) was launched with the CloudSat satellite and designed to fly close to each other in the A-Train satellite constellation to make synergistic observations. The Cloud-Aerosol Lidar with Orthogonal

Polarization (CALIOP) carried on the CALIPSO is a dual-wavelength polarized lidar, providing 532 nm and 1064 nm
120 backscatter profiles with a footprint of 75 m cross-track and 1 km along-track (Winker et al., 2009).

The CloudSat and CALIPSO Ice Cloud Characterization product (2C-ICE) contains retrieved estimates of IWC, effective radius and extinction coefficient for identified ice clouds measured by CPR and CALIOP with orthogonal polarization. The 2C-ICE cloud product uses a combined input of the radar reflectivity factor measured by the CPR and the attenuated backscatter coefficient measured by the Lidar at 532 nm to constrain the ice cloud retrieval more tightly than using only the
125 radar product and to produce more accurate results (Mace and Deng, 2019). The combination of CPR and CALIOP provides a more complete measurement of the ice clouds than any other current spaceborne sensor measurements. Further study showed that this combined retrieval method is less sensitive to the changes in the assumed microphysical properties than CPR or CALIOP single retrieval (Delanoë and Hogan, 2010).

The 2C-ICE retrieval relies on forward model assumptions. Lidar is sensitive to small particles near the top of the cloud,
130 but cannot measure that deep in the cloud which can lead to an unquantifiable error (Mace et al., 2009). A sensitivity study shows that multiple scattering, assumptions regarding particle habits and size distribution shapes are critical to the accuracy of the retrieval (Deng et al., 2010). The research also finds that the ratio between IWC product and in-situ measurements is similar to the ratio between two independent in-situ measurements (around a factor of 2) and conclude that the retrieval agrees well with in-situ data. Since 2C-ICE is used to train the retrieval network in this work, the trained network directly
135 inherits all the systematic errors and limitations of the product.

2.2 Collocation

Collocated measurement is the occurrence where two or more sensors observe the same regions at the same time. One factor for the collocation window requirements is the specific observation target. Ice clouds is a fast-changing (minutes to hours) atmospheric parameter that needs a window of short time and small space. Another considered factor in defining the
140 collocation window is the number of meaningful statistics for training.

The ascending node time of CloudSat is between 13:30 and 13:45 at the local solar time (LST) which is close to that of FY-3B (13:30 LST). Because of the close orbits and the ascending time between FY-3B and CloudSat, the number of collocated measurements is large. In this study, a collocation data set of MWHS and 2C-ICE was created by setting the collocation window to 15 min in time and 15 km in space. Since the footprint of MWHS is an order of magnitude larger than
145 that of CPR, multiple 2C-ICE pixels can be found within one MWHS measurement. Thus, the IWP values of 2C-ICE within a circular window (with a radius of 7.5 km) were averaged to represent the mean IWP for the MWHS measurement pixel. According to this collocation strategy, 1207731 collocations have been found between the FY-3B/MWHS and the CloudSat/2C-ICE for the year 2014. Since the different observation methods of MWHS and CPR/CALIOP, only 14 pixels of 2C-ICE are contained in the best case of collocations (See Fig. 1a). Thus, the CloudSat footprints cover at most 13.75% of
150 the area of an MWHS footprint, an error from imprecise collocation is unavoidable and the representation of the data set must be considered. ~~The criteria discussed in Holl et al. (2010) is applied to reduce the sampling effect of collocations.~~

~~However, in the case of highly inhomogeneous clouds, larger uncertainty for the IWP within MWHS pixels cannot be eliminated.~~

Figure 1 illustrates the statistics of 2C-ICE IWP within the MWHS footprints in the collocations. In most cases, more than 10 pixels of 2C-ICE were averaged in the corresponding MWHS pixel. However, there are still many MWHS pixels that only cover a small quantity of 2C-ICE pixels which means the collocations are poorly represented. The coefficient of variation of each collocation pixel is manifested in Fig. 1b. The coefficient of variation is used to represent the IWP dispersion of 2C-ICE pixels in each MWHS pixel. When the coefficient of variation is small, it means the IWP of 2C-ICE pixels averaged in this MWHS pixel are homogeneous and represent the scene that MWHS observed relatively well. Since the collocation error cannot be estimated, the criteria discussed in Holl et al. (2010) is applied to reduce the sampling effect of collocations. In this study, an MWHS pixel with more than 10 pixels of 2C-ICE and less than 0.6 coefficients of variation were selected for subsequent processing. However, in the case of highly inhomogeneous clouds existing outside the CloudSat field of view, larger uncertainty for the IWP within MWHS pixels cannot be eliminated. After the reduction of inhomogeneous collocations, 665519 collocations were retained. ~~Overall, the collocation dataset is available for subsequent processing.~~

(a)   (b)

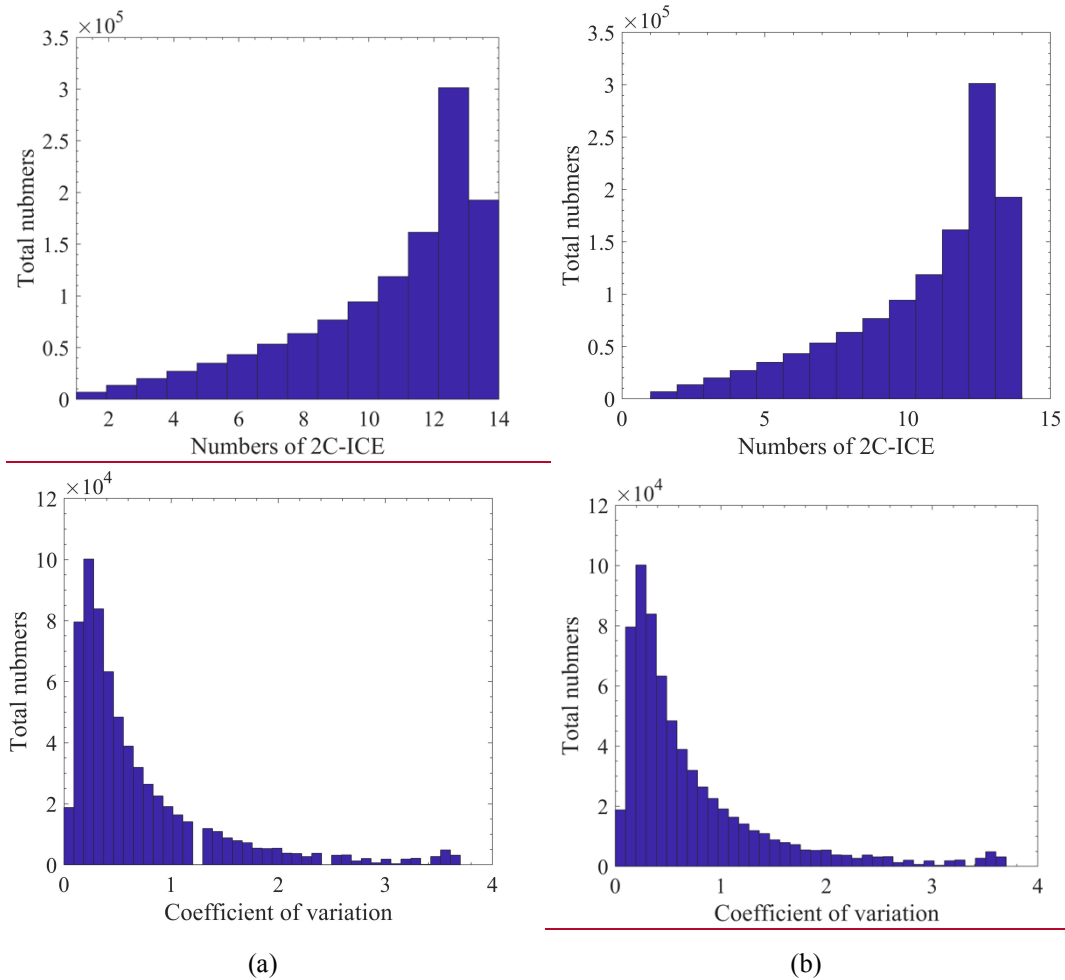
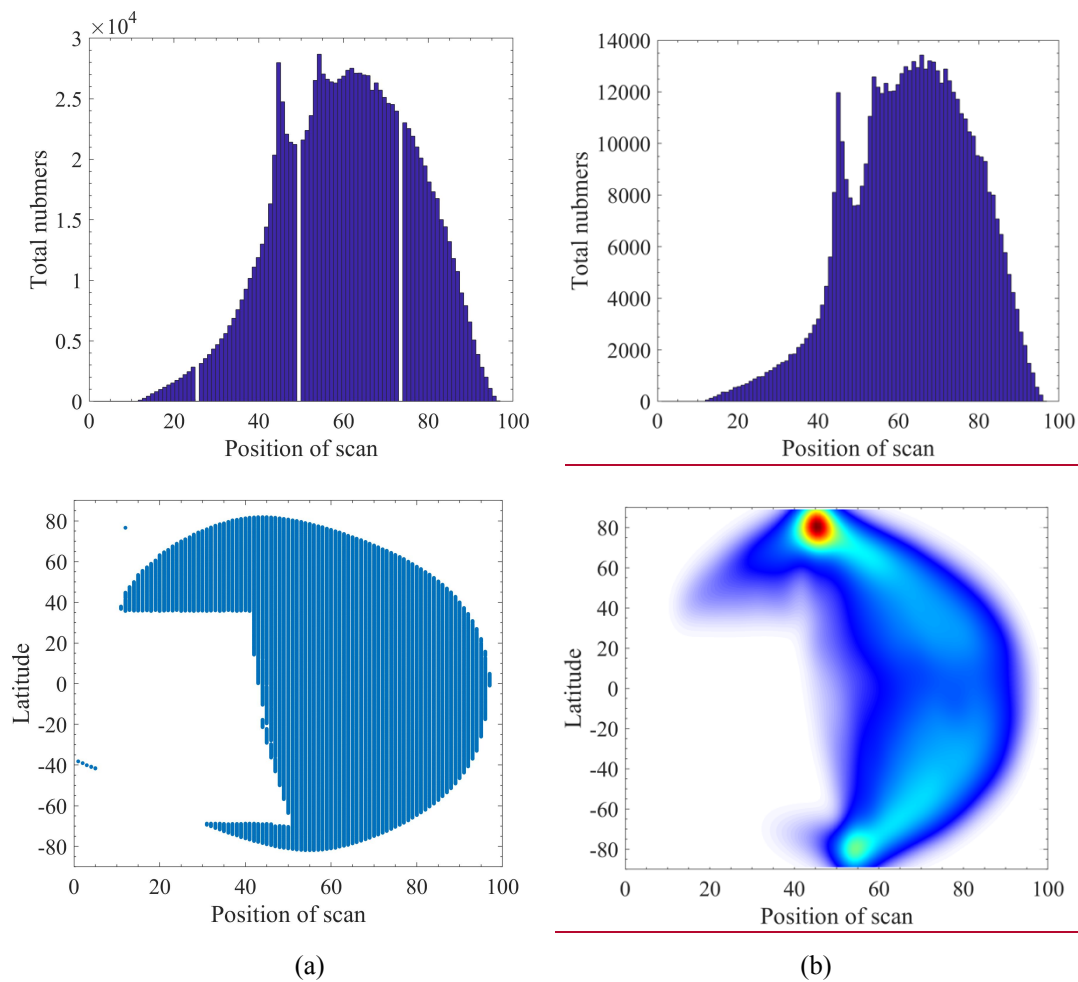



Figure 1. Statistical information of MWHS and 2C-ICE collocations in 2014. (a) Histogram of the number of 2C-ICE pixels within an MWHS pixel. (b) Histogram of the coefficient of variation of the collocations.

170   Figure 2 and 3 give statistical information on the scan angle, latitude and time of the MWHS measurements in this data set. Since the data set is used for global retrieval, it must have sufficient samples and their distribution must represent the real world. According to the statistical results of the collocated MWHS pixels shown in Fig. 2, most of the collocations occurred on one side of the flight direction (from the 40th to 90th scan pixel). In terms of observation latitude, the collocations near the nadir scan (the 49th pixel) cover the latitude from 80°S to 80°N, while at the edge of the observation (the 90th pixel) they

175   only cover the tropical regions. In terms of observation time and latitude, Figure 3 illustrates that there is an obvious lack of data above 60°S from April to September, and there are also few data between 0° and 30°S in December. The data distribution suggests that the training in polar regions may be inadequate. Due to the high number of collocations near the poles, 121500 observations at high latitude were randomly excluded to obtain a balanced data set. For IWP retrieval, collocations should be classified into two bins (clear-sky scene and cloudy scene) according to a specific IWP threshold. A

 threshold of IWP >100 g m^{-2} is preliminarily selected to classify cloudy scenes. Thus, 81490 collocations are recognized to be cloudy scenes and 462529 collocations are clear sky scenes in this data set.

(a)   (b)

185   Figure 2. Statistical information of scan angle and latitude of MWHS observations in the collocation data set.

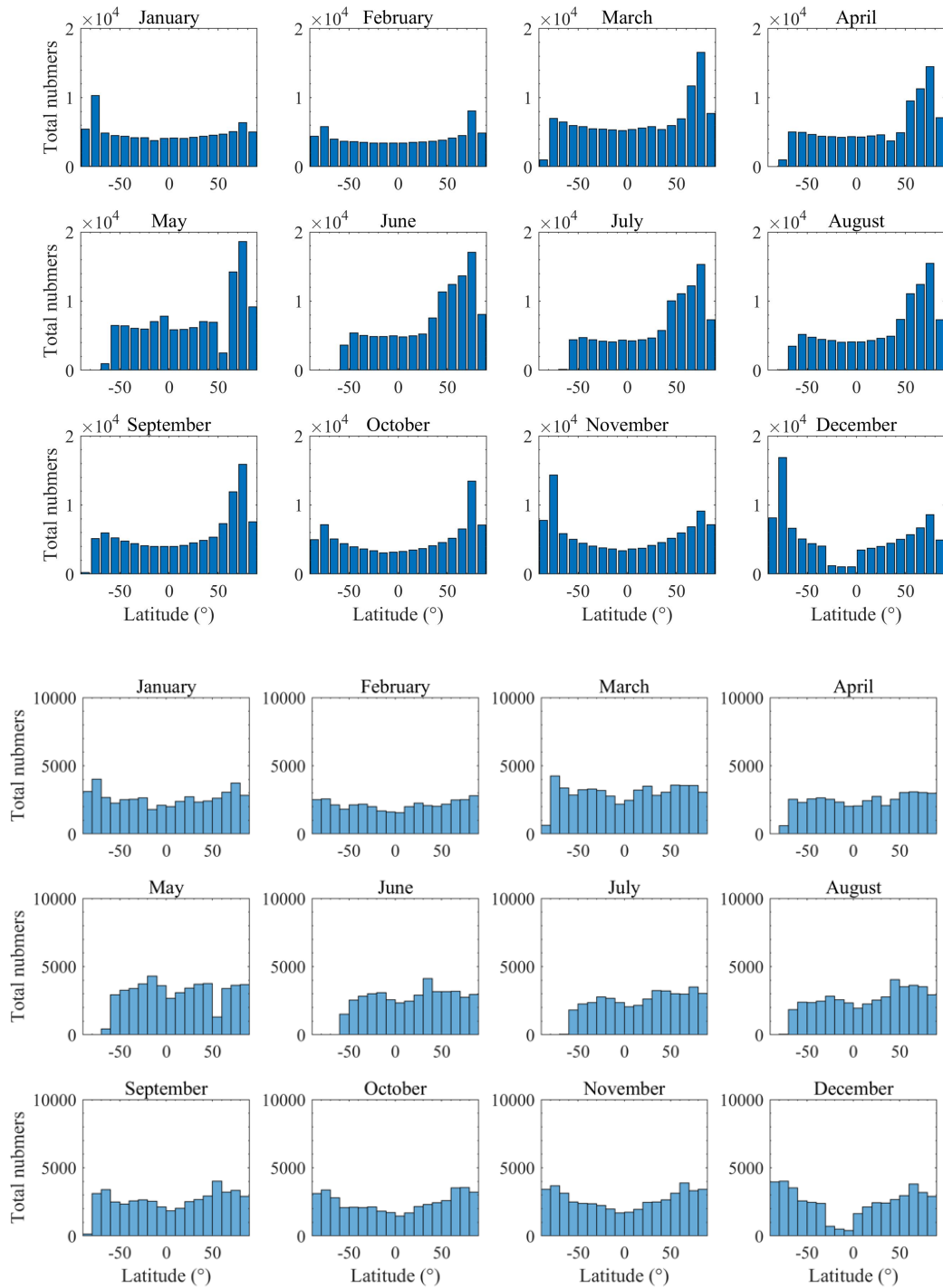



Figure 3. MWHS measurements distribution of time and latitude in the collocation data set. ~~For IWP retrieval, the collocations should be classified into two bins (clear sky scene and cloudy scene) according to a specific IWP threshold. A~~

190 threshold of IWP >10 g/m2 is preliminarily selected to classify cloudy scenes. Thus, 401520 collocations are recognized to be cloudy scenes in this dataset.

The density plots show the statistics of of the PDPD and TB at 150 GHz and TB (clear-sky and cloudy) at 150 GHz(clear-sky and cloudy scene) and the corresponding IWP from 2C-ICE over the ocean and land over the ocean and the land in 2014are depicted in Fig. 4 and 5 . Scan angles from ±40.15° to ±53.35° are selected to compare the results with observations

195 from conical scanners. The scatter points are color-coded by IWP from 2C-ICE. There is 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 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

200 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. a clear difference in the shape of the scatter plot between the ocean and the land. There is a significant peak of PD at TB of 210 K to 250 K over the ocean while there is no peak of PD over the land. This may be related to the strong polarization signal at the ocean surface. Compared the cloudy

205 scene with the clear-sky scene, the scatter shapes are quite similar over the ocean and the land except when IWP is larger than 1000 g/m2. The PD (around 5 K) in both cloudy and clear-sky may be caused by the scan angle due to the quasi-polarization measurement of MWHS. It also can be found that the PD which may be caused by ice particles is about 5 K when IWP is larger than 1000 g/m2. This is consistent with the retrieval results shown in section 4.2.

210

(a)   (b)

(c)   (d)

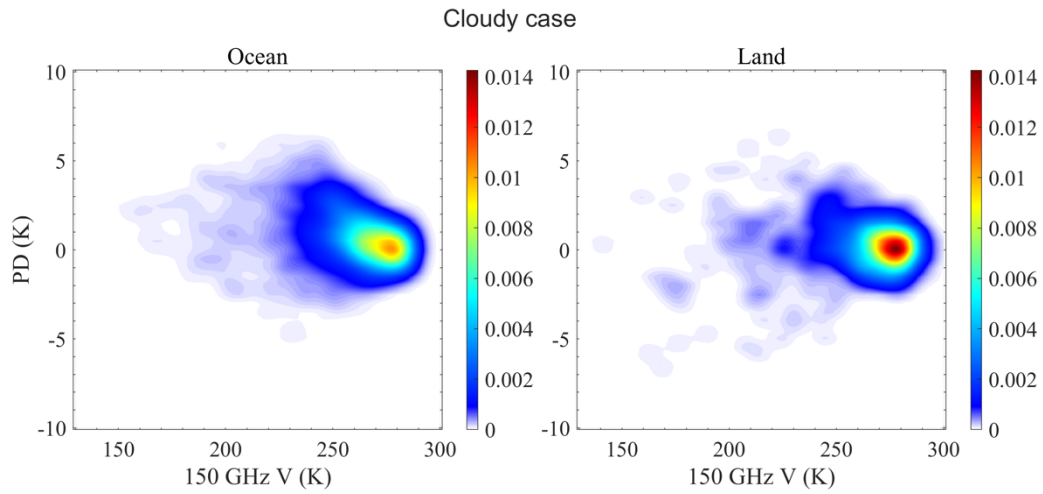
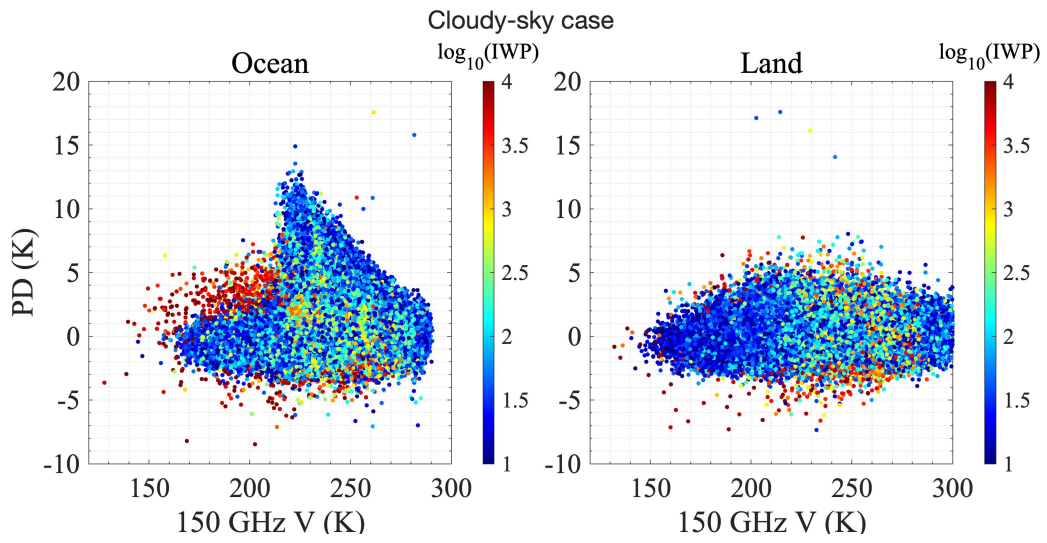
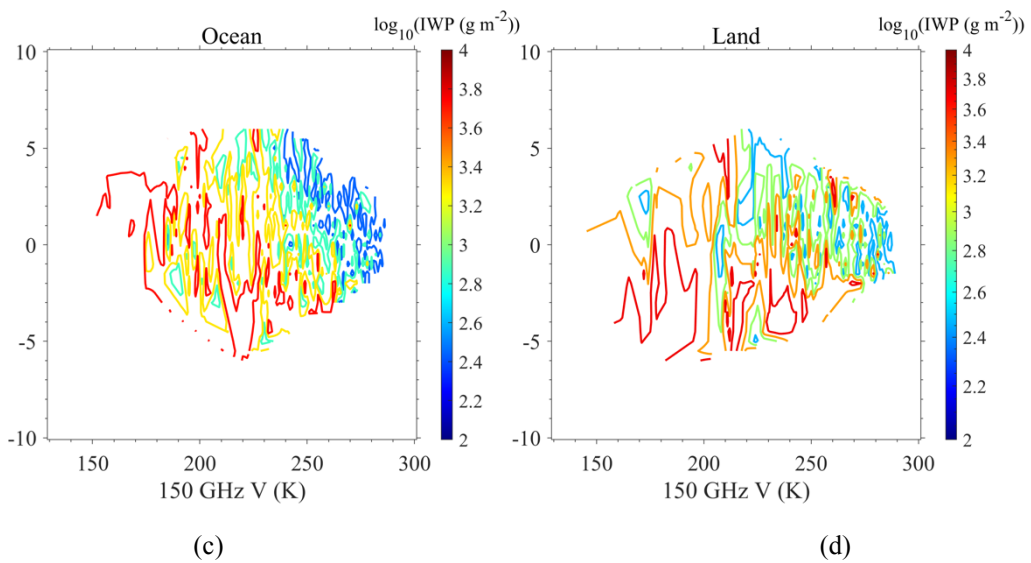



Figure 4. The PD–TB$_{150v}$ density plots for the collocations in the cloudy scenes over the ocean (a) and land (b). The (c) and (d) show the corresponding IWP from 2C-ICE. ~~Figure 4. The PD-TB$_{150v}$ scatter plots for the collocations in the cloudy scenes over the ocean (a) and land (b). The points are color-coded by IWP from 2C-ICE.~~

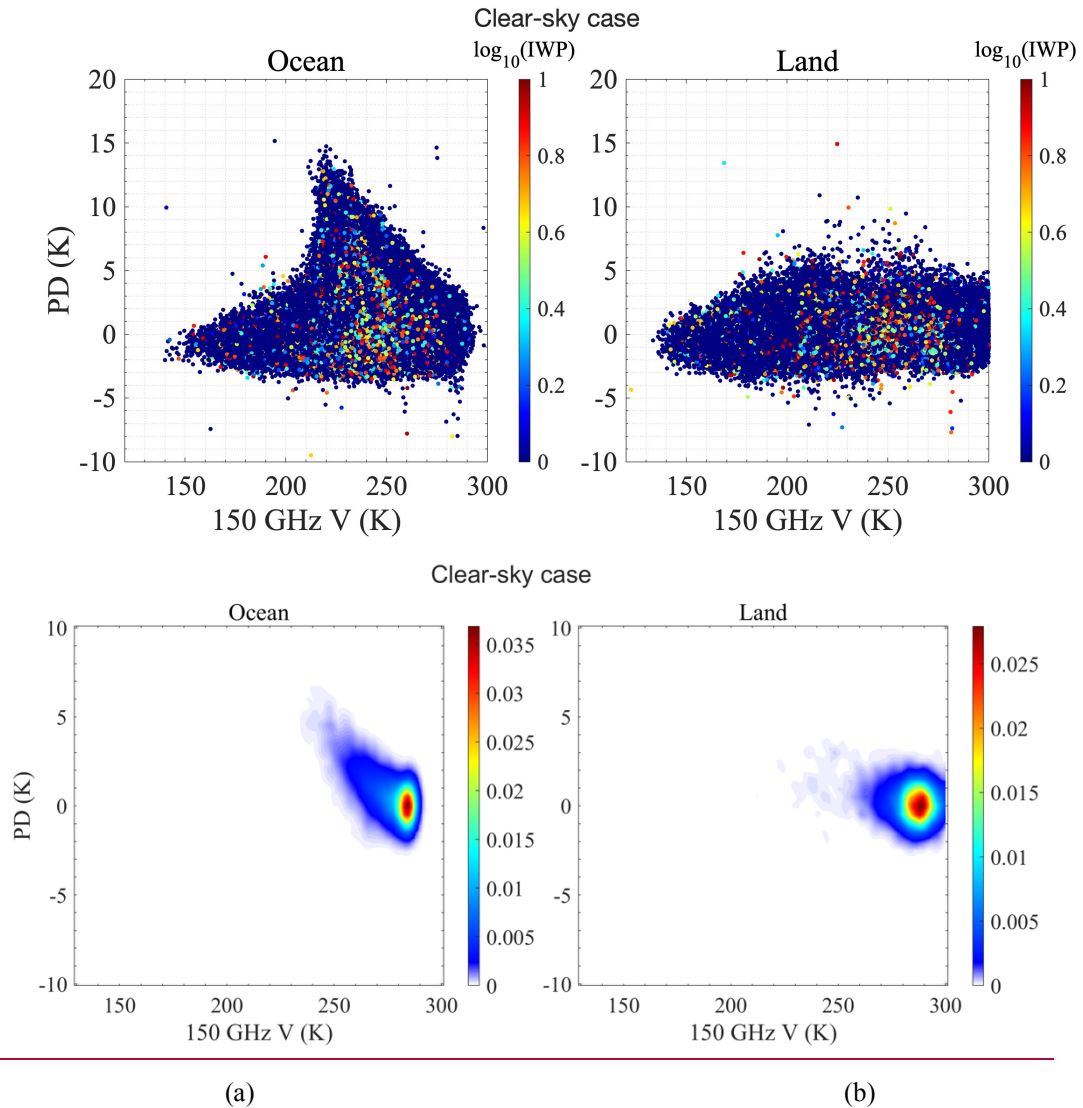



Figure 5. Same as Fig. 4 but for clear-sky scenes.

3 Retrieval method

The collocations are used as a retrieval database to train the networks, the processing flow is shown in Fig. 6. The DNN is a
220   feed-forward neural network which contains an input layer, several hidden layers, and an output layer. The DNN is a fully
connected network, neurons in each layer connect with all neurons in the next layer. The hidden layers are used to perform
the nonlinear calculation to achieve a nonlinear mapping of the relationship between input and output data. DNN is based on
backpropagation learning algorithms to search for a minimum loss function (such as the mean squared error between

prediction data and reference data) and then adjust the thresholds and weights iteratively to close the reference data. The

225 outstanding nonlinear mapping capability makes DNN popular for geophysical retrieval.

In this study, DNN with 6 layers is selected. The first layer is the input layer, and each input quantity uses a neuron to connect with the next layer. The second to fifth layers are the hidden layers, in which 300 256 neurons are used for each layer, and the tanh and the Rectified Linear Unit (ReLU) is selected as the activation function for the cloud filtering network and the IWP retrieval network, respectively. Since networks are prone to overfitting in the training, the early stopping and

230 dropout method is used to improve the training. To remove the effect of the order of data, random assignation and normalization are performed in the front of the hidden layers. The final layer is the output layer which uses the IWP of 2C-ICE (transfer to log space) as reference. The activation function of the last layer is selected according to the target of the network. For the determination of cloudy and clear-sky scenes, the sigmoid function is used for binary classification. For the IWP retrieval, the results are output directly. Due to the imbalanced data set of the clear-sky and cloudy scenes, the "focal

235 loss" function which can solve the problem of serious imbalance of positive and negative sample ratio in one-stage object detection is used instead of the cross-entropy loss function (Lin et al., 2017). In the iterative training of the networks, the models with the best results in the validation data will be retained. The hyperparameters were chosen by comparing the performance of DNNs with different hidden layers, number of hidden neurons and regularization parameters. Each network mentioned in the next section uses the same hyperparameters of the model to ensure that the performance of the network is

240 only affected by the input parameters.

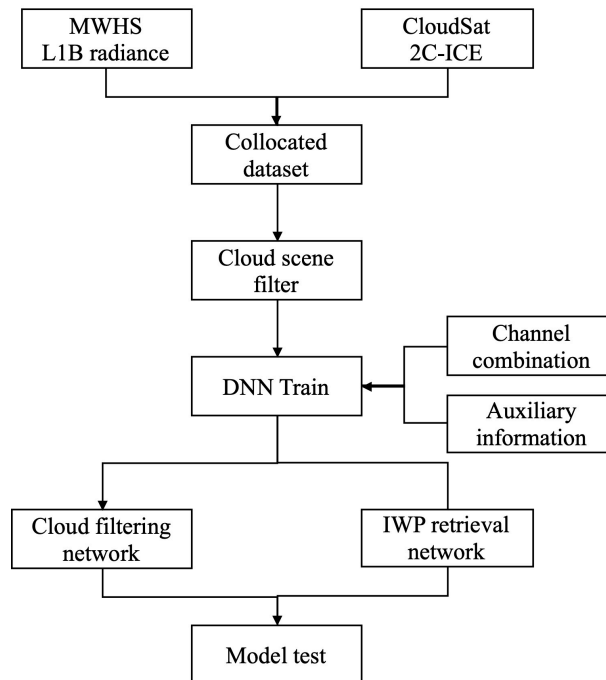



Figure 6. The schematic of the MWHS retrieval based on the DNN model.

The sensitivity of ice clouds is discussed by Holl et al. (2010) and Eliasson et al. (2013), their studies show no significant radiance signals at IWP <100 $\mathrm{g\ m^{-2}}$ ~~g/m²~~ for MHS measurements. Thus, ~~two~~ it is used as the threshold for the cloud filtering network. ~~thresholds for cloud filtering are considered respectively with 2C-ICE averaged IWP of 10 g/m² and 100 g/m² within the MWHS footprint. For the IWP threshold of 10 g/m², 401520 collocations are left. For the IWP threshold of 100 g/m², 168898 collocations are left.~~

From those collocations, we randomly assign 75% to be used for training and 25% to be used for validation. The training data are used as a sample of data for model fitting. The validation data can be used to tune the hyperparameters of the network and for preliminary evaluation of the model. Collocations during January 2015 are used for testing. These data are not used to train the networks and adjust the hyperparameters but serve as independent data to test the performance of the final obtained networks.

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

$$M = \begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix} \tag{1}$$

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

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$FAR = \frac{FP}{TP + FP} \tag{3}$$

$$POD = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{5}$$

$$CSI = \frac{TP}{TP + FN + FP} \tag{6}$$

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

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

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|y_{pred,i} - y_{valid,i}\right|}{y_{valid,i}} \times 100\% \tag{8}$$

270

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

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

4 Results

To retrieve the IWP from the MWHS measurements, two networks were trained for different capabilities. The first one
275 allows classifying a scene according to whether it is clear or cloudy. The second is to retrieve the IWP. The two networks are used separately, and the IWP of the scene considered clear is set to 0. Due to the randomness of the neural network in the assigned training and validation data, 20 models were trained for each combination to ensure the stability of the model results. ~~To retrieve the IWP from the MWHS measurements, two networks were trained for different capabilities. The first one allows is to classify classifying a scene according to whether it is clear or cloudy whether is clear sky or cloudy. The other one second is to retrieve the IWP. The two networks are used separately, and the IWP of the scene considered clear which is~~
280 ~~considered to be clear sky is set to 0. the ed~~

4.1 Cloud Filtering Network

The network structure, training data set and cloud IWP threshold are discussed above. The sigmoid activation function can vary the output of the network from 0 to 1, which represents the probability of cloud occurrence. Thus, a threshold value of

cloud probability must be assigned to determine the cloudy scene. After testing, a threshold value of 0.4 is the most appropriate for this cloud filtering. 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

4.2 IWP Retrieval Network

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. These cases are divided into three parts, their details and mean retrieval errors are listed in Table 2. To avoid the effect of random factors, each combination was run twenty times and the best result was taken.

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 MWHS 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 first part (Case 1-5) considers the distribution of the training data. Collocations selected from IWP >10 g/m² and IWP >100 g/m² are considered separately (Case 1 and Case 2). The results show that Case 1 is not converged well in the training. Its mean error is much larger than that of other cases. Thus, the IWP threshold of 100 g/m² is

315 ~~used in the following cases. It is consistent with the results of cloud filtering. There are 47976 collocations over land and 112085 collocations over the ocean in this dataset. Data over land and ocean are also retrieved separately (Case 3 and Case 4). Case 5 uses the better representative collocations (i.e. sample optimization discussed in section 2.2) to train the network. In this case, MWHS pixel with more than 10 pixels of 2C-ICE and less than 0.6 coefficients of variation are selected which lead to the collocations reducing to 100249. Observations over the ocean show a mean relative error of 81.77% while retrieval over land shows a poorer result of 100.92%. This may be due to the insufficient amount of data over the land. Applying the criteria of sample optimization, the result is improved compared with Case 2 which means the collocations are sampled well. Case 5 have the mean and median relative error of 96.72% and 52.84% respectively and it is used in Part II as the basic input. Figure 7 shows the mean and median relative errors of Case 2-5 using MWHS all channels with reference~~

320 ~~IWP divided into several bins from 100 g/m² to 10000 g/m². The performance of the retrieval is poor (mean errors >150%) between IWP of 100 g/m² and 200 g/m² while the errors reduce to 45-70% at IWP >200 g/m². It should be noted that there are only 13 samples in the last bin (i.e. IWP >10000 g/m²).~~ 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 g m^{-2} and IWP >1000 g m^{-2}.

325

Table 3. Errors of IWP retrieval using different auxiliaries ~~Table 2. Mean errors of IWP retrieval~~

	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

~~Part~~	~~Combination Case~~	~~Mean Error (%)~~	~~Median Error (%)~~
~~I. Data distribution~~	~~1. IWP >10 g/m²~~	~~170.17~~	~~68.60~~
	~~2. IWP >100 g/m²~~	~~98.01~~	~~54.23~~
	~~3. Only Land~~	~~100.92~~	~~57.91~~
	~~4. Only Ocean~~	~~81.77~~	~~51.45~~
	~~5. Sample Optimize~~	~~96.72~~	~~52.84~~

II. Auxiliary information	6. Latitude	83.49	49.65
	7. Scan Angle	101.16	52.74
	8. Mask	80.42	50.41
	9. Latitude + Scan Angle	81.23	49.86
	10. Latitude + Mask	79.28	48.38
	11. Scan Angle + Mask	86.27	52.15
III. Channel selection	12. CH.1-CH.5	72.02	46.29
	13. CH.2-CH.5+PD	74.68	46.45
	14. CH.2-CH.5	73.20	47.91
	15. CH.3-CH.5+PD	82.79	50.13
	16. CH.3-CH.5	82.21	50.99

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 g m^{-2}. This conclusion is close to the analysis in Figure 4. In general, all channels of MWHS contribute to ice cloud retrieval. The second part (Case 6-11) aims to select the important auxiliary information which is useful for the retrieval. Based on Case 5, additional information including latitude, scan angle and ocean/land mask is added to train the networks. In terms of the mean errors, a significant improvement of retrieval performance is achieved by adding latitude or ocean/land mask information and even more improvement when they are all added. In MWHS measurements, the signal from ice clouds is a reduction in TB through scattering effects. When lacking the latitude information, it is difficult to distinguish whether the decrease of TB is due to the ice particles or the low radiance from the surface or atmosphere. So is the ocean/land mask information. Case 10 (Latitude + Mask) shows a mean error of 79.28% which is improved by 17.44% compared to Case 5. The scan angle seems to have a negative impact on the retrieval (Case 7). However, when it is combined with latitude or mask information, the results are also improved (Case 9, 11). The retrieval error of each bin is shown in Fig. 8. Compared to the first part, adding auxiliary information can significantly reduce the retrieval errors, especially when the IWP is small.

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

The third part (Case 12-16) compares the performance of different channel combinations (all the auxiliary information is added). In this part, the influence of the 150 GHz channel and PD is mainly focused on. According to the retrieval errors, the water vapor channels (CH.3-CH.5) at 183 GHz show good sensitivity to IWP (Case 16) while adding the window channel (CH.2) give a large improvement to the retrieval (Case 14). It might be due to the sensitivity of 150 GHz to the deep convective clouds and precipitating ice. Case 12 and Case13 exhibit similar retrieval performance which means the network can discriminate the polarization information well. Figure 9 illustrates the errors in different IWP bins which give a more detailed comparison. Compare Case 15 with Case 16, adding PD give an obvious improvement on the retrieval results at IWP >1000 g/m². The mean and median errors are reduced by ~10% and ~15% at IWP >2000 g/m², respectively. However, it increases a little mean and median error at IWP <500 g/m². It can be explained that PD at the window channel is also sensitive to the surface which will make the network misidentify the PD from the surface as IWP when the cloud is thin. When the cloud is thick, more oriented ice particles appear to cause a relatively large PD. It can be improved by adding the surface sensitive channel simultaneously. Case 13 illustrates that adding CH.2 significantly reduces the retrieval error at IWP <400 g/m² and maintains the benefit from adding PD at IWP >1000 g/m² (compare Case 13 and Case 15). Case 13 and Case 14 also demonstrate the improvement in retrieval performance by adding PD at IWP >1000 g/m². Figure 10 shows the scatter plot between MWHS IWP and 2C-ICE IWP in January 2015.

Since the 2C-ICE retrievals are proved to be mostly no more than a factor of 2 from the in situ measurements (i.e. 100% error), it gives a lower bound of the network retrieval error. According to the law of error propagation, if both are random and uncorrelated, the true mean error of the best network retrieval (Case 13) can be calculated as $\sqrt{1^2 + 0.7202^2} \times 100\% = 123.24\%$ and the true median error can be calculated as $\sqrt{1^2 + 0.4629^2} \times 100\% = 110.10\%$.

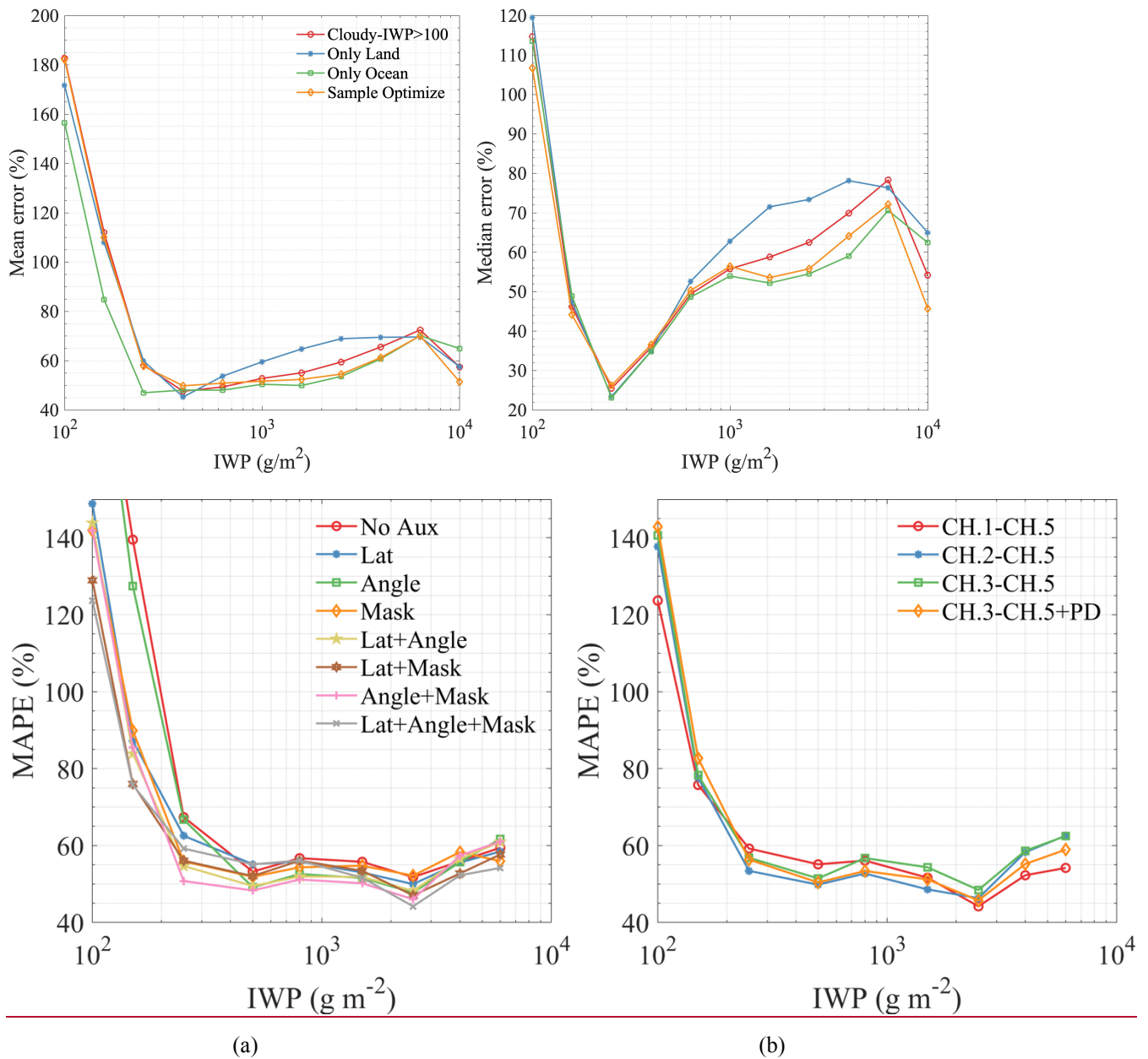



Figure 7. Comparison between the performance of the IWP retrieval networks using different auxiliary and channel combinations of input. Figure 9. Comparison between the performance of the IWP retrieval networks using different channel combinations of input data. All the auxiliary information is added.

370

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 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

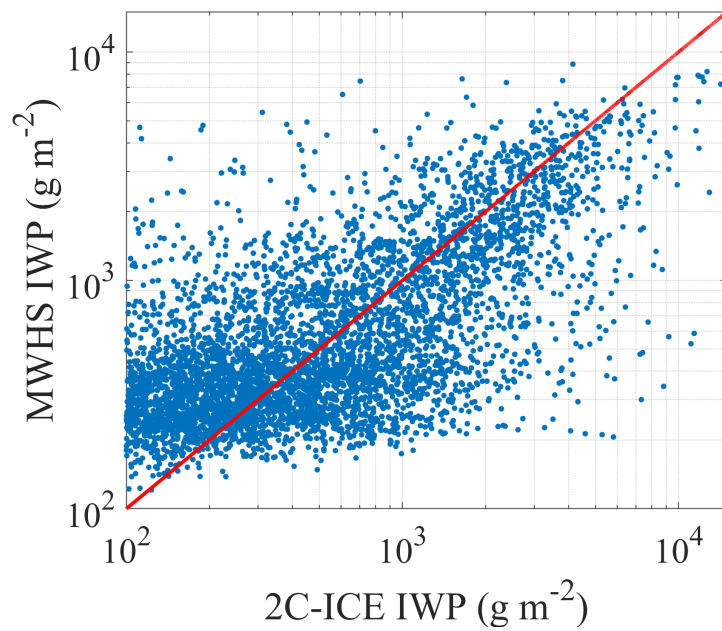
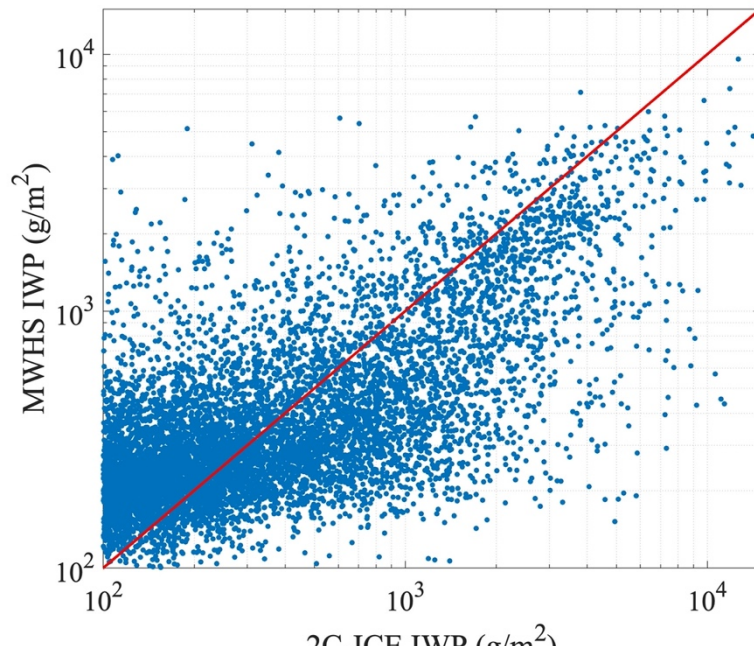



Figure 8. Comparison between 2C-ICE and MWHS IWP. The red line represents the diagonal 1:1 line. Clear-sky scenes are not shown. ~~Figure 10. Comparison between 2C-ICE and MWHS IWP. The red line represents the diagonal 1:1 line. Clear-sky scenes are excluded.~~

4.3 Network application

4.3.1 Tropical Cyclone IWP retrieval

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. ~~MWHS observed TB in the lifetime of the typhoon Rammasun that occurred in July 2014 is manifested in Fig. 11. Three regions of typhoon (eye, eyewall, and spiral rain bands) is obvious. Quite low TB (as low as 150 K) can be found in the regions of the eyewall and spiral rain bands which is mainly caused by the scattering of ice particles in the clouds. There is no significant difference between the TB over the ocean and the land. According to t~~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. ~~he PD at 150 GHz ($TB_V - TB_H$), its distribution characteristics are the same as the typhoon. 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. All channels show similar typhoon characteristics.~~

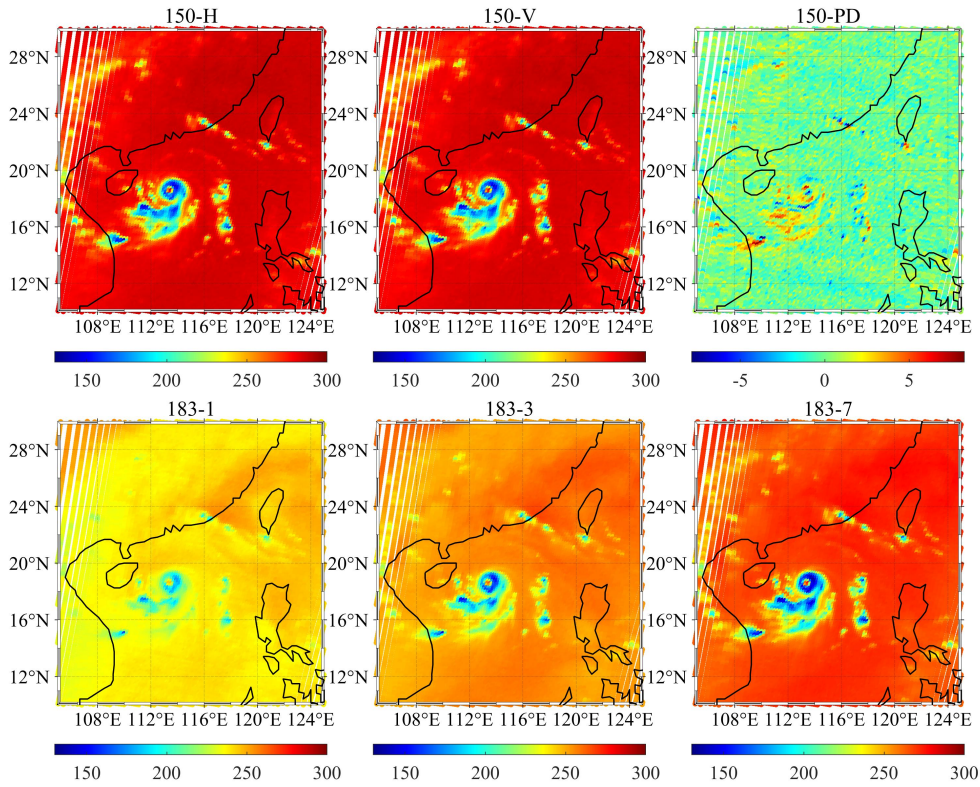
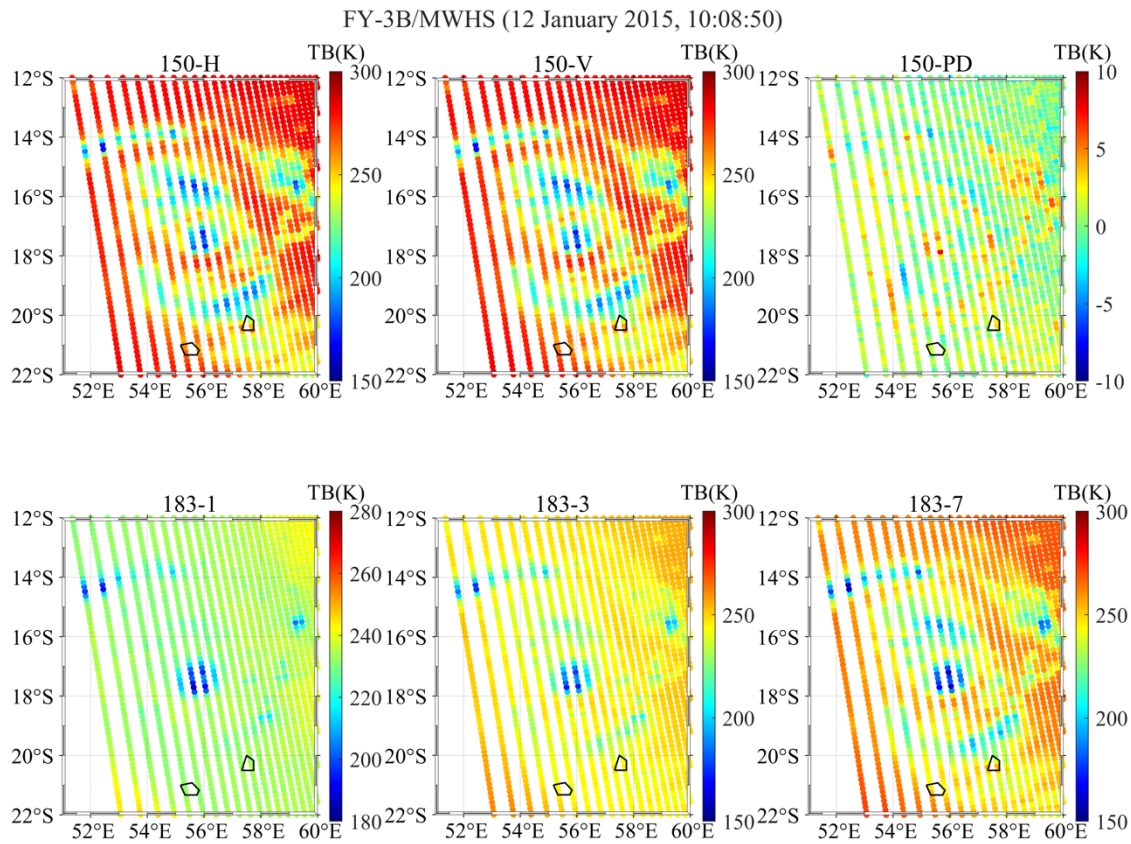



Figure 9. Tropical cyclone Bansi on 12 January –2015 as observed with FY-3B/MWHS channels.Figure 11. Typhoon Rammasun on 17 July 2014 as observed with FY-3B/MWHS channels. The polarimetric differences are shown separately.

Comparing the 37 GHz and 89 GHz TB of Microwave Radiometer Imager (MWRI) onboard the same satellite, the 89 GHz channels of MWRI show similar but higher TB in the typhoon regions which means the 89 GHz TB is not only sensitive to cloud ice but also contaminated by liquid water and ocean surface. The 37 GHz TB shows the complete opposite characteristics to the 89 GHz TB since it is only sensitive to liquid water which greatly increases the TB. By comparison, the 37 GHz TB shows a significantly high temperature in the typhoon regions, the 150 GHz TB shows a significantly low temperature in the typhoon regions due to its sensitivity to ice particles, and the 89 GHz TB is lower than the 37 GHz but higher than the 150 GHz in the typhoon regions due to its sensitivity to both ice particles and liquid water. Thus, the surface temperature and emissivity are quite important if the 89 GHz measurements are used in the retrieval.

The PD from 37 GHz channels shows a clear division between the land, ocean and typhoon which is opposite to the TB map. There is no polarization signal from the land and the polarization difference from the ocean is large. The PD in the typhoon region is almost zero because the 37 GHz TB is not affected by ice particles. And the polarization signal from the ocean surface cannot reach the sensor due to the influence of liquid water. The PD from 89 GHz measurements contain a joint contribution from the ocean surface and ice particles which show about 0 K at land and 5-10 K in the typhoon regions and much larger at the ocean surface. There is no significant difference in the PD between the ocean (non typhoon regions) and land which means the 150 GHz channel is not strongly affected by the surface polarization signal and can provide a good representation of the polarization information of cloud ice particles. From the comparison above, PD of 150 GHz in the typhoon region does come from the ice cloud rather than the ocean surface. The PD from MWRI 37 GHz, 89 GHz and

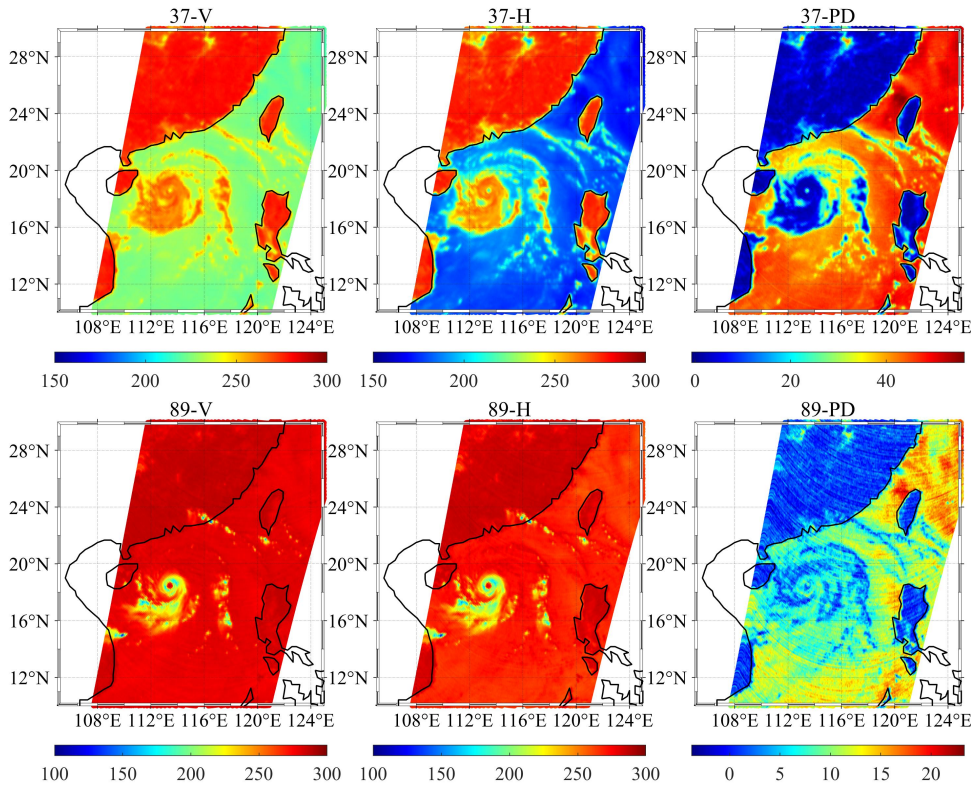
FY-3B/MWRI (2014-07-17, 18:13:39)

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. Figure 12. Typhoon Rammasun on 17 July 2014 as observed with FY-3B/MWRI 37 GHz and 89 GHz channels. The polarimetric differences are shown separately.

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

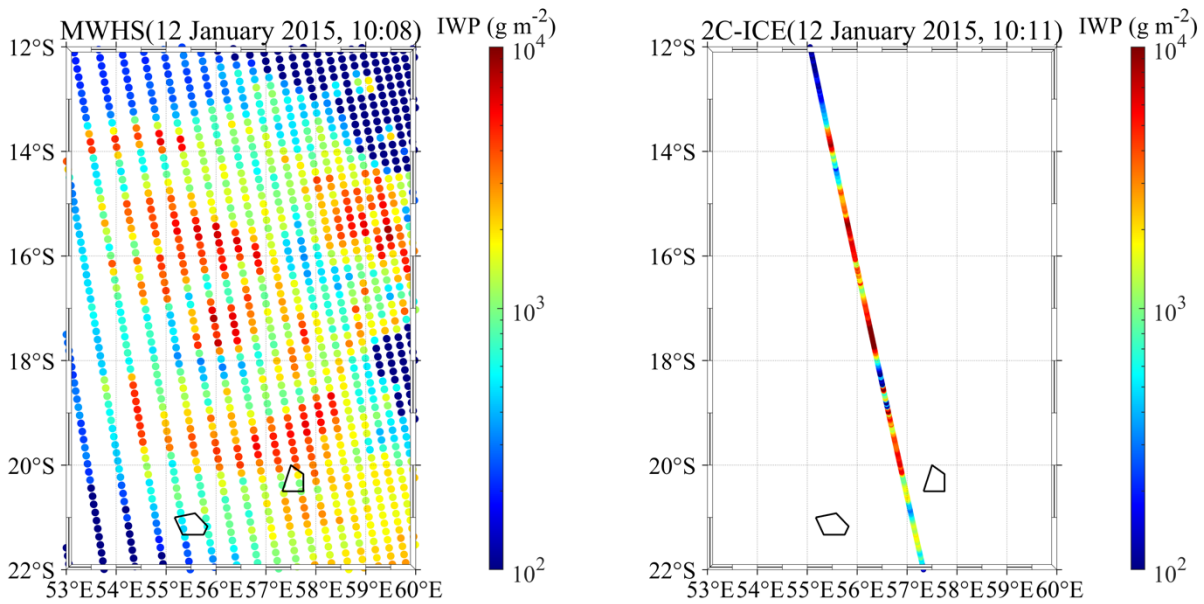
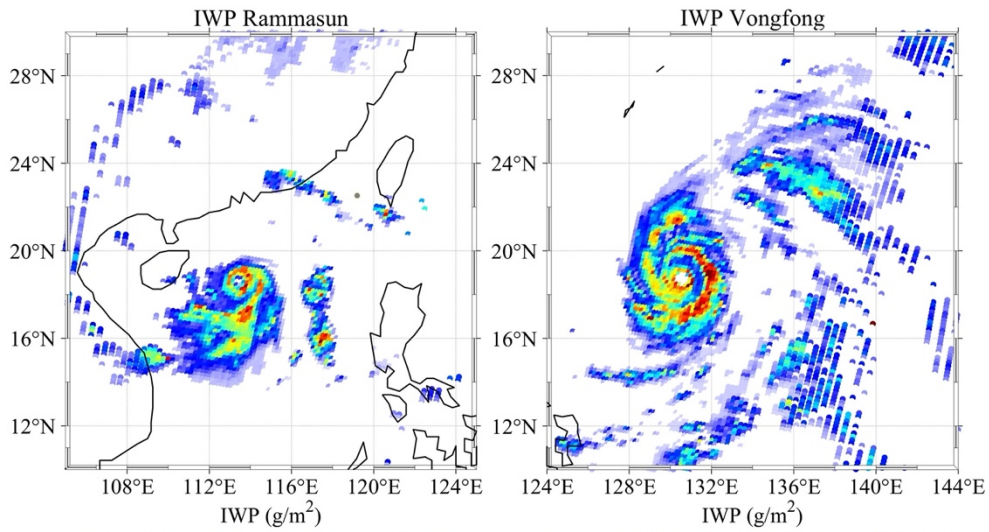
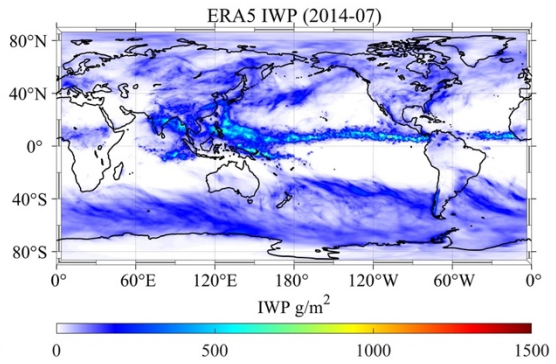
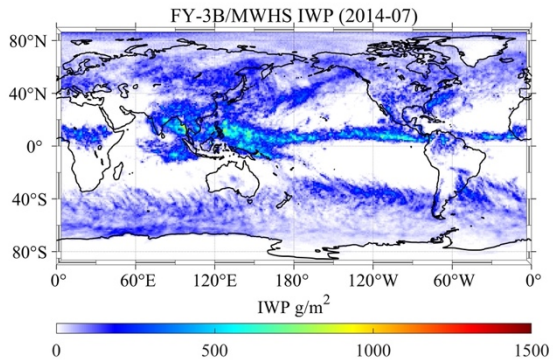
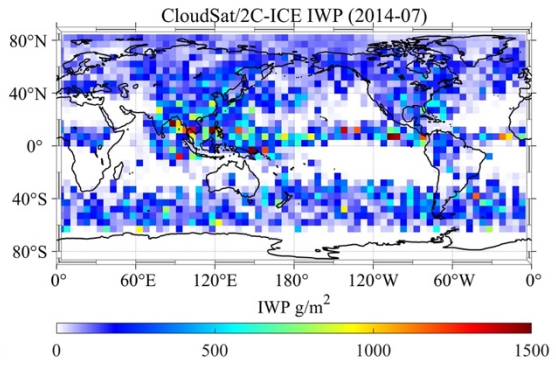
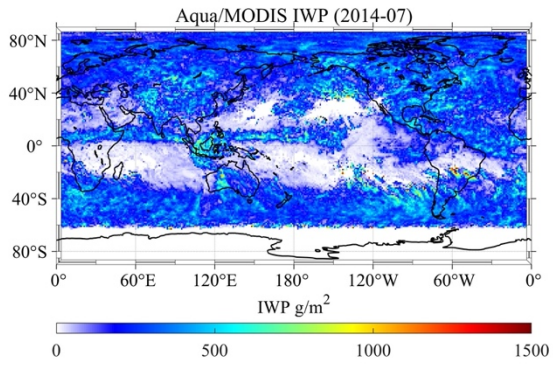
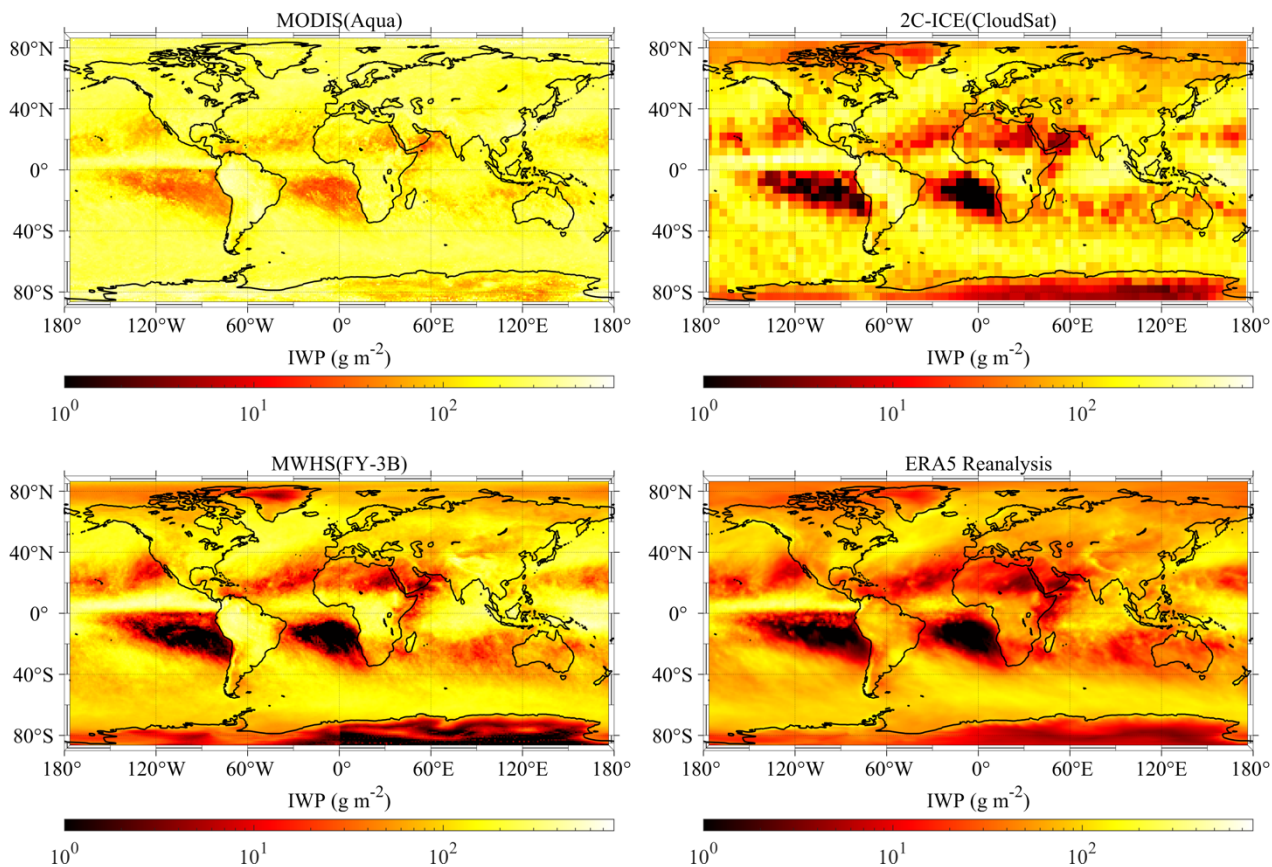
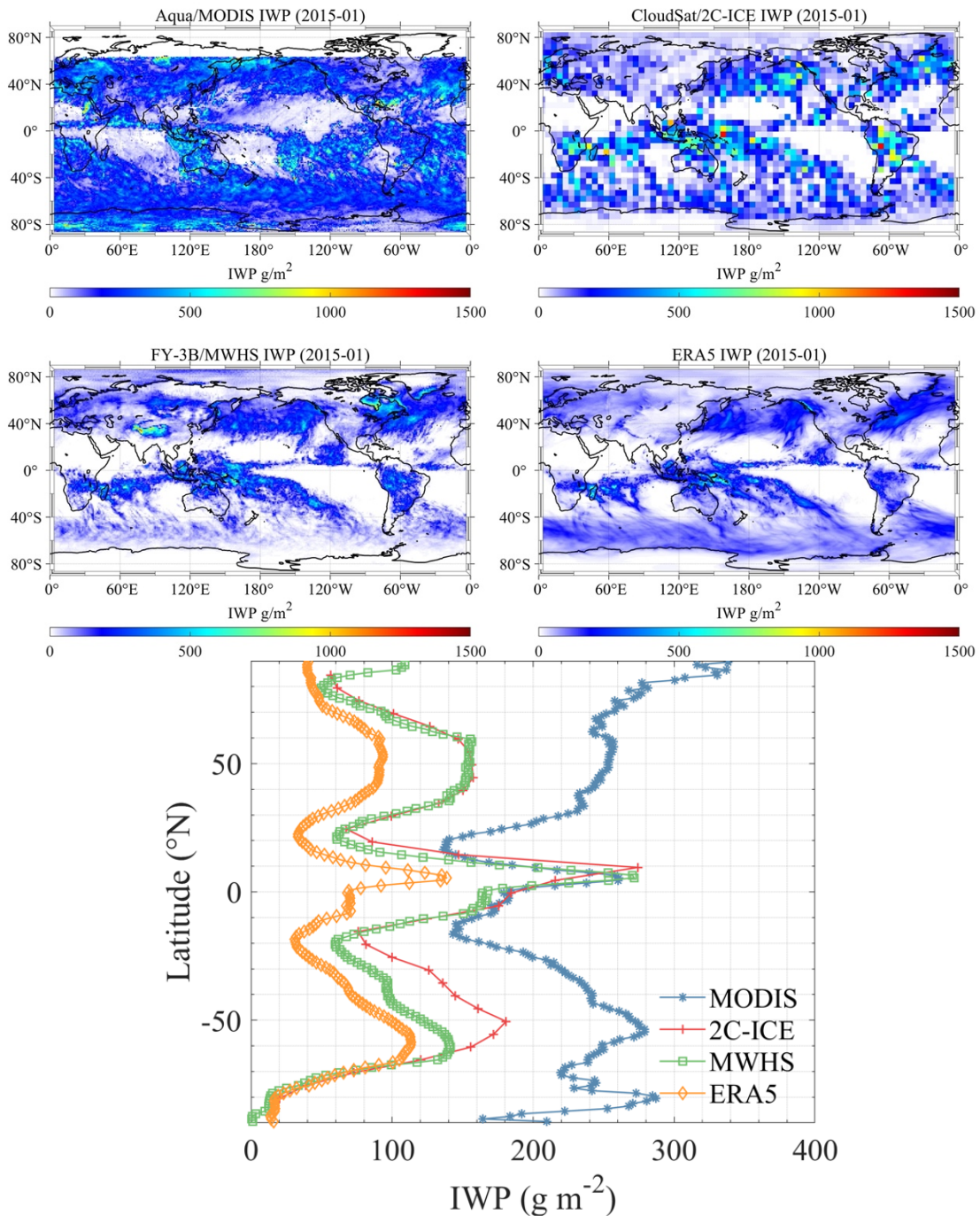



Figure 10. IWP comparison of MWHS and 2C-ICE at the tropical cyclone Bansi. ~~Figure 14. MWHS IWP retrieval results at two typhoon scenes shown in Fig. 11 and Fig. 13. The white regions are the clear-sky distinguished by the cloud filtering network.~~

4.3.2 Global mean IWP comparison

Figure 11 shows the global mean IWP for 2015~~Figure 15 and Figure 16 show the monthly global mean IWP for summer (July 2014) and winter (January 2015)~~ from Aqua/MODIS L3 product (MYD08_M3, C61, Platnick et al., ~~2015~~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.

~~Since 2C-ICE is used to train the networks, MWHS IWP is certainly approaching the 2C-ICE. However, due to the low cover of CloudSat, the monthly mean IWP of 2C-ICE with 5° grid still lack sufficient observations. The MWHS IWP is in general consistent with the range and trend of the ERA5 IWP while there are two obvious regions with large IWP found by MWHS in Tibetan Plateau and Hudson Bay, January 2015. Compared to the MODIS IWP, MWHS IWP is lower in parts of the midlatitudes and polar areas. This may be due to the different measurement resolutions or the bright surfaces which is covered by snow and ice. Because there is no independent in situ measurement, it is difficult to validate which one is more correct. From an overall perspective, it is indicated that the performance of the cloud filtering and IWP retrieval network is good.~~

475   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 15. Monthly global mean IWP 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.

480   Figure 12. Zonal means of IWP for 2015. ~~Figure 16. Same as Fig. 15 but for January 2015.~~

4.4 Discussion

Ice cloud misidentification is an important and unavoidable problem in this study. One reason is that the microwave channels detect ice clouds through the large decrease in TB. However, the low temperature in high altitude regions or other temperature anomaly phenomena can also lead to low TB. In the final results above, although the geographic information is added to the training data, there are still many misclassification cases, such as on the Tibetan Plateau in winter. Therefore, knowing the surface temperature or the near-surface air temperature will help the ice cloud detection. The other reason is due to the mismatch between the CloudSat and the MWHS footprints spatially and temporarily. Since the CloudSat pixels only cover less than 15% of the MWHS pixel, the 2C-ICE scenes cannot fully represent the MWHS observations, especially in the case of thin clouds.

For the IWP retrieval, the 150 GHz window channel has a significant ice cloud response which in combination with 183 GHz channels provides a better retrieval of IWP. The PD at 150 GHz, although contaminated by polarization from the ocean surface, also contributes positively to the retrieval especially when the IWP is larger than 1000 g m^{-2}. In addition, the PD of quasi-polarization channels from MWHS is related to the scan angle and does not fully represent the polarization information of the ice particles, especially near the 45° scan angle. From the perspective of polarization measurements only, a cross-track scanner does not provide as much polarization information as a conical scanner but is more convenient for data assimilation.

However, there are some limitations to 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.

From the error curves shown in Fig. 7-9, the trend is the same for almost all combinations: the errors are largest at IWP =100 g/m², then decrease rapidly and keep smooth, and finally, there is a small increase again at IWP >2000 g/m². The largest error (at IWP =100 g/m²) is supposed to be caused by the insensitivity of the MWHS channels to the thin clouds. The increasing error at IWP >2000 g/m² could be due to the insufficient training of the network (fewer samples). Ice cloud misidentification is an unavoidable problem mainly due to the mismatch between CloudSat and MWHS footprints spatially and temporarily. Since CloudSat pixels only cover less than 15% of the MWHS pixel, the 2C-ICE scenes cannot fully represent the MWHS observations, especially in the case of thin clouds. However, the results show that the neural network

does well in IWP retrieval, with the mean difference from 2C-ICE at about 70%, which is acceptable for passive radiometer
observation. The wide coverage and day/night availability of MWHS facilitates continuous and global IWP observation. PD
at 150 GHz seems not to play a significant role in the overall retrieval errors, this is because the PD mainly occur at
IWP >1000 g/m², at which the error is smaller than the error at IWP =100 g/m² and the samples of IWP >1000 g/m² are also
much smaller, so the effect of PD is greatly diluted in the overall average. In addition, the PD of quasi-polarization channels
is related to the scan angle and does not fully represent the polarization information of the ice particles, especially near the
45° scan angle. However, from the error bins, PD does have a positive effect on retrieval.

However, there are also some difficulties of IWP retrieval using neural networks, including the representative of the
collocations, the requirement of a sufficient number of samples, the unexplainable of the network hyperparameters, etc. And
in the case of not being able to obtain a large number of true samples, the use of neural networks can only converge to a
certain observed product which has the highest accuracy. If the model simulation results (typical profile) are used for
training, the generalization ability of the network will strongly depend on the model itself and the input fields. The
microwave band below 200 GHz is only sensitive to large ice particles and thick clouds, adding the infrared channels like
SPARE-ICE does can improve the retrieval. Information about surface temperature and emissivity also seem important for
IWP retrieval.

5 Conclusions

In this paper, an analysis of global IWP retrieval from FY-3B/MWHS radiance measurements based on neural networks is
presented. MWHS onboard FY-3B satellite has two quasi-polarization channels at 150 GHz which can provide more
information about ice clouds. was overlooked in previous studie. For IWP retrieval, CloudSat/2C-ICE is chosen as the
reference data set for neural networks because it is publicly available and it meets the requirements in terms of data numbers
and measurement accuracy. The collocation between MWHS and 2C-ICE is performed by using a window with 15 min in
time and 15 km in space (a radius of 7.5 km). Comparing the TB and PD at cloudy and clear-sky scenes, PD caused by ice
particles can be easily identified at IWP >1000 g/m² with a typical value of 5 K. Two types of networks are trained to
retrieve the IWP from MWHS measurements. Two types of networks (cloud filtering and IWP retrieval) are trained using
the collocation data set of MWHS and 2C-ICE. A cloud filtering network is trained to classify the cloudy and clear-sky
scenes. For the IWP threshold of 100 $g m^{-2}$ g/m², all channels of MWHS show sensitivity to ice clouds, and CH. 4 is the most
powerful for cloud detection. The FAR and POD of the final network are 0.26 and 0.63, respectively. cloud filtering is 86.48%
and 94.22% respectively. IWP retrieval networks with different combinations of channels and auxiliary information as input
are compared to find the best retrieval strategy. The retrieval results show that adding the 150 GHz channel gives an obvious
improvement in IWP retrieval and the PD also make a positive impact. Comparing the MWHS IWP with 2C-ICE, the CC =
0.65, RMSE = 916.76 $g m^{-2}$, MAPE = 92.90%, and BIAS = -213.12 $g m^{-2}$. Applying the networks to the cyclone Bansi, the
results show a relatively high correlation (0.73) between MWHS IWP and 2C-ICE. The 2015 annual mean IWP from

MWHS shows a similar overall shape to that of MODIS, 2C-ICE and ERA5, and is very close to 2C-ICE in magnitude making the retrieval IWP more credible.

~~The trained two networks show good performance in the application cases of Typhoon Rammasun and Vongfong. The monthly mean IWP from MWHS is similar to that of CloudSat and ERA5, which makes the retrieval results more credible.~~
550 ~~Compared to the IWP product from MODIS, the MWHS IWP is significantly lower in the middle to high latitudes.~~

Neural networks are widely used to statistically characterize the mapping between radiometric measurements and related geophysical variables. The advantages of neural networks are their simplicity and ease of use, their ability to effectively learn the complex nonlinear mapping relationships in samples, and their better robustness to noisy data. By using the collocated measurements, there is no need to establish a complicated radiative transfer model with many possible sources of
555 error. Although the retrieval accuracy can never be as good as 2C-ICE, the spatial and temporal coverage will be much larger which is important for long time series of climate research.

Code and data availability. FY-3B MWHS data can be downloaded from http://satellite.nsmc.org.cn/portalsite/. CloudSat 2C-ICE product can be downloaded from https://www.cloudsat.cira.colostate.edu/data-products. Aqua/MODIS L3 product can be downloaded from https://ladsweb.modaps.eosdis.nasa.gov/search/order. ERA5 reanalysis data can be downloaded from
560 https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset.

Author contribution. Zhenzhan Wang and Wenyu Wang designed the study. Wenyu Wang performed the implementation and wrote the manuscript. Qiurui He and Lanjie Zhang provided the training data and established the network model. Zhenzhan Wang revised the article.

Competing interests. The authors declare that they have no conflict of interest.

565 *Acknowledgements.* The authors would like to thank National Satellite Meteorological Center, China Meteorological Administration for providing the FY-3B MWHS data. The authors thank the CloudSat and CALIPSO science teams for their hard efforts in providing the 2C-ICE product. The authors also thank the MODIS and ERA5 teams for providing the Aqua/MODIS L3 product and ERA5 reanalysis data. We would also like to thank the reviewers and the editors for their valuable and helpful suggestions.

Financial support. This work was supported by the National Natural Science Foundation of China under Grant No. 41901297 and
570 No. 42105130, the Science and Technology Key Project of Henan Province under Grant No. 202102310017.

References

Austin, R. T., Heymsfield, A. J., and Stephens, G. L.: Retrieval of ice cloud microphysical parameters using the CloudSat millimeter-wave radar and temperature, J. Geophys. Res., 114, D00A23, https://doi.org/10.1029/2008JD010049, 2009.

Bennartz, R. and Bauer, P.: Sensitivity of microwave radiances at 85-183 GHz to precipitating ice particles: SENSITIVITY
575 OF MICROWAVE RADIANCES, Radio Sci., 38, n/a-n/a, https://doi.org/10.1029/2002RS002626, 2003.

Boukabara, S.-A., Garrett, K., Chen, W., Iturbide-Sanchez, F., Grassotti, C., Kongoli, C., Chen, R., Liu, Q., Yan, B., Weng, F., Ferraro, R., Kleespies, T. J., and Meng, H.: MiRS: An All-Weather 1DVAR Satellite Data Assimilation and Retrieval System, IEEE. T. Geosci. Remote., 49, 3249–3272, https://doi.org/10.1109/tgrs.2011.2158438, 2011.

Brath, M., Fox, S., Eriksson, P., Harlow, R. C., Burgdorf, M., and Buehler, S. A.: Retrieval of an ice water path over the ocean from ISMAR and MARSS millimeter and submillimeter brightness temperatures, Atmos. Meas. Tech., 11, 611–632, https://doi.org/10.5194/amt-11-611-2018, 2018.

Brath, M., Ekelund, R., Eriksson, P., Lemke, O., and Buehler, S. A.: Microwave and submillimeter wave scattering of oriented ice particles, Atmos. Meas. Tech., 13, 2309–2333, https://doi.org/10.5194/amt-13-2309-2020, 2020.

Buehler, S. A., Jimenez, C., Evans, K. F., Eriksson, P., Rydberg, B., Heymsfield, A. J., Stubenrauch, C. J., Lohmann, U., Emde, C., John, V. O., Sreerekha, T. R., and Davis, C. P.: A concept for a satellite mission to measure cloud ice water path, ice particle size, and cloud altitude, Q. J. Roy. Meteor. Soc., 133, 109–128, https://doi.org/10.1002/qj.143, 2007.

Buehler, S. A., Defer, E., Evans, F., Eliasson, S., Mendrok, J., Eriksson, P., Lee, C., Jiménez, C., Prigent, C., Crewell, S., Kasai, Y., Bennartz, R., and Gasiewski, A. J.: Observing ice clouds in the submillimeter spectral range: the CloudIce mission proposal for ESA's Earth Explorer 8, Atmos. Meas. Tech., 5, 1529–1549, https://doi.org/10.5194/amt-5-1529-2012, 2012.

Coy, J. J., Bell, A., Yang, P., and Wu, D. L.: Sensitivity Analyses for the Retrievals of Ice Cloud Properties From Radiometric and Polarimetric Measurements in Sub-mm/mm and Infrared Bands, J. Geophys. Res. Atmos., 125, https://doi.org/10.1029/2019JD031422, 2020.

Defer, E., Galligani, V. S., Prigent, C., and Jimenez, C.: First observations of polarized scattering over ice clouds at close-to-millimeter wavelengths (157 GHz) with MADRAS on board the Megha-Tropiques mission, J. Geophys. Res. Atmos., 119, 12,301-12,316, https://doi.org/10.1002/2014jd022353, 2014.

Delanoë, J. and Hogan, R. J.: Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds, J. Geophys. Res., 115, D00H29, https://doi.org/10.1029/2009JD012346, 2010.

Deng, M., Mace, G. G., Wang, Z., and Okamoto, H.: Tropical Composition, Cloud and Climate Coupling Experiment validation for cirrus cloud profiling retrieval using CloudSat radar and CALIPSO lidar, J. Geophys. Res., 115, D00J15, https://doi.org/10.1029/2009JD013104, 2010.

Deng, M., Mace, G. G., Wang, Z., and Lawson, R. P.: Evaluation of Several A-Train Ice Cloud Retrieval Products with In Situ Measurements Collected during the SPARTICUS Campaign, J. Appl. Meteorol. Clim., 52, 1014–1030, https://doi.org/10.1175/JAMC-D-12-054.1, 2013.

Duncan, D. I. and Eriksson, P.: An update on global atmospheric ice estimates from satellite observations and reanalyses, Atmos. Chem. Phys., 18, 11205–11219, https://doi.org/10.5194/acp-18-11205-2018, 2018.

Eliasson, S., Buehler, S. A., Milz, M., Eriksson, P., and John, V. O.: Assessing observed and modelled spatial distributions of ice water path using satellite data, Atmos. Chem. Phys., 11, 375–391, https://doi.org/10.5194/acp-11-375-2011, 2011.

Eliasson, S., Holl, G., Buehler, S. A., Kuhn, T., Stengel, M., Iturbide-Sanchez, F., and Johnston, M.: Systematic and random errors between collocated satellite ice water path observations: BIAS AND RANDOM ERRORS OF OBSERVED IWP, J. Geophys. Res. Atmos., 118, 2629–2642, https://doi.org/10.1029/2012JD018381, 2013.

Eriksson, P., Ekstroem, M., Rydberg, B., and Murtagh, D. P.: First Odin sub-mm retrievals in the tropical upper troposphere: ice cloud properties, Atmos. Chem. Phys., 7, 471–483, https://doi.org/10.5194/acp-7-471-2007, 2007.

Eriksson, P., Ekelund, R., Mendrok, J., Brath, M., Lemke, O., and Buehler, S. A.: A general database of hydrometeor single scattering properties at microwave and sub-millimetre wavelengths, Earth. Syst. Sci. Data, 10, 1301–1326, https://doi.org/10.5194/essd-10-1301-2018, 2018.

Eriksson, P., Rydberg, B., Mattioli, V., Thoss, A., Accadia, C., Klein, U., and Buehler, S. A.: Towards an operational Ice Cloud Imager (ICI) retrieval product, Atmos. Meas. Tech., 13, 53–71, https://doi.org/10.5194/amt-13-53-2020, 2020.

Field, P. R. and Heymsfield, A. J.: Importance of snow to global precipitation, Geophys. Res. Lett., 42, 9512–9520, https://doi.org/10.1002/2015GL065497, 2015.

Fox, S.: An Evaluation of Radiative Transfer Simulations of Cloudy Scenes from a Numerical Weather Prediction Model at Sub-Millimetre Frequencies Using Airborne Observations, Remote Sens., 12, 2758, https://doi.org/10.3390/rs12172758, 2020.

Gong, J. and Wu, D. L.: Microphysical properties of frozen particles inferred from Global Precipitation Measurement (GPM) Microwave Imager (GMI) polarimetric measurements, Atmos. Chem. Phys., 17, 2741–2757, https://doi.org/10.5194/acp-17-2741-2017, 2017.

Gong, J., Zeng, X., Wu, D. L., and Li, X.: Diurnal Variation of Tropical Ice Cloud Microphysics: Evidence from Global Precipitation Measurement Microwave Imager Polarimetric Measurements, Geophys. Res. Lett., 45, 1185–1193, https://doi.org/10.1002/2017GL075519, 2018.

Gong, J., Zeng, X., Wu, D. L., Munchak, S. J., Li, X., Kneifel, S., Ori, D., Liao, L., and Barahona, D.: Linkage among ice crystal microphysics, mesoscale dynamics, and cloud and precipitation structures revealed by collocated microwave radiometer and multifrequency radar observations, Atmos. Chem. Phys., 20, 12633–12653, https://doi.org/10.5194/acp-20-12633-2020, 2020.

He, J. and Zhang, S.: Research on cirrus clouds in Tibetan Plateau using MWHS onboard Chinese FY3B/C meteorological satellite, in: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, Beijing, China, 11–15 July 2016, 581–584, https://doi.org/10.1109/IGARSS.2016.7729145, 2016.

Heymsfield, A., Krämer, M., Wood, N. B., Gettelman, A., Field, P. R., and Liu, G.: Dependence of the Ice Water Content and Snowfall Rate on Temperature, Globally: Comparison of in Situ Observations, Satellite Active Remote Sensing Retrievals, and Global Climate Model Simulations, J. Appl. Meteorol. Clim., 56, 189–215, https://doi.org/10.1175/JAMC-D-16-0230.1, 2017.

Holl, G., Buehler, S. A., Rydberg, B., and Jiménez, C.: Collocating satellite-based radar and radiometer measurements – methodology and usage examples, Atmos. Meas. Tech., 3, 693–708, https://doi.org/10.5194/amt-3-693-2010, 2010.

Holl, G., Eliasson, S., Mendrok, J., and Buehler, S. A.: SPARE-ICE: Synergistic ice water path from passive operational sensors, J. Geophys. Res. Atmos., 119, 1504–1523, https://doi.org/10.1002/2013JD020759, 2014.

King, M. D., Tsay, S.-C., Platnick, S. E., Wang, M., & Liou, K.-N.: Cloud retrieval algorithms for MODIS: Optical thickness, effective particle radius, and thermodynamic phase, ATBD-MOD-05, NASA, Washington, D. C., 1998

Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

650   Liou, K.-N.: Influence of Cirrus Clouds on Weather and Climate Processes: A Global Perspective, Mon. Weather Rev., 114, 1167–1199, https://doi.org/10.1175/1520-0493(1986)114<1167:IOCCOW>2.0.CO;2, 1986.

Liu, G. and Curry, J. A.: Remote Sensing of Ice Water Characteristics in Tropical Clouds Using Aircraft Microwave Measurements, J. Appl. Meteorol., 37, 337–355, https://doi.org/10.1175/1520-0450(1998)037<0337:RSOIWC>2.0.CO;2, 1998.

655   Liu, G. and Curry, J. A.: Determination of Ice Water Path and Mass Median Particle Size Using Multichannel Microwave Measurements, J. Appl. Meteorol., 39, 1318–1329, https://doi.org/10.1175/1520-0450(2000)039<1318:DOIWPA>2.0.CO;2, 2000.

Mace, G. G., Zhang, Q., Vaughan, M., Marchand, R., Stephens, G., Trepte, C., and Winker, D.: A description of hydrometeor layer occurrence statistics derived from the first year of merged Cloudsat and CALIPSO data, J. Geophys.
660   Res., 114, D00A26, https://doi.org/10.1029/2007JD009755, 2009.

Mace, G. G, & Deng, M.: Level 2 CloudSat-CALIPSO Combined Ice Cloud Property Retrieval Product Process Description and Interface Control Document. NASA CloudSat Project Rep. P1_R05, NASA, Washington, D. C., 2019

Millán, L., Read, W., Kasai, Y., Lambert, A., Livesey, N., Mendrok, J., Sagawa, H., Sano, T., Shiotani, M., and Wu, D. L.: SMILES ice cloud products, J. Geophys. Res. Atmos., 118, 6468–6477, https://doi.org/10.1002/jgrd.50322, 2013.

665   Pfreundschuh, S., Eriksson, P., Buehler, S. A., Brath, M., Duncan, D., Larsson, R., and Ekelund, R.: Synergistic radar and radiometer retrievals of ice hydrometeors, Atmos. Meas. Tech., 13, 4219–4245, https://doi.org/10.5194/amt-13-4219-2020, 2020.

Platnick, S., M. King, P. Hubanks. MODIS Atmosphere L3 Monthly Product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, https://10.5067/MODIS/MYD08_M3.061, 2017Platnick, S., et al.: MODIS Atmosphere L3
670   Monthly Product [Dataset]. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA. http://dx.doi.org/10.5067/MODIS/MOD08_M3.006., 2015.

Stephens, G. L. and Kummerow, C. D.: The Remote Sensing of Clouds and Precipitation from Space: A Review, J. Atmos. Sci., 64, 3742–3765, https://doi.org/10.1175/2006JAS2375.1, 2007.

Stephens, G. L., Vane, D. G., Tanelli, S., Im, E., Durden, S., Rokey, M., Reinke, D., Partain, P., Mace, G. G., Austin, R.,
675   L'Ecuyer, T., Haynes, J., Lebsock, M., Suzuki, K., Waliser, D., Wu, D., Kay, J., Gettelman, A., Wang, Z., and Marchand, R.: CloudSat mission: Performance and early science after the first year of operation, J. Geophys. Res., 113, D00A18, https://doi.org/10.1029/2008JD009982, 2008.

Stubenrauch, C. J., Rossow, W. B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., Girolamo, L. D., Getzewich, B., Guignard, A., Heidinger, A., Maddux, B. C., Menzel, W. P., Minnis, P., Pearl, C., Platnick, S., Poulsen, C., Riedi, J., Sun-
680   Mack, S., Walther, A., Winker, D., Zeng, S., and Zhao, G.: Assessment of Global Cloud DatasetData sets from Satellites: Project and Database Initiated by the GEWEX Radiation Panel, B. Am. Meteorol. Soc., 94, 1031–1049, https://doi.org/10.1175/BAMS-D-12-00117.1, 2013.

Sun, N. and Weng, F.: Retrieval of Cloud Ice Water Path from Special Sensor Microwave Imager/Sounder (SSMIS), J. Appl. Meteorol. Clim., 51, 366–379, https://doi.org/10.1175/JAMC-D-11-021.1, 2012.

685   Wang Y, Fu Y, Fang X, et al. Estimating ice water path in tropical cyclones with multispectral microwave data from the FY-3B satellite[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(9): 5548-5557.

Wang, Z., Zhang, S., Li, J., Li, Y., & Wu, Q.: Thermal/vacuum calibration of microwave humidity sounder on FY-3B satellite. Strategic Study of CAE, 15(10), 33-46+53. https://doi.org/10.3969/j.issn.1009-1742.2013.10.005, 2013.

Weng, F. and Grody, N. C.: Retrieval of Ice Cloud Parameters Using a Microwave Imaging Radiometer, J. Appl. Meteorol., 57, 1069–1081, https://doi.org/10.1175/1520-0469(2000)057<1069:ROICPU>2.0.CO;2, 2000.

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H., and Young, S. A.: Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms, J. Atmos. Ocean. Tech., 26, 2310–2323, https://doi.org/10.1175/2009JTECHA1281.1, 2009.

Wu, D. L., Jiang, J. H., Read, W. G., Austin, R. T., Davis, C. P., Lambert, A., Stephens, G. L., Vane, D. G., and Waters, J. W.: Validation of the Aura MLS cloud ice water content measurements, J. Geophys. Res. Atmos., 113, https://doi.org/10.1029/2007jd008931, 2008.

Wu, D. L., Austin, R. T., Deng, M., Durden, S. L., Heymsfield, A. J., Jiang, J. H., Lambert, A., Li, J. L., Livesey, N. J., McFarquhar, G. M., Pittman, J. V., Stephens, G. L., Tanelli, S., Vane, D. G., and Waliser, D. E.: Comparisons of global cloud ice from MLS, CloudSat, and correlative data sets, J. Geophys. Res. Atmos., 114, https://doi.org/10.1029/2008jd009946, 2009.

Zhao, L. and Weng, F.: Retrieval of Ice Cloud Parameters Using the Advanced Microwave Sounding Unit, J. Appl. Meteorol., 41, 384–395, https://doi.org/10.1175/1520-0450(2002)041<0384:ROICPU>2.0.CO;2, 2002.

Zou, X., Chen, X., and Weng, F.: Polarization signature from the FengYun-3 Microwave Humidity Sounder, Front. Earth Sci., 8, 625–633, https://doi.org/10.1007/s11707-014-0479-y, 2014.