Retrieval of Oceanic Total Precipitable Water Vapor and Cloud Liquid Water from Fengyun-3D Microwave Sounding Instruments

Yang HAN1,2, Jun YANG1*, Hao HU1, and Peiming DONG1

1 State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, China Meteorological Administration, Beijing 100081
2 National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081

(Received May 12, 2020; in final form October 17, 2020)

ABSTRACT

Fengyun-3D (FY-3D) satellite is the latest polar-orbiting meteorological satellite launched by China and carries 10 instruments onboard. Its microwave temperature sounder (MWTS) and microwave humidity sounder (MWHS) can acquire a total of 28 channels of brightness temperatures, providing rich information for profiling atmospheric temperature and moisture. However, due to a lack of two important frequencies at 23.8 and 31.4 GHz, it is difficult to retrieve the total precipitable water vapor (TPW) and cloud liquid water path (CLW) from FY-3D microwave sounder data as commonly done for other microwave sounding instruments. Using the channel similarity between Suomi National Polar-orbiting Partnership (NPP) advanced technology microwave sounder (ATMS) and FY-3D microwave sounding instruments, a machine learning (ML) technique is used to generate the two missing low-frequency channels of MWTS and MWHS. Then, a new dataset named as combined microwave sounder (CMWS) is obtained, which has the same channel setting as ATMS but the spatial resolution is consistent with MWTS. A statistical inversion method is adopted to retrieve TPW and CLW over oceans from the FY-3D CMWS. The intercomparison between different satellites shows that the inversion products of FY-3D CMWS and Suomi NPP ATMS have good consistency in magnitude and distribution. The correlation coefficients of retrieved TPW and CLW between CMWS and ATMS can reach 0.95 and 0.85, respectively.

Key words: precipitable water vapor, cloud liquid water, Fengyun-3D (FY-3D), microwave, machine learning (ML)

Citation: Han, Y., J. Yang, H. Hu, et al., 2021: Retrieval of oceanic total precipitable water vapor and cloud liquid water from Fengyun-3D microwave sounding instruments. J. Meteor. Res., 35(2), 371–383, doi: 10.1007/s13351-021-0084-4.

1. Introduction

Clouds and water vapor are vital for regulating the global radiation budget, and their distribution and evolution in the atmosphere play an extremely important role in global weather and climate change (Stephens, 2005; Yang et al., 2016, 2017, 2018). The total precipitable water vapor (TPW) and cloud liquid water path (CLW) are two key physical quantities for assessing cloud water resources in the air. The radiosonde observation (RAOB) can obtain direct measurement of water vapor, but the operational RAOB measurements are only performed twice a day at several hundred fixed stations around the world, which makes its temporal and spatial resolution have great limitations (Gui et al., 2017). Therefore, remote sensing inversion has become an important means to obtain TPW and CLW. Both ground-based and satellite retrieval methods have been widely developed in the microwave range to obtain local or global cloud water content. Ground-based microwave radiometer can obtain continuous TPW and CLW measurements with a very high temporal resolution for a stationary observation site (Lin et al., 2001; Cadeddu et al., 2009; Cimini et al., 2010; Yang and Min, 2018). The limitation of ground observation lies in its poor spatial distribution, especially over ocean.

Compared with ground-based observation, satellite measurement can provide TPW and CLW inversion...
products on a global scale although its temporal resolution is relatively low. The large-scale TPW and CLW distribution obtained by satellite inversion is used for global cloud water resource assessment; in addition, another important function is to use as cloud detection criterion for quality control in satellite data assimilation. The brightness temperature measured by satellites is affected by many factors such as surface emissivity, water vapor, and cloud liquid water content. On the ocean, the surface emissivity is relatively uniform. Normally, TPW and CLW can be retrieved by using two microwave channel measurements. In the early days of satellite applications, Grody et al. (1980) applied the data from 21- and 31-GHz channels of scanning microwave spectrometer mounted on Nimbus 6 to retrieve CLW over the Pacific Ocean. Prabhakara et al. (1983) developed a method to retrieve CLW from the data on Nimbus 7 scanning multichannel microwave radiometer at 6.6 and 10.7 GHz. After the launch of special sensor microwave/imager (SSM/I) aboard the defense meteorological satellite program, a large number of satellite inversion algorithms based on this passive microwave sensor were reported to obtain globe distribution of TPW and CLW over ocean (Greenwald et al., 1993; Liu and Curry, 1993; Weng and Grody, 1994; Ferraro et al., 1996). There are four frequencies (at 19.35, 22.35, 37, and 85.5 GHz) in SSM/I, and the existing algorithms mainly adopt the combination of 22.235 GHz and another frequency to invert TPW and CLW. The successive launch of advanced microwave sounding unit and advanced technology microwave sounder (ATMS) further promoted the application of satellite inversion of TPW and CLW over ocean (Dyras and Serafin-Rek, 2002; Mo and Liu, 2008; Weng et al., 2012). Both statistical (Grody et al., 2001) and physical inversion (Weng et al., 2003) algorithms had been proposed to improve the inversion accuracy of TPW and CLW based on two window channels (at 23.8 and 31.4 GHz, respectively).

Fengyun-3D (FY-3D) satellite, which was launched in November 2017, is the fourth satellite of China’s second-generation polar-orbiting satellite. There are 10 sets of advanced remote sensing devices on the FY-3D, among which microwave temperature sounder (MWTS) and microwave humidity sounder (MWHS) are mainly designed to obtain atmospheric temperature and water vapor profiles (Yang et al., 2019). The MWTS operates in an oxygen absorption band of 50–60 GHz and is subdivided into 13 channels. There are total 15 channels in MWHS, which are located in oxygen absorption zone (8 channels near 118.75 GHz), water vapor absorption zone (5 channels near 183.31 GHz), and window zone (at 89 and 150 GHz). So far, FY-3D’s microwave sounding data still fail to be assimilated into numerical weather prediction system, and an important reason is the lack of effective quality control methods to obtain clear sky pixels. Using the dual oxygen absorption bands in MWTS and MWHS, several pairs of oxygen channels can be applied to compute cloud emission and scattering index at different height levels (Han et al., 2015), where each pair of oxygen channels has a similar peak weighting function height. The strong water vapor absorption channels at high frequency in MWHS are very sensitive to small changes of water vapor, and it is only suitable for inversion of TPW below 7 kg m$^{-2}$ (Melsheimer and Heygster, 2008). Therefore, it is often used for water vapor retrieval in the polar regions. Two window channels in MWHS can be used to retrieve cloud scattering index (Bennartz et al., 2002) and cloud ice water path (Zhao and Weng, 2002). Although some cloud water information can be obtained by using different algorithms based on the existing microwave channels of FY-3D, it is still very difficult to give specific values of TPW and CLW. The main reason is the lack of two important window channels, 23.8 and 31.4 GHz, in FY-3D.

In this paper, we introduced a channel simulation algorithm from the machine learning (ML) to generate these two window channels, which makes it possible to retrieve TPW and CLW from FY-3D microwave sounding instruments. In the following section, the channel simulation method based on ML is first introduced. Some ML simulation results and accuracy assessments are presented in Section 3. In Section 4, the inversion results of TPW and CLW based on FY-3D are compared with those of ATMS. The brief conclusions and recommendations for future work are discussed in Section 5.

2. Channel simulation algorithm based on ML

For any field of view (FOV), the measurements among different channels on the same satellite sensor should have a certain correlation, because their corresponding surface types and atmospheric environment are exactly the same. Therefore, we can establish the relationship between two low-frequency window channels and other channels of ATMS through the ML model for all FOVs. Since MWTS and MWHS contain all channel settings in ATMS except for the two low-frequency window channels, we can match FY-3D data to ATMS level by cross calibration, thus realizing the prediction of missing channel values in FY-3D using ATMS training model. It should be pointed out that the channel simulation...
method proposed in this paper is mainly to facilitate the transplantation of ATMS inversion algorithm to FY-3D data, but the two simulated channels can not be used to replace the real observation because they only combine the existing observation information of other channels. The flow chart of the proposed channel simulation algorithm is shown in Fig. 1, which includes three main steps: footprint matching, cross calibration, and ML. First, each channel of MWHS transforms to a new channel consistent with the spatial resolution of MWTS by performing footprint matching with MWTS. Then, we can obtain a 28-channel dataset, 20 of which have the same channel properties as ATMS. Second, these 20 channels will be cross-calibrated with the corresponding channels in ATMS. Third, the typical ATMS samples are trained by using ML method, and the model relationship between other channels and low-frequency window channels in ATMS can be established. Next, the corresponding 20 channels in FY-3D are input into the model to obtain two simulated low-frequency channels. It will combine 20 channels in FY-3D to form a new dataset, named combined microwave sounder (CMWS), which is identical to the channel settings of the ATMS, but has the same observation range and resolution as the MWTS. Several key steps will be introduced in the following subsections.

### 2.1 Footprint matching

Although both MWTS and MWHS are mounted on the FY-3D satellite, they are two completely independent sensors, which also result in the differences of FOV size and FOV number on each scan line between MWTS and MWHS. For example, there are 98 FOVs on each scan line in MWHS, while only 90 FOVs in MWTS. In order to make the pixels corresponding to MWTS and MWHS have the same observation position and instantaneous FOV, we need to perform footprint matching for the original observation data. Since the spatial resolution of MWHS is higher than that of MWTS, an alternative method is to use B–G method (Backus and Gilbert, 1968) to re-sample MWHS to produce observations with the same resolution as MWTS. However, on the FY-3D satellite, the scan times of MWTS and MWHS do not match well, which make the brightness temperatures observed by them not overlap well in space. Therefore, we adopted a simple weighted average to match the MWHS observations to the MWTS resolution level. By setting a distance threshold (equal to half FOV of each MWTS pixel), for each FOV in the MWTS, we find all points below the threshold in the MWHS, and calculate the average of these points to obtain the matched MWHS brightness temperature values. Figure 2 shows two typical MWHS channels (89 and 183.31 ± 1.0 GHz) on 9 July 2018 and their results after footprint matching with MWTS. The left column represents the original measurements of MWHS, and the right column is the results after matched. The matched results are very close to the original observations in both intensity and distribution of brightness temperature, which ensures that our footprint matching does not lose too much information.

### 2.2 Cross calibration

Although the channel settings between ATMS and CMWS are basically the same, the brightness temperatures of ATMS and CMWS come from different satellite measurements, which inevitably lead to a certain deviation between the two observations due to the differences in hardware setting and radiometric calibration. Since our training samples are all from ATMS, and the input data of the model prediction come from CMWS, it becomes indispensable to perform a cross-calibration process, which will ensure that the measured values on each channel between the two instruments are as consistent as possible.

The sub-satellite trajectories of Suomi National Polar-orbiting Partnership (NPP) and FY-3D satellites are very close to each other on 1–2 February 2018, which allows us to use the data of these two days to cross-calibrate ATMS and CMWS. For each FOV in CMWS, the arc length and observation time difference between the pixel and all observation points in ATMS are calculated. The pair of FOV with the shortest distance may be the observation of the same point by two sensors. By setting a dis-
tance threshold and a time threshold, all matching pairs of FOVs satisfying the threshold conditions can be considered as the same observation point and used for cross calibration. Successfully matched pixel pairs need to meet the following parameters: imaging time difference is shorter than 30 min, space distance is less than 15 km, satellite height angle difference is less than 10°, and scanning angle difference is no more than 20°. A simple linear regression method is adopted to derive a linear regression equation for each channel, which can match CMWS observations to the same level as ATMS. Table 1 provides the linear fitting coefficients (slope and intercept) and mean absolute error (MAE) for each channel.

2.3 ML

After completing the footprint matching of MWTS and MWHS and the cross calibration between ATMS and CMWS, the next key step of channel simulation is the application of ML algorithm. ML can be divided into supervised learning and unsupervised learning (Russell and Norvig, 2010). Under supervised learning, each group of training data has a clear mark or result. In unsupervised learning, data are not specifically identified, and learning model is designed to infer some of the intrinsic structure of data. The issue to be solved here is a regression problem, so supervised learning should be adopted to train the model. Recently, the ensemble algorithms have been developed, which solve the single prediction problem by establishing a combination of several models (Rokach, 2010). It works by generating multiple classifiers/models, each of which learns and makes predictions independently. These predictions are ultimately combined into a single prediction, which is usually better than any single estimator. Random forest (RF) is an ML algorithm that integrates multiple trees by the idea of ensemble learning (Breiman, 2001). Its basic unit is the decision tree, and each tree is bootstrapping random sampling from the training set. There is no correlation between each decision tree in an RF model. The basic schematic diagram of the RF is shown in Fig. 3. RF has...
the following main advantages: 1) training can be highly parallelized, which makes training speed of the model have great advantages, especially for large samples such as satellite data; 2) because the decision tree node splitting feature can be selected randomly, the training model can still be efficient even if the dimension of sample features is very high; and 3) the training model has small variance and strong generalization ability because of random sampling.

3. Accuracy evaluation and channel simulation examples

In ATMS, there are remaining 20 channels besides two low-frequency window channels. To determine whether all 20 channels should be involved in the training of the RF model, we conducted some sensitivity tests based on a very popular ML software package scikit-learn (https://scikit-learn.org/stable/index.html). In order to facilitate the pixel-by-pixel quantitative evaluation of the simulation results, the data in the training set and the test set are all from ATMS. In order to increase the representativeness of training samples, our training set consists of two categories; one is to select one-day data from each month in the past year, which mainly represents the basic global climate trends in different seasons. The other type is the data of the day before the forecast sample, which mainly on behalf of the global approaching weather system. Here, the 12-day global climate data and the approaching weather data on 1 February 2018 were chosen as the training set, and the data of 2 February 2018 were adopted as the test sample. By comparing the simulation results with the actual observations, the MAE of the two low-frequency window channels (represented by Ch1 and Ch2, respectively) in different channel combinations and different model parameters can be obtained. The simulated error of each single channel is displayed in Fig. 4a. It is clear that MAEs of channels 3, 4, 5, and 16 are significantly lower than those of other channels. Considering that the weighting functions and peak heights of these four channels are very similar to those of Ch1 and Ch2 (see Fig. 2 in Weng et al., 2012), it has natural advantages to simulate Ch1 and Ch2 using these four channels. In addition to these four channels, other channels in ATMS are added to the training model one by one to test their effects on the simulation accuracy of Ch1 and Ch2. Figure 4b denotes different MAEs when each channel is progressively added to the training model. The trend of simulation accuracy for Ch1 and Ch2 is basically the same. When the channel used in the model increases from channels 6 to 11, the MAE is decreasing; and from channels 12 to 15, the MAE is starting to increase again. Channels 17–22 are high frequency channels in ATMS. When channels 17 and 18 are added to the training model, the simulation errors of Ch1 and Ch2 can be further reduced, but the introduction of channels 19–22 makes the MAEs of Ch1 and Ch2 slightly increased again. These sensitivity tests show that channels 12–15 and 19–22 cannot play a positive role in the simulation of Ch1 and Ch2. Therefore, in the next

| Channel | ATMS Center frequency (GHz) | CMWS Center frequency (GHz) | Slope | Intercept | MAE (K) |
|---------|-----------------------------|-----------------------------|-------|-----------|--------|
| 1       | 23.8                        | 23.8                        |       | /         | /      |
| 2       | 31.4                        | 31.4                        |       | /         | /      |
| 3       | 50.3                        | 50.3                        | 1.0124| -4.6015   | 2.7139 |
| 4       | 51.76                       | 51.76                       | 0.9665| 8.8694    | 1.4912 |
| 5       | 52.8                        | 52.8                        | 0.9980| 1.9720    | 1.7411 |
| 6       | 53.596 ± 0.115              | 53.596 ± 0.115              | 1.0021| 0.0328    | 0.9459 |
| 7       | 54.4                        | 54.4                        | 0.9944| 1.7225    | 0.9955 |
| 8       | 54.94                       | 54.94                       | 0.9533| 10.8517   | 0.7353 |
| 9       | 55.5                        | 55.5                        | 0.9978| -0.7035   | 1.8547 |
| 10      | 57.290344 (fo)              | 57.290344 (fo)              | 1.0258| -6.7331   | 1.8826 |
| 11      | fo ± 0.217                  | fo ± 0.217                  | 1.0201| -5.2703   | 1.6061 |
| 12      | fo ± 0.3222 ± 0.048         | fo ± 0.3222 ± 0.048         | 1.0163| -4.2106   | 1.2949 |
| 13      | fo ± 0.3222 ± 0.022         | fo ± 0.3222 ± 0.022         | 1.0277| -8.5121   | 2.6138 |
| 14      | fo ± 0.3222 ± 0.010         | fo ± 0.3222 ± 0.010         | 0.9827| 4.3781    | 1.4696 |
| 15      | fo ± 0.3222 ± 0.0045        | fo ± 0.3222 ± 0.0045        | 1.0255| -6.6283   | 2.2318 |
| 16      | 88.2                        | 89.0                        | 0.9374| 18.6856   | 5.0148 |
| 17      | 165.5                       | 150.0                       | 0.9392| 22.8063   | 8.2390 |
| 18      | 183.31 ± 7                  | 183.31 ± 7                  | 0.9747| 8.0082    | 1.9217 |
| 19      | 183.31 ± 4.5                | 183.31 ± 4.5                | 1.0631| -10.4440  | 2.9500 |
| 20      | 183.31 ± 3                  | 183.31 ± 3                  | 0.9592| 12.2276   | 1.9781 |
| 21      | 183.31 ± 1.8                | 183.31 ± 1.8                | 0.9311| 19.6531   | 2.5483 |
| 22      | 183.31 ± 1                  | 183.31 ± 1                  | 0.9588| 13.2642   | 3.2390 |
In an RF model, in order to obtain the highest prediction accuracy and reduce the computation time, it is necessary to optimize some input parameters. We did some sensitivity tests to determine the values of these parameters (see Fig. 5). The more decision trees are used in the RF model, the accuracy of the model is higher, but the number of trees seriously affects the training speed. It is a reasonable choice to adopt 130 decision trees in our RF model. The maximum depth of each tree determines the extent of tree expansion. More depth does not significantly increase the accuracy of the model, so we set the maximum depth to 30. The number of features to consider when looking for the best split obviously affects the accuracy of the model, and the simulation error is the smallest when all 12 channels are used. The other parameters, such as the maximum number of leaf nodes, the minimum number of samples required to split an internal node, and the minimum number of samples required to be at a leaf node, can obtain the best simulation accuracy by directly using the default values of the software package scikit-learn.

Using the training channels and ML parameters determined by the aforementioned sensitivity tests, the two ATMS low frequency window channels on 10 July 2018 were simulated. The pixel-to-pixel accuracy evaluation results are shown in Fig. 6. The top row of Fig. 6 depicts the evaluation results of Ch1, and the bottom row of Fig. 6 shows the evaluation results of Ch2. The left column of Fig. 6 shows the difference between observation and sim-
ulation, while the scatter density of observed and simulated brightness temperatures are given in the right column. Visually, the range of error distribution of Ch1 and Ch2 simulation is very similar, which mainly distributed on the high latitude regions, coastlines, and the edge of some heavy rainfall. The errors on the high latitude regions and coastlines are mainly affected by the inhomogeneity of the underlying surface type, which makes the surface emissivity very uncertain. In addition, the simulation errors of the two channels at the edge of the weather system are mainly affected by the uncertainty of cloud and rain scattering. It should be noted that the MAE of Ch1 simulation is slightly bigger than that of Ch2 simulation, corresponding to 3.5252 and 3.4151 K, respectively.

A quantitative assessment of the simulation results of the two channels based on ATMS has been completed previously. We can get Ch1 and Ch2 simulation results based on FY-3D observation data by replacing the predictive input of ML model with FY-3D measurements. Figure 7 shows the simulation results of the two low-frequency window channels based on FY-3D observations on 13 September 2018. The left column is the actual observations of ATMS, and the right column is the simulation results based on FY-3D observations. The upper row is the observation and simulation of 23.8 GHz and the lower row is the related results of 31.4 GHz. It can be seen that the simulation results of the two channels are very similar to the observations of the ATMS, especially on the ocean. On 13 September 2018, Super Typhoon Mangkhut (about at 15°N, 130°E) developed vigorously in the Northwest Pacific Ocean. It is clear that our simulation results for both Ch1 and Ch2 can give the accurate locations and basic form of the typhoon, except that the simulated brightness temperature in the typhoon area is a bit weak.

It should be pointed out that the results of the simulation using FY-3D will definitely be lower than the accuracy of quantitative evaluation by ATMS. This is because the cross calibration between ATMS and FY-3D will inevitably introduce some new errors (see Table 1). Since ATMS and CMWS have different FOV and satellite transit times, to perform pixel-to-pixel accuracy assessments, we need to collocation all pixels to ensure that the same pixels are evaluated. The specific collocation threshold has been given in Section 2.2. Most of the time, the same observation point that two sensors can match is very limited. To quantitatively evaluate the simulated brightness temperature accuracy when FY-3D observations are input into ML model, a total of 5-day data were selected as data sources from different months. Considering that the simulation of Ch1 and Ch2 has large errors in high latitudes, we only collocated all the ocean pixels from 60°S to 60°N. There are a total of 180,906 pixels used for quantitative evaluation. The top row of Fig. 8 shows the scatter density plots between ATMS and CMWS of two corresponding channels. Quantitative evaluation result for five independent days can be found in Table 2. Overall, the accuracy and stability of two-channel
simulation are satisfactory. According to the 5-day observation results from different months, the correlation coefficient of Ch1 is more than 0.9, and the correlation coefficient of Ch2 is also close to 0.9. The MAEs of the two channels between ATMS and CMWS are 6.74 and 5.73 K, respectively.

4. Retrieval of TPW and CLW

Traditionally, retrieving TPW and CLW from satellite measurements is mainly aimed at the ocean surface because of its relatively low and uniform surface emissivity. Although several methods have been reported to retrieve TPW from land surface (Aires et al., 2001; Boukabara et al., 2010; Zhou et al., 2016), it is still very difficult to retrieve TPW and CLW from non-oceanic surface because of its pixel-by-pixel high uncertainty of surface emissivity. The purpose of this paper is not to propose a new inversion method, mainly to test the feasibility of TPW and CLW inversion based on FY-3D microwave sounding instruments. Therefore, the inversion of TPW and CLW mainly focuses on the ocean. After the high-precision simulation of 23.8- and 31.4-GHz channels based on FY-3D measurements, the inversion of TPW and CLW over ocean becomes relatively easy. There are two inversion methods over ocean widely used in the field of microwave remote sensing: one is Grody’s statistical inversion method (Grody et al., 2001), and the other is Weng’s physical inversion method (Weng et al., 2003). In theory, physical inversion method should have higher retrieval accuracy than statistical inversion method. However, in Weng’s physical inversion model (Weng et al., 2003), some precise ocean and cloud parameters need to be given, such as sea surface temperature, wind speed, and cloud layer temperature. These parameters are usually difficult to obtain precise measurements. An alternative is to obtain them from global forecast system or from reanalysis data. Considering the root mean square error (RMSE) for non-precipitation CLW retrieval is no more than 0.05 mm (Grody et al., 2001), we select Grody’s statistical method to retrieve TPW and CLW over ocean from FY-3D in this paper. The equations for statistical in-

Fig. 6. Pixel-to-pixel accuracy evaluation results on 10 July 2018. (a, b) Ch1 and (c, d) Ch2, and (a, c) difference results and (b, d) scatter density plots.
version can be expressed as,

\[ TPW = \cos(\theta) \left[ a_0 + a_1 \ln(T_s - T_{b23}) + a_2 \ln(T_s - T_{b31}) \right], \]  

(1)

\[ CLW = \cos(\theta) \left[ b_0 + b_1 \ln(T_s - T_{b23}) + b_2 \ln(T_s - T_{b31}) \right], \]  

(2)

where \( \theta \) is the local zenith angle, \( T_s \) is the surface temperature, and \( T_{b23} \) and \( T_{b31} \) represent the observed brightness temperatures of 23.8 and 31.4 GHz, respectively. To ensure that \( T_s \) is larger than \( T_{b23} \) and \( T_{b31} \) over ocean, the value of \( T_s \) is set to 285 K. The coefficients \( a_0, a_1, a_2, b_0, b_1, \) and \( b_2 \) can be obtained by performing regression analysis on the simulated channel measurements. Since we have calibrated the CMWS data to the level of ATMS through cross calibration, these coefficients can be transplanted from the ATMS algorithm directly (Weng et al., 2012). The specific coefficients are as follows,

\[ a_0 = 247.92 - (69.235 - 44.177 \cos(\theta)) \cos(\theta), \]  

(3)

\[ a_1 = -116.27, \quad a_2 = 73.409, \]  

(4)

\[ b_0 = 8.240 - (2.622 - 1.846 \cos(\theta)) \cos(\theta), \]  

(5)

\[ b_1 = 0.754, \quad b_2 = -2.265. \]  

(6)

Using the above formulas and combining the brightness temperature values of 23.8- and 31.4-GHz channels simulated in CMWS, the specific values of TPW and

---

**Table 2.** Quantitative evaluation results between ATMS and CMWS in 2018

| Date       | Matched pixel | Ch1 | Ch2 | TPW | CLW   | Ch1 (K) | Ch2 (K) | TPW (mm) | CLW (mm) | MAE    |
|------------|---------------|-----|-----|-----|-------|---------|---------|----------|----------|--------|
| 2 June     | 54,831        | 0.90| 0.82| 0.94| 0.85  | 7.27    | 8.66    | 5.43     | 0.15     |        |
| 2 July     | 53,322        | 0.94| 0.90| 0.95| 0.89  | 6.75    | 4.40    | 5.34     | 0.08     |        |
| 2 August   | 40,565        | 0.94| 0.90| 0.95| 0.89  | 6.24    | 4.60    | 5.11     | 0.08     |        |
| 2 September| 22,955        | 0.93| 0.88| 0.96| 0.87  | 6.37    | 4.62    | 4.48     | 0.07     |        |
| 2 October  | 8,936         | 0.88| 0.88| 0.89| 0.86  | 6.61    | 3.77    | 4.03     | 0.07     |        |
| Total      | 180,609       | 0.92| 0.85| 0.95| 0.85  | 6.74    | 5.73    | 5.14     | 0.10     |        |

---

Fig. 7. Compared between the simulation and observation on 13 September 2018 at (a, b) 23.8 GHz and (c, d) 31.4 GHz. (a, c) The original measurements for Ch1 and Ch2, and (b, d) simulation results of two corresponding channels based on FY-3D observation.
CLW on the ocean can be obtained efficiently. The validation of TPW retrieved from satellite can be carried out by various methods. On the one hand, it can be compared with the direct measurements of RAOB or the inversion results of ground-based microwave radiometer; on the other hand, it can also be done by intercomparing with the inversion product of other satellites. The accuracy verification of the CLW is mainly compared with the retrieval results of the inversion of ground-based microwave radiometer or the retrieval product of other satellites. Here, we mainly verify the inversion effect of TPW and CLW based on FY-3D CMWS are very similar to those of ATMS inversion both in intensity and distribution.

Similarly, we quantitatively evaluated the difference between the TPW and CLW inversions of FY-3D CMWS and the results of ATMS inversion. The results of quantitative evaluation show that the correlation coefficients of TPW and CLW between CMWS and ATMS are 0.95 and 0.85, respectively, and the MAEs are 5.14 and 0.1 mm (see Figs. 8c, d). Moreover, the correlation coefficients and MAEs during the five independent days are also very close, which shows that the method proposed in this paper has good stability and robustness (see Table 2).

5. Conclusions and discussion

FY-3D satellite is the latest polar-orbiting meteorological satellite launched by China, which carries 10 sets of advanced monitor sensors. The MWTS and MWHS mounted on the FY-3D provide a total of 28 channel observations that greatly improve the inversion accuracy of temperature and humidity profiles. In particular, the 50–60-GHz oxygen channels in the MWTS and the oxygen channels near the 118.75-GHz channel in the MWHS can be used not only for mutual backup but also for re-
trieving of clouds and precipitation using dual oxygen absorption channels. However, due to the lack of two important low-frequency window channels, at frequencies 23.8 and 31.4 GHz, it is hard to retrieve TPW and CLW using FY-3D microwave sounding measurements. The similarity of channel settings between NPP ATMS and FY-3D microwave sounding instruments provides a good opportunity for us to simulate these two missing channels. First, using the correlation between measurements of all channels on the same sensor, the model relationship between other channels and two low-frequency window channels in ATMS is established by ML algorithm. Then, cross-calibration between MWTS and MWHS measurements and corresponding ATMS channels is carried out. Finally, the calibrated FY-3D microwave sounding measurements are input into the training model to obtain the simulated results. Quantitative evaluation results show that the correlation coefficients of Ch1 and Ch2 are 0.92 and 0.85, respectively. The MAEs of the two channels between ATMS and CMWS can also reach 6.74 and 5.73 K, respectively. Although the simulation errors of the two missing channels still seem to be a little big, the accuracy and stability of the two-channel simulation are still satisfactory, especially the simulation accuracy is sufficient to meet the quality control requirements of satellite data assimilation and atmospheric profile parameter inversion. In the next step, we will build training models for ocean and land, respectively, so as to further improve the accuracy of channel simulation.

After simulating the 23.8- and 31.4-GHz channels based on FY-3D microwave sounding measurements and ML model, the statistical inversion method was adopted to reverse TPW and CLW over ocean. The adopted statistical inversion method has been successfully applied to multiple satellite data [AMSU (Advanced Microwave Sounding Unit) and ATMS], and the inversion results have been widely used as satellite data quality control in numerical weather prediction systems. By comparing with different satellite inversion products, it can be found that the inversion results of TPW and CLW based on FY-3D microwave sounding instruments are in good agreement with those of ATMS in both strength and distribu-

\[ \text{Fig. 9. Comparisons of retrieved (a, b) TPW and (c, d) CLW between (a, c) ATMS and (b, d) CMWS for the descending orbit measurement on 8 July 2018.} \]
The correlation coefficients of TPW and CLW between CMWS and ATMS can reach 0.95 and 0.85, respectively. Actually, the channel settings of CMWS dataset established by ML are basically consistent with ATMS, which makes that the existing inversion algorithms based on ATMS can be seamlessly transplanted into CMWS. In addition, after successfully simulating 23.8- and 31.4-GHz channels, we can even build a new CMWS dataset contained 30 channels (plus 13 channels of MWTS and 15 channels of MWHS), which will provide more details for the inversion of temperature and humidity profiles in the vertical direction. Thus, the new CMWS dataset has great application prospects by combining the fast radiative transfer model independently developed by China (Yang et al., 2020), especially in the inversion of typhoon warm core structure, typhoon location and intensity determination, and precipitation estimation. At the same time, the next FY-3 satellite FY-3E will soon be launched, in which two window channels (at 23.8 and 31.4 GHz) will be included in the new microwave sounder instrument. The simulation and retrieval method proposed in this paper can also provide a good proxy test for FY-3E.

REFERENCES

Aires, F., C. Prigent, W. B. Rossow, et al., 2001: A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations. *J. Geophys. Res. Atmos.*, **106**, 14887–14907, doi: 10.1029/2001JD900085.

Backus, G., and F. Gilbert, 1968: The resolving power of gross earth data. *Geophys. J. Int.*, **16**, 169–205, doi: 10.1111/j.1365-246X.1968.tb00216.x.

Bennartz, R., A. Thoss, A. Dybbroe, et al., 2002: Precipitation analysis using the Advanced Microwave Sounding Unit in support of nowcasting applications. *Meteor. Appl.*, **9**, 177–189, doi: 10.1017/S1350482702002037.

Boukabara, S.-A., K. Garrett, and W. C. Chen, 2010: Global coverage of total precipitable water using a microwave variational algorithm. *IEEE Trans. Geosci. Remote Sens.*, **48**, 3608–3621, doi: 10.1109/TGRS.2010.2048035.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, doi: 10.1023/A:1010933404324.

Cadeddu, M. P., D. D. Turner, and J. C. Liljegren, 2009: A neural
network for real-time retrievals of PWV and LWP from arctic millimeter-wave ground-based observations. *IEEE Trans. Geosci. Remote Sens.*, 47, 1887–1900, doi: 10.1109/TGRS.2009.2013205.

Cimini, D., E. R. Westwater, and A. J. Gasiewski, 2010: Temperature and humidity profiling in the Arctic using ground-based millimeter-wave radiometry and 1DVAR. *IEEE Trans. Geosci. Remote Sens.*, 48, 1381–1388, doi: 10.1109/TGRS.2009.2030500.

Dyras, I., and D. Serafin-Rek, 2002: The use of AMSU data from NOAA-15 satellite for meteorological products generation. *Adv. Space Res.*, 30, 2461–2466, doi: 10.1016/S0273-1177(02)80306-4.

Ferraro, R. R., F. Z. Weng, N. C. Grody, et al., 1996: An eight-year (1987–1994) time series of rainfall, clouds, water vapor, snow cover, and sea ice derived from SSM/I measurements. *Bull. Amer. Meteor. Soc.*, 77, 891–905, doi: 10.1175/1520-0477(1996)077<0891:ATROPJ>2.0.CO;2.

Greenwald, T. J., G. L. Stephens, T. H. V. Haar, et al., 2017: A physical retrieval of cloud liquid water over the global oceans using Special Sensor Microwave/Imager (SSMI) observations. *J. Geophys. Res. Atmos.*, 122, 18471–18488, doi: 10.1029/2017JD027396.

Grody, N. C., A. Gruber, and W. C. Shen, 1980: Atmospheric water content over the tropical pacific derived from the Nimbus-6 scanning microwave spectrometer. *J. Appl. Meteor.*, 19, 986–996, doi: 10.1175/1520-0450(1980)019<0986:AWCT>2.0.CO;2.

Han, Y., X. L. Zou, and F. Z. Weng, 2015: Cloud and precipitation features of Super Typhoon Neoguri revealed from dual microwave sounder for numerical weather prediction and tropical cyclone applications. *J. Geophys. Res. Atmos.*, 120, 9163–9186, doi: 10.1002/2014JD022832.

Yang, J., Q. L. Min, W. T. Lu, et al., 2017: An RGB channel operation for removal of the difference of atmospheric scattering and its application on total sky cloud detection. *Atmos. Meas. Tech.*, 10, 1191–1201, doi: 10.5194/amt-10-1191-2017.

Yang, J., W. T. Lyu, Y. Ma, et al., 2018: Analyzing of cloud macroscopic characteristics in the Tibetan Plateau using the total-sky images. *J. Appl. Meteor. Climatol.*, 57, 1977–1987, doi: 10.1175/JAMC-D-18-0095.1.

Yang, J., S. G. Ding, P. M. Dong, et al., 2020: Advanced radiative transfer modeling system developed for satellite data assimilation and remote sensing applications. *J. Quant. Spectrosc. Radiat. Transfer*, 251, 107043, doi: 10.1016/j.jqsrt.2020.107043.

Yang, J., W. T. Lu, Y. Ma, et al., 2018: Radiative transfer modeling system developed for satellite data assimilation and remote sensing applications. *J. Climate*, 31, 4126393, doi: 10.1175/JCLI-D-18-0590.1.

Zhou, F.-C., X. N. Song, P. Leng, et al., 2020: An algorithm for retrieving precipitable water vapor over land based on passive microwave satellite data. *Adv. Meteor.*, 4126393, doi: 10.1155/2016/4126393.