

# Rainfall Algorithms Using Oceanic Satellite Observations from MWHS-2

Ruiyao CHEN<sup>\*1</sup> and Ralf BENNARTZ<sup>1,2</sup>

<sup>1</sup>*Earth & Environmental Sciences Department, Vanderbilt University, Nashville, TN 37235, USA*

<sup>2</sup>*Space Science & Engineering Center, University of Wisconsin-Madison, Madison, WI 53706, USA*

(Received 12 August 2020; revised 10 November 2020; accepted 13 November 2020)

## ABSTRACT

This paper describes three algorithms for retrieving precipitation over oceans from brightness temperatures (TBs) of the Micro-Wave Humidity Sounder-2 (MHWS-2) onboard Fengyun-3C (FY-3C). For algorithm development, scattering-induced TB depressions ( $\Delta$ TBs) of MWHS-2 at channels between 89 and 190 GHz were collocated to rain rates derived from measurements of the Global Precipitation Measurement's Dual-frequency Precipitation Radar (DPR) for the year 2017.  $\Delta$ TBs were calculated by subtracting simulated cloud-free TBs from bias-corrected observed TBs for each channel. These  $\Delta$ TBs were then related to rain rates from DPR using (1) multilinear regression (MLR); the other two algorithms, (2) range searches (RS) and (3) nearest neighbor searches (NNS), are based on k-dimensional trees. While all three algorithms produce instantaneous rain rates, the RS algorithm also provides the probability of precipitation and can be understood in a Bayesian framework. Different combinations of MWHS-2 channels were evaluated using MLR and results suggest that adding 118 GHz improves retrieval performance. The optimal combination of channels excludes high-peaking channels but includes 118 GHz channels peaking in the mid and high troposphere. MWHS-2 observations from another year were used for validation purposes. The annual mean  $2.5^\circ \times 2.5^\circ$  gridded rain rates from the three algorithms are consistent with those from the Global Precipitation Climatology Project (GPCP) and DPR. Their correlation coefficients with GPCP are 0.96 and their biases are less than 5%. The correlation coefficients with DPR are slightly lower and the maximum bias is ~8%, partly due to the lower sampling density of DPR compared to that of MWHS-2.

**Key words:** rainfall retrievals, 118 GHz, FY-3C, MWHS-2, multilinear regression, k-d tree

**Citation:** Chen, R., and R. Bennartz, 2021: Rainfall algorithms using oceanic satellite observations from MWHS-2. *Adv. Atmos. Sci.*, **38**(8), 1367–1378, <https://doi.org/10.1007/s00376-020-0258-5>.

## Article Highlights:

- Including 118 GHz channels in addition to other channels (89 and 183 GHz) improves the rain rate retrieval performance.
- Rainfall retrieval from channels between 89 and 190 GHz relies on the indirect scattering signature of ice particles.
- All three algorithms were developed based on scan positions and the resulting rainfall estimates are stable across the same scanline.
- The rainfall retrievals are highly consistent with other such products, particularly the comparison with GPCP, which achieves a correlation coefficient of 0.96.

## 1. Introduction

Advancing global precipitation observations are important for both scientific and operational purposes. Spaceborne sounding systems provide uniform global coverage relative to ground-based precipitation monitoring systems. The widely used sounding systems include the DMSP Special Sensor Microwave Imager/Sounder (SSMIS) series, National Oceanic and Atmospheric Administration (NOAA)

and MetOp Advanced Microwave Sounding Unit (AMSU) and Microwave Humidity Sounder (MHS) series, and the Suomi National PolarOrbiting Partnership (S-NPP) and NOAA-20 Advanced Technology Microwave Sounder (ATMS) series. Sounding observations made simultaneously in transparent and opaque water-vapor and oxygen absorption bands have been utilized extensively to exploit the capability of satellite-borne passive microwave sensors for precipitation detection (Staelin and Chen, 2000; Grody et al., 2001; Chen and Staelin, 2003; Weng et al., 2003; Ferraro et al., 2005; Vila et al., 2007; Surussavadee and Staelin, 2010; Laviola and Levizzani, 2011; Boukabara et al., 2013;

\* Corresponding author: Ruiyao CHEN  
Email: [ruiyao.chen@vanderbilt.edu](mailto:ruiyao.chen@vanderbilt.edu)

Sanò et al., 2015). The use of opaque bands also reduces errors caused by surface emissivity uncertainties.

Comparing these various precipitation retrieval algorithms, in general, there are three common major features that a conventional precipitation retrieval algorithm should follow. First, the algorithm retrieves precipitation intensity using the contrast of brightness temperature (TB) from hydrometeors and the radiatively cool ocean. Secondly, the algorithm is associated with statistical approaches. For example, a study by Ferraro and Marks (1995) proposed both linear and nonlinear regression methods to retrieve instantaneous rain rates from SSM/I observations. Sanò et al. (2015) proposed an algorithm based on neural networks for precipitation rate estimation from AMSU/MHS observations. Lastly, the algorithm requires historical precipitation datasets from measurements of spaceborne radars, ground radar networks or cloud-resolving model output. Widely used spaceborne radars include the Tropical Rainfall Measuring Mission's (TRMM) Precipitation Radar (PR) (Surussavadee and Staelin, 2010; Kummerow et al., 2015), the Global Precipitation Measurement's (GPM) Dual-Frequency Radar (DPR) and CloudSat profiling radar (Kidd et al., 2016).

The two principal physical mechanisms that permit the measurement of rain with microwave radiometers are emission and scattering. Precipitation retrievals derived from channels under 89 GHz are emission-based, where liquid precipitation causes TB to increase over a radiometrically cold background. Because channels above 89 GHz can hardly "see" through the atmosphere to the surface when precipitation occurs, the precipitation algorithms are scattering-based, where precipitation, especially that above freezing level, causes brightness temperature to decrease over a radiometrically warm or cold background (Wilheit, 1986; Spencer et al., 1989). While other temperature and humidity absorption bands between 89 and 190 GHz have been well-investigated based on satellite observations, the 118 GHz oxygen absorption line has not been in full use because none of these spaceborne instruments employs the 118 GHz channels. To augment existing retrieval algorithms, this study proposes to use TB depression caused by scattering associated with precipitation-sized ice particles to retrieve indirectly related surface rainfall information. The channels used in this study are between 89 and 190 GHz, of which some are near the 118 GHz oxygen absorption line.

The China Meteorological Administration's (CMA) Fengyun-3C (FY-3C) MicroWave Humidity Sounder-2 (MWHS-2) is characterized by 15 microwave channels ranging in frequency from 89 to 190 GHz. In addition to carrying more traditional channels around 89, 150 and 183 GHz, the MWHS-2 carries eight temperature sounding channels near 118 GHz (Dong et al., 2009; Zhang et al., 2012). It is the first space-borne instrument carrying the 118 GHz channels, and expected to provide new information for not only temperature sounding but also for precipitation retrievals (Bauer and Mugnai, 2003; He and Chen, 2019).

This study builds upon the work from Chen and Bennartz (2020) that investigated the sensitivities of TBs at channels between 89 and 190 GHz to ice scattering caused by precipitation-sized ice particles using MWHS-2 observations. Based on the findings from this work, we will continue using the MWHS-2 observations and focus on the channels that were discovered to show sensitivities to ice scattering in varying degrees for rainfall retrieval algorithm development. This study will provide new perspectives for other instruments that have frequency coverage at 118 and 183 GHz, such as NASA's Time-Resolved Observations of Precipitation structure and storm Intensity with a Constellation of Smallsats (TROPICS) and EUMETSAT's Microwave Imager (MWI) (Holmlund et al., 2017; Blackwell et al., 2018; Mattioli et al., 2019).

The remainder of this paper is structured as follows. The instruments and dataset used, as well as the method of preprocessing the data, are described in section 2. In section 3 we present the bias correction process for the MWHS-2 observations, and the responses of TB and scattering-induced TB depression to rainfall. Next, we describe three precipitation retrieval algorithms and evaluate the rain rates produced by each algorithm by comparing them against other well-established rainfall products in section 4. Finally, the conclusions are presented in section 5.

## 2. Instruments, data and methods

The description of the datasets and methods of preprocessing the data parallels that of Chen and Bennartz (2020). The MWHS-2 instrument provides the observed TBs at 15 channels between 89 and 190 GHz, and the channels and their frequencies as well as polarizations are given in Table 1. The MWHS-2 observation information was extracted from the MWHS-2 Level-1 files provided by the CMA/National Satellite Meteorological Center (CMA/NSMC) (<http://satellite.nsmc.org.cn/portalsite/default.aspx>).

The GPM core spacecraft hosts two instruments: the DPR and the GPM Microwave Imager (GMI) (Hou et al., 2014). The GPM operates in a circular orbit at an altitude of 407 km and inclination of 65°. This orbit was chosen because it can ensure sufficient overlap with sun-synchronous satellites, such as FY-3C, for cross-calibration and covering a large portion of Earth's surface with minimal repetition of ground track. It also allows for the gathering of samples at latitudes where most precipitation occurs in terms of absolute amount at various times of the day. The DPR instrument combines a Ku- and Ka-band precipitation radar capable of making accurate rainfall measurements from the ground to 19 km in altitude. The surface rain rates retrieved from the DPR were collocated with the MWHS-2 observations to produce matchups for investigating the implicit relationships in between at various channels including the newly added 118 GHz channels. The DPR rainfall retrieval was obtained from the GPM 2BCMB product provided by the NASA Precipitation Processing System

**Table 1.** MWHS-2 channel frequencies and polarization at nadir used in RTTOV.

Channel number	Frequency (GHz)	Polarization at nadir used in RTTOV
1	89	H
2	$118.75 \pm 0.08$	V
3	$118.75 \pm 0.2$	V
4	$118.75 \pm 0.3$	V
5	$118.75 \pm 0.8$	V
6	$118.75 \pm 1.1$	V
7	$118.75 \pm 2.5$	V
8	$118.75 \pm 3.0$	V
9	$118.75 \pm 5.0$	V
10	150	H
11	$183.31 \pm 1.0$	V
12	$183.31 \pm 1.8$	V
13	$183.31 \pm 3.0$	V
14	$183.31 \pm 4.5$	V
15	$183.31 \pm 7.0$	V

archived at the NASA GES DISC (<https://doi.org/10.5067/GPM/DPRGMI/CMB/2B/06>).

MWHS-2 observations and DPR profiles were projected into each  $0.25^\circ$  latitude  $\times$   $0.25^\circ$  longitude grid with their maximum time difference being 15 minutes. This process was applied to oceans only due to the more complicated simulation of the surface emissivity over land. To eliminate the impact of the high zenith angles and the low spatial resolution of MWHS-2 at the outer edge of each scan, the ten outermost scan positions (five on each side) were excluded from the collocated dataset. Eventually, a total of over 1.5 million samples were achieved for the year of 2017.

To validate the developed rainfall retrieval algorithms, two different precipitation products were used as benchmarks. The first consists of rain rates in the aforementioned GPM 2BCMB product. The other source of rain rate data is the Global Precipitation Climatology Project (GPCP) formed by the World Climate Research Program in 1986 (WCRP 1986) to exploit the capabilities of satellite-borne instruments along with gauges for producing monthly and finer temporal resolution global precipitation in the long term (Adler et al., 2003). It has three products on different scales:  $2.5^\circ \times 2.5^\circ$ ,  $1^\circ \times 1^\circ$ , and pentad (5 days). In this study, the GPCP monthly global precipitation data on  $2.5^\circ \times 2.5^\circ$  scales for the year 2016 were used. The GPCP data were provided by NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their website at <https://psl.noaa.gov/>.

### 3. Bias correction and brightness temperature analysis

#### 3.1. Bias correction and radiative transfer simulation

It is important to ensure that the MWHS-2 TBs are bias-free before we use them for further analysis. Chen and

Bennartz (2020) proposed a bias-correction method based on the idea that the mode of the histogram of the TB differences corresponds to the observations affected by precipitation at a minimum level, and therefore this mode can be regarded as an estimate of the bias. We applied this method by first calculating the differences between observed and simulated TBs for each MWHS-2 channel. We then calculated the mode of the histograms of TB differences per channel and per scan position. The resulting bias-correction values were subtracted from observations to produce bias-free observed TBs.

The TIROS Operational Vertical Sounder Radiative Transfer (RTTOV, Version 12.2) Model (Saunders et al., 2007; Saunders et al., 2018; Hocking et al., 2019) was used to simulate clear-sky background TBs for all the 15 channels of MWHS-2. The ERA-Interim data from the European Centre for Medium-Range Weather Forecasts (ECMWF) provided the 6-hourly surface and vertically resolved moisture and temperature field products (Dee et al., 2011). This dataset was obtained from the National Center for Atmospheric Research (downloaded from <https://rda.ucar.edu/datasets/ds627.0/>). The MWHS-2 observations were collocated with the ERA-Interim reanalysis profiles and their maximum time difference is 3 h. The resulting ERA-Interim profiles, together with sea surface temperature and surface wind speed, were then used as inputs to the RTTOV cloud-free radiative transfer simulations.

#### 3.2. Brightness temperature response to rainfall

We define the scattering-induced brightness temperature depression ( $\Delta TB$ ) as the difference between bias-corrected microwave observations,  $TB_{obs}$ , and simulated clear-sky background brightness temperatures,  $TB_{sim}$ :

$$\Delta TB = TB_{obs} - TB_{sim}. \quad (1)$$

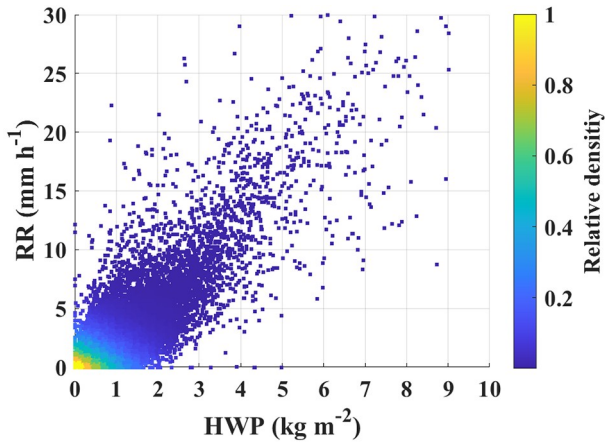
Chen and Bennartz (2020) investigated the relation between  $\Delta TB$  of the individual MWHS-2 channels to the presence of hydrometeors and concluded that the oxygen and water vapor sounding channels exhibit a strong dependency on how close each channel is to the center of its corresponding absorption line. It was also found that the actual scattering intensity of ice particles monotonically increases with frequency. Based on these findings, we first examine the relation between the hydrometeor water path and the surface rain rate. Figure 1 shows a strong linear relationship between the two, which indicates that the surface rain rate is highly associated with the quantity of hydrometeors in a vertical column. This reinforces the implicit yet virtual relationship between the surface rain rates and scattering-induced  $\Delta TB$ , which forms the basis of the subsequent rainfall retrieval algorithm development.

Next, we explore the pattern of rain rates derived from the DPR in terms of variations in  $\Delta TB$  and  $TB_{obs}$ . Figure 2 presents the two-dimensional rainfall distribution relative to the  $TB_{obs}$  and  $\Delta TB$  for all 15 MWHS-2 channels. The highest-peaking channels 2–4 only exhibit slight deviations

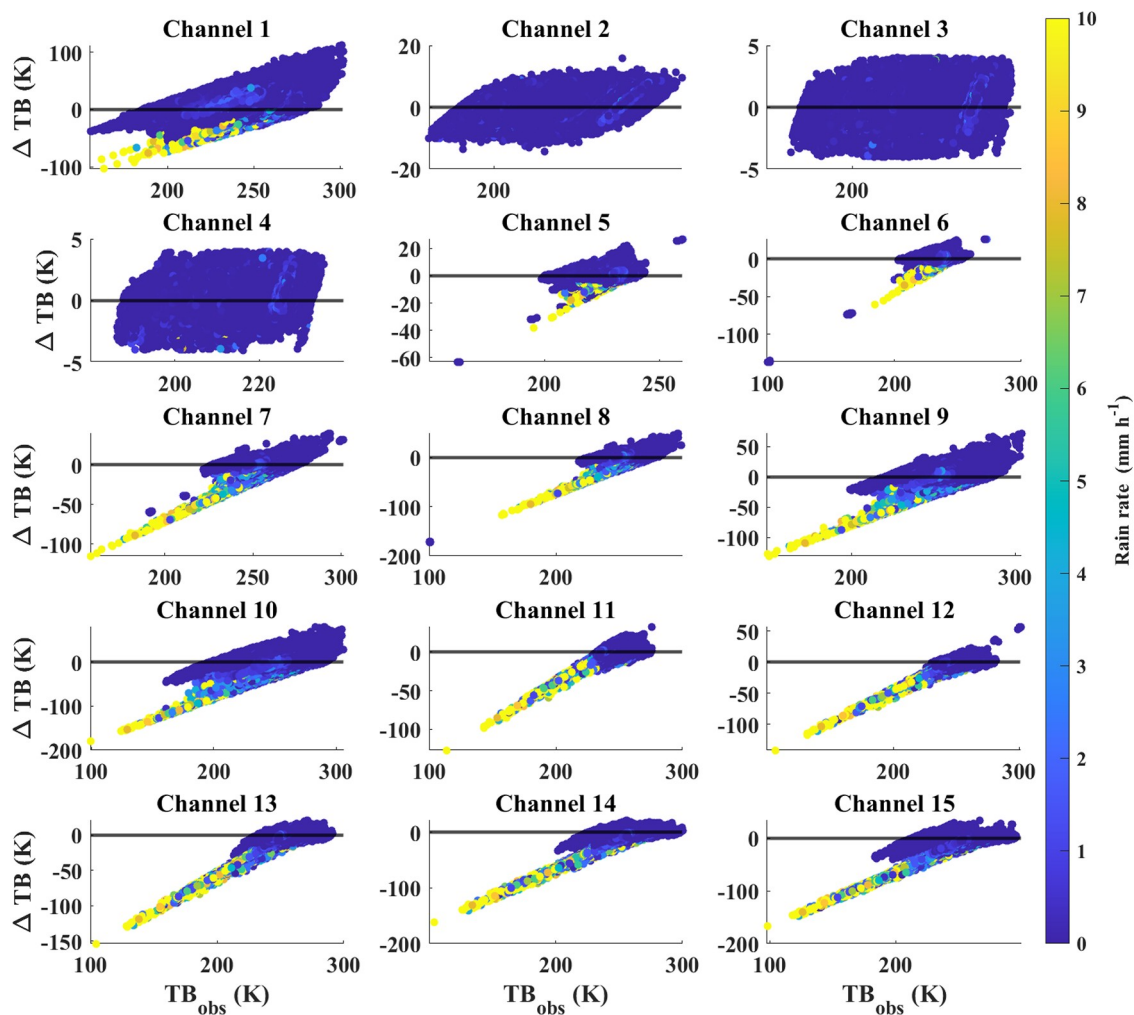
of  $\Delta TB$  from zero. Because of their insensitivity to ice particle scattering, channels 2–4 will be excluded in subsequent analysis. Channel 5 presents the weakest sensitivity among the rest channels (channels 1, 5–15) and therefore this channel is non-essential for deriving the rainfall

retrieval algorithms. In this study, we will include channel 5 in only one of the three algorithms that will be described in section 4.

For all the 12 channels, heavier rainfall occurs at colder  $TB_{obs}$  and larger negative  $\Delta TB$ s, while warm  $TB_{obs}$  and near-zero  $\Delta TB$  are mostly accompanied by near-zero rain rates. The latter highlights that a perfect radiative transfer model with a perfect clear-sky input would produce near-zero values in  $\Delta TB$  for all cloud-free conditions regardless of how warm the  $TB_{obs}$  is. Also noticeable is that, in several channels, including channels 7–10 and 14–15, the rainfall distribution shows a bifurcation between those data following the horizontal zero line and those for which  $\Delta TB$  decreases approximately linearly with decreasing  $TB_{obs}$ . Among the two groups of data, given the same  $TB_{obs}$  the latter occurs with larger negative  $\Delta TB$  and heavier rainfall, and the former is mostly with much smaller negative (or near zero)  $\Delta TB$  and very light (or near-zero) rainfall. In other words, scattering reduces the amount of radiation and results in large negative  $\Delta TB$ , which provides more substantial information than the  $TB_{obs}$  regarding the measurement



**Fig. 1.** Linear relationship between hydrometeor water path (HWP) and surface rain rate (RR).



**Fig. 2.** Two-dimensional rain rate distributions of  $TB_{obs}$  and  $\Delta TB$  for all 15 channels of MWHs-2 and for all collocated data. Note the different scales of both the x- and y-axis.



of rainfall.

The large positive  $\Delta TB$  in each of channels 1 and 7–10 represents an emission signal caused by liquid clouds and rain to different extents, with the largest of over 100 K in channel 1. For cases of little or no ice as scatterers in the atmosphere, this will allow us to still be able to retrieve rainfall using these channels based on their emission signals.

## 4. Precipitation retrieval algorithms

### 4.1. Multilinear regression

#### 4.1.1. Algorithm description

Based on the above analysis, we develop four different multilinear regression (MLR) models for each of the 44 scan positions using different sets of channels of MWHS-2.  $\Delta TB$ s of different combinations of channels and rain rates are considered as independent and response variables, respectively, to model their relationships. Because dry snow can scatter significantly like the precipitation ice particles, the signal can be misinterpreted as rainfall. To avoid this issue, the retrieval methods in this study are limited to the tropics and midlatitudes between 35°N and 35°S. The channel sets of the four models are: (1) channels 1, 6–15; (2) channels 1, 10–15; (3) channels 1 and 10; (4) channel 1 only. Therefore, for each set of channels, we have built 44 different MLR sub-models for 88 symmetrical scan beams of MWHS-2 (five scan beams on each side are removed for the purpose of quality control). The regression performances in terms of the correlation coefficient ( $R$ ), mean absolute error (MAE) and root-mean-square error (RMSE) for each of the models are presented in Table 2. Model 1 performs better than the other models in terms of  $R$  and RMSE. The MAEs of Models 1, 3 and 4 are the same and slightly lower than that of Model 2. The better performance of Model 1 indicates that the addition of the lower peaking channels near 118 GHz, channels 6–9, is necessary to improve rainfall retrieval.

#### 4.1.2. Algorithm evaluation

We further apply the regression coefficients derived from Model 1 in the above analysis to another full year (2016) of MWHS-2 observations over oceans between 35°N and 35°S. The resulting rain rates are compared with the DPR-derived rain rates as well as with the GPCP gridded

rain rates. All comparisons are performed on an annual-averaged  $2.5^\circ \times 2.5^\circ$  grid that is also used by the GPCP. We note that comparisons between DPR- and MWHS-2-derived annual means are not entirely independent as DPR values are also chosen for training the MWHS-2 regression retrievals, although a different year was used for the collocated dataset that underlies the training.

Figure 3 shows scatterplots of annual mean surface rain rates for all four MWHS-2 retrievals against the DPR and GPCP, respectively. The following conclusions can be drawn from these scatterplots:

(1) The scatterplots provided in Fig. 3 show generally strong correlations ( $R > 0.82$  in all cases) between the DPR-derived annual mean rain rates and all four retrieval versions of MWHS-2. A degradation can be observed, however, both in terms of RMSE and in terms of the linear relation between the two quantities (see red lines in Fig. 3). Going from V01 to V04, the regression uses fewer channels (Table 2), and the relation between MWHS-2-derived retrievals and DPR-derived retrievals deviates more strongly from the 1:1 line. It thus appears that all four bands (89, 118, 150, and 183 GHz) provide independent information that contributes to improved rain rate retrievals.

(2) When comparing MWHS-2 to GPCP, one can see that the scatter between the two different datasets is significantly smaller than the scatter between DPR and MWHS-2 partly because the data density for DPR is lower (only 25 independent beams per scan, as opposed to 88 for MWHS-2). This increased noise in DPR gridded estimates will also be observed in the following analysis.

(3) The correlation between the GPCP and MWHS-2 exceeds 0.93 for all four versions of the retrievals. Similar to the comparison between DPR and MWHS-2, the inclusion of more bands (V01) illustrates a greater sensitivity than the MWHS-2 retrievals with fewer bands (e.g., V04).

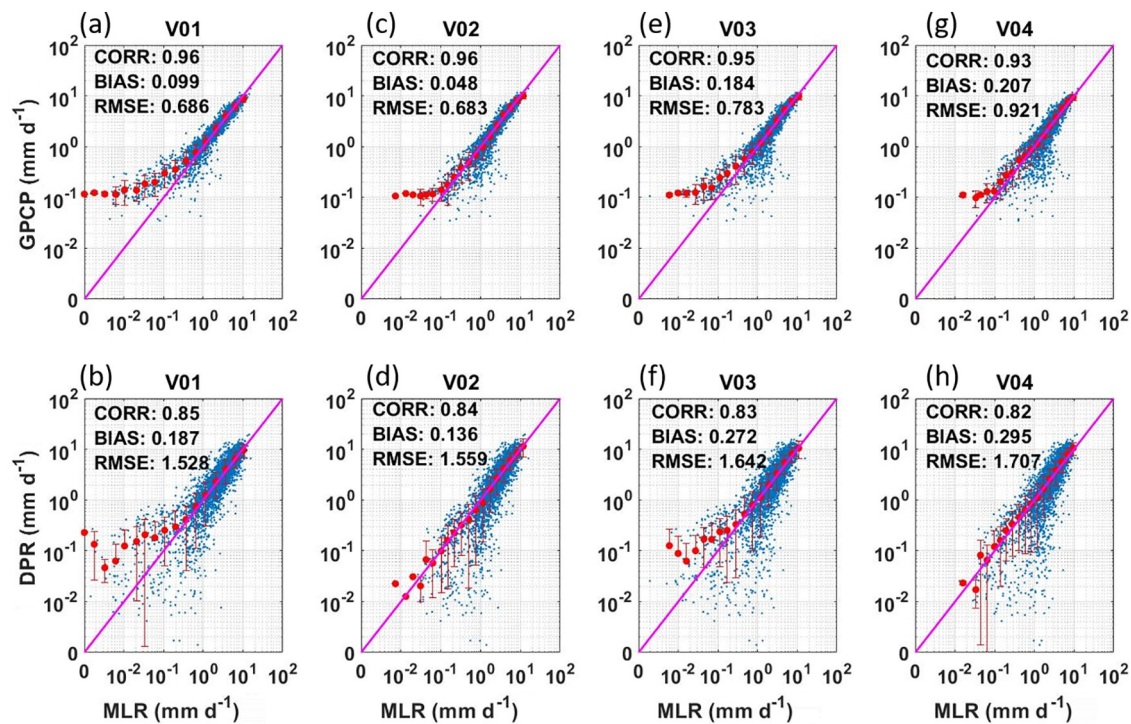
In particular, the MWHS-2 V01 retrievals compare well against the GPCP, with the regression curve (red line) falling nearly on the 1:1 line and a correlation of 0.96, whereas a slight underestimation occurs at light rainfall ( $< 0.3 \text{ mm d}^{-1}$ ).

A comparison of the spatial distribution of the annual mean surface rain rates between MWHS2, GPCP and DPR (Fig. 4) yields the following key points:

(1) The spatial distribution of both MWHS-2 and GPCP reflects well the major areas of deep convection in

**Table 2.** Performance metrics summary of RR regression models. Reported are the correlation coefficient ( $R$ ), the mean absolute error (MAE) and the root-mean-square error (RMSE) for all four retrieval models. All regressions were performed on the precipitation-induced brightness temperature depressions  $\Delta TB$ s. Coefficients for each model were derived individually for each scan position.

Model	$R$	MAE ( $\text{mm h}^{-1}$ )	RMSE ( $\text{mm h}^{-1}$ )	MWHS-2 channels	Channel selection
1	0.64	0.23	0.69	1, 6–15	89 GHz and 150 GHz window channels and 118 GHz and 183 GHz sounding channels
2	0.61	0.22	0.71	1, 10–15	Excluding 118 GHz sounding channels
3	0.59	0.23	0.73	1 and 10	Only 89 GHz and 150 GHz
4	0.57	0.23	0.74	10	Only 150 GHz



**Fig. 3.** Scatterplots of annual mean  $2.5^\circ \times 2.5^\circ$  gridded rain from GPCP [y-axis of (a, c, e, g)] and DPR [y-axis of (b, d, f, h)] compared against the four different retrievals from MWHS-2 over oceans between  $35^\circ\text{N}$  and  $35^\circ\text{S}$ .

the tropics. Differences between DPR and MWHS-2 appear to show an overestimation by MWHS-2 near Indonesia and an underestimation in areas such as the central Pacific ITCZ. Interestingly, this behavior differs from that observed in the ice water path derived from MWHS-2 observations (for brevity, not shown here), indicating that the relation between the ice water path and surface rain rates itself differs in these two areas. This result could conceivably be caused by higher aerosol loading near Indonesia, which would, compared to cleaner air, lead to reduced surface rain rates for given hydrometeor water paths. Such a mechanism over Indonesia was first described by Rosenfeld (1999). The DPR-derived rain rates are noisier than the MWHS-2 derived rain rates, which supports the above explanation that the lower DPR data density is at least partly responsible for the increased scatter.

(2) Comparing GPCP and MWHS-2, the scatter is generally lower, with a similar overestimation over Indonesia and a few other coastal regions. The central Pacific ITCZ also shows slight underestimation that is evident in the comparisons with the DPR.

The comparison illustrated here suggests that the high-frequency microwave channels between 89 GHz and 183 GHz can successfully be used to derive rain rates. These channels can also be used for precipitation retrieval with the caveat that the retrieval relies on the indirect scattering signature of ice particles higher up in the atmosphere that are not directly linked to surface precipitation. Thus, if the relation between the ice water path and surface rain rates itself changes, the surface rain rate retrievals will be adversely

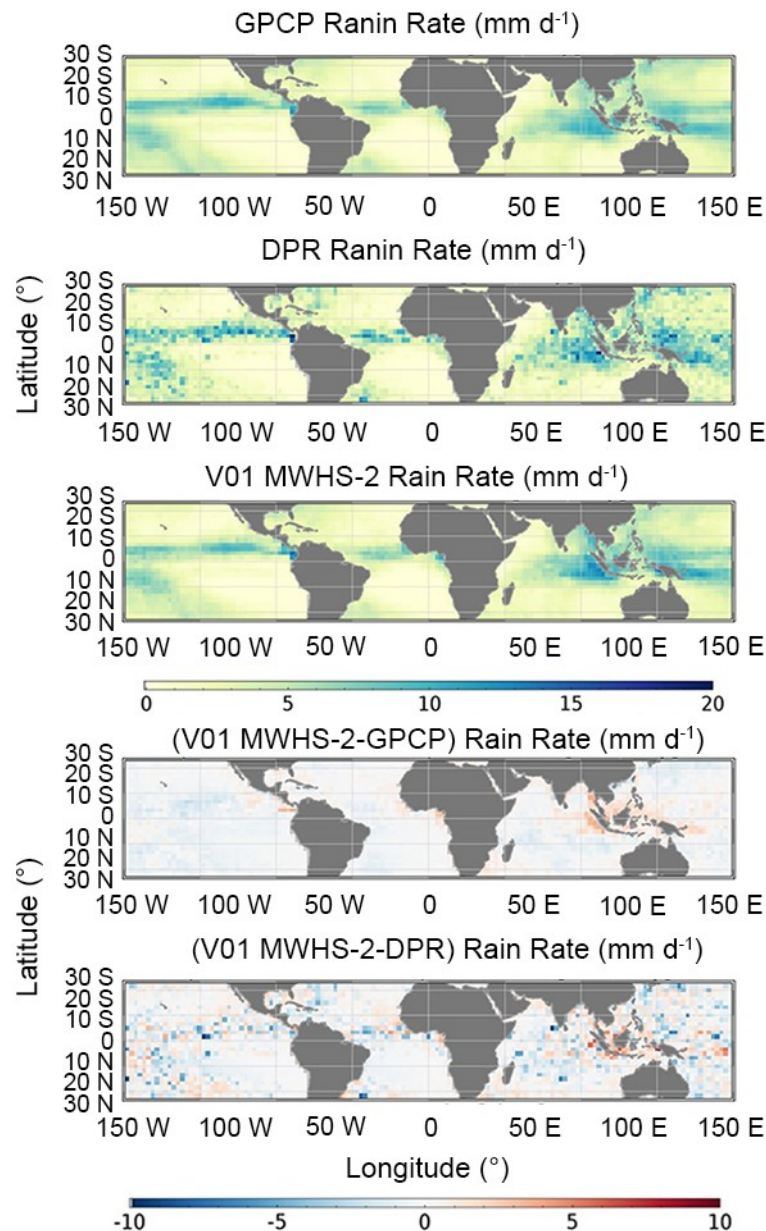
affected, as shown above in the case of Indonesia. In addition, the algorithm used to perform the regression based on scan positions has the advantage of eliminating the concern that different footprints of a cross-track radiometer have different local zenith angles.

Based on the above analysis, channels 1 and 6–15 are selected for our study going forward (channel 5 will be carried for some cases, but its impact is negligible because of its high peaking weighting function).

## 4.2. Range searches and nearest neighbor searches

### 4.2.1. Description of algorithms

A K-Dimensional tree (or k-d tree, where  $k$  is the dimensionality of the search space) is a hierarchical structure built by partitioning the data recursively along the dimension of maximum variance. At each iteration, the variance of each column is computed and the data is split into two parts on the column with maximum variance. It is a very useful structure, especially for searches involving a multi-dimensional search key, e.g., range searches and nearest neighbor searches (Bentley, 1980). As a simple example, assume that  $k = 2$  and one needs to build a 2D tree which is also regarded as a generalization of a binary search tree. The idea is to build a binary search tree with points in the nodes using the  $x$ - and  $y$ -coordinates of the points as keys in strictly alternating sequence. Starting with the  $x$ -coordinate at the root, if the point to be inserted has a smaller  $x$ -coordinate than the point at the root, it goes left; otherwise it goes right. At the next level, the insertion is switched to the other



**Fig. 4.** Comparison of monthly mean surface rain rates between MWHS-2 (V01 regression only), DPR and GPCP. The upper three plots show the annual mean surface rain rates. The lower two plots show MWHS-2 minus GPCP and DPR, respectively.

coordinate ( $y$ -coordinate). If the point to be inserted has a smaller  $y$ -coordinate than the point in the node, it goes left, otherwise it goes right. The coordinate is then switched again and so on and so forth, until the insertion of the last point.

For the purpose of rainfall retrieval, we use 12 channels (channels 1 and 5–15) of MWHS-2 observations and radiative transfer simulations as well as matched rain rates derived directly from the DPR to build a 12-dimensional tree. We first divide the  $\sim 1.5$  million collocated data for the full year of 2017 into two sub-datasets: 70% for training and 30% for testing. Each sub-dataset included the  $\Delta TB$ s of

channels 1 and 5–15 from MWHS-2 observations and radiative transfer simulations. To address the slant path impact on the MWHS-2 observations, we further stratify the training dataset into four subsets uniformly based on the relative air-mass [ $1 / \cos(\theta)$ ] that is calculated from the zenith angle ( $\theta$ ) of MWHS-2. Considering the training subset  $i$  ( $i = 1, 2, 3$  or 4) has  $n_i$  points, the k-d tree algorithm partitions this  $n_i$ -by-12 dataset by recursively splitting the  $n_i$  points in 12-dimensional space into a binary tree known as a model object, which is a convenient way of storing information of the grown tree. Four individual k-d trees (model objects) are then created and passed to the subsequent process of search-



ing neighbors.

Next, two different search mechanisms are adopted to estimate the rain rates from MWH-2 observations based on the four k-d trees built earlier:

(1) Range searches (RS): Given a range (hypersphere radius) of  $r$  Kelvin and a point in the query data (testing data), we search for all points in the model object that are within a Euclidean distance  $r$  Kelvin from that query point and consider them as neighbors. The indices of these neighboring points are then used to map the corresponding DPR rain rates in the training data. This will allow us to obtain a set of neighboring rain rate estimates for each query point. The average value over this neighboring set represents the estimated k-d rain rate, henceforth called the RS rain rate. Excluding the zero rain rates, the rest of the neighboring rain rates, which are precipitation cases, are averaged to represent the conditional rain rate. An advantage of this method is that the percentage of non-zero rain rates over this neighboring set provides an estimation of the probability of precipitation.

(2) Nearest neighbor searches (NNS): Given a point in the query data, we find the point in the k-d tree that is nearest to that query point in terms of the Euclidean distance. The index of the nearest neighbor then enables the mapping of the corresponding DPR rain rate in the training data. This mapping then yields the nearest neighboring rain rate, which serves as another way of representing the estimated k-d rain rate of that query point, henceforth called the KD NN rain rate. Compared to RS, NNS can be done efficiently by using the tree's properties to quickly eliminate large portions of the search space, especially in a study such as the present one that deals with high dimensional data (12 dimensions).

For both the RS and NNS method, we use the zenith

angles in the testing data to determine which k-d tree out of the four should be used for searching neighbors. For RS, we set the radius to be proportional to the MWH-2 noise equivalent temperature (NE $\Delta T$ ), as the following equation shows:

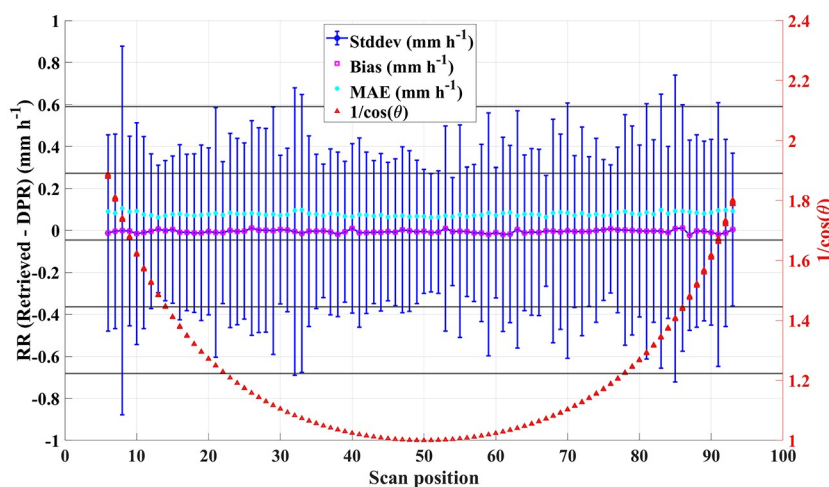
$$r = \text{NE}\Delta T \sqrt{k}. \quad (2)$$

The NE $\Delta T$  of MWH-2 is initially set to be 1 K and  $k$  is the dimension of the query dataset (here, 12). For points found to be without neighbors within the initial radius, we extend the search to a larger hypersphere by continuously increasing the NE $\Delta T$  with increments of 1 K until the maximum value of 5 K was reached. As such, more than 98% of the points are found with at least one neighbor. The NNS is concluded once the first neighbor is found with the search range within up to 5 K.

The statistics of the RS rain rates compared to the DPR rain rates for testing data per scan position are shown in Fig. 5. These statistics include the mean bias, standard deviation of the bias, and MAE. Despite different slant paths at various scan positions, the rain rate estimates are very stable across the same scanline. The largest deviation from the DPR rain rates is about 0.04 mm h<sup>-1</sup> and the largest MAE is less than 0.1 mm h<sup>-1</sup>.

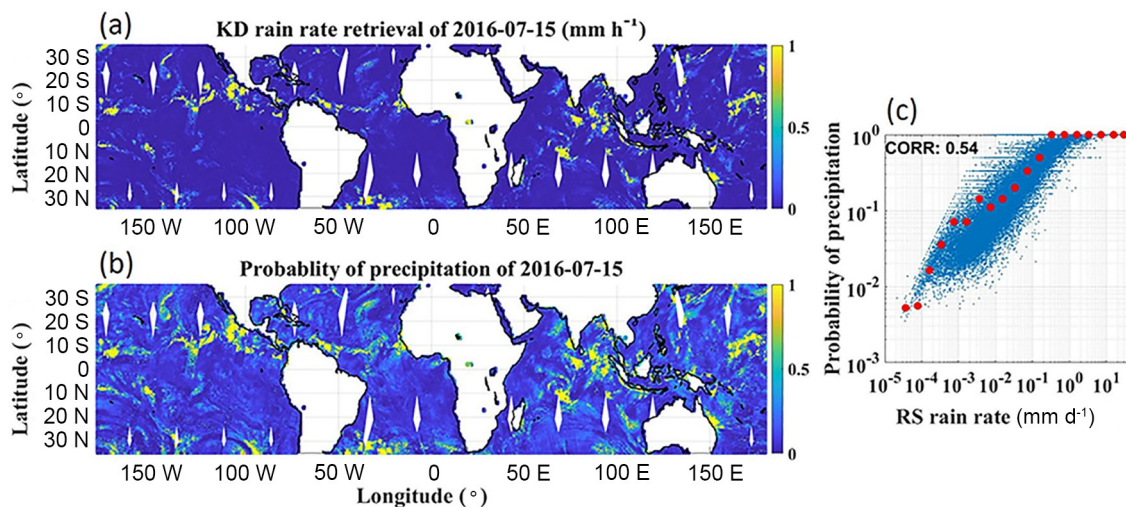
#### 4.2.2. Evaluation of algorithms

Like the procedure for validating the MLR method, we apply the created k-d trees to the MWH-2 observations for the year 2016 to retrieve the rain rates based on either RS or NNS. An example of the rain rate retrievals with the unit of mm h<sup>-1</sup> using the RS method for the day of 15 July 2016 is shown in Fig. 6a. The corresponding probability of precipitation is also shown, in Fig. 6b, in which the tropical regions with deep convection generally have a higher chance of pre-



**Fig. 5.** Statistics of retrieved rain rates compared to DPR rain rates, for testing data from MWH-2 by searching neighbors within a fixed hypersphere radius per scan position. The statistics include bias (magenta dots), mean absolute error (MAE, mint-green dots) and standard deviation of the bias (blue line). The brick-red dots are for relative airmass, which is used to stratify the training data when creating the four k-d trees and to determine which k-d tree is used in neighbor-searching for testing data.



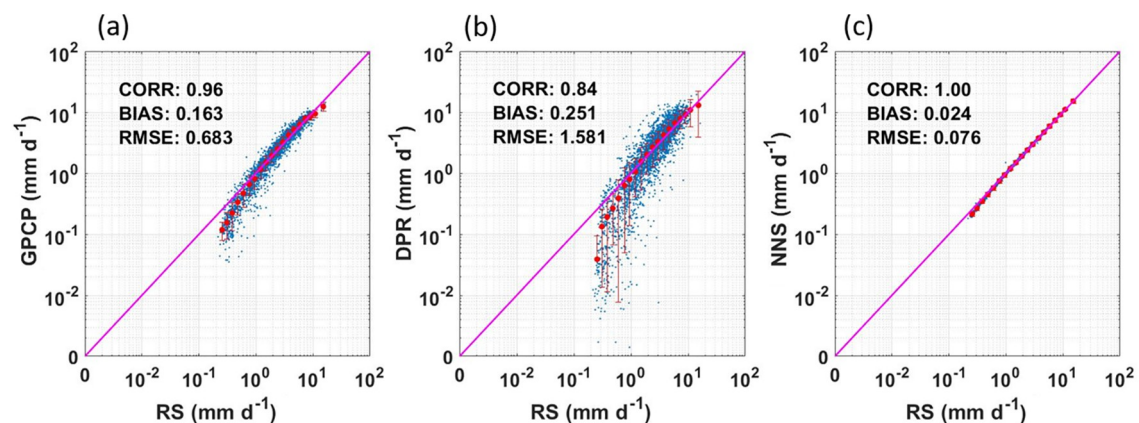


**Fig. 6.** Spatial distribution of (a) rain rate retrieval and (b) probability of precipitation, and (c) scatterplot of (a) versus (b) on double logarithmic scale (blut dota) with their bin average (x-axis, RS rain rate) or mode (y-axis, probability of precipitation, red dots), from MWHS-2 observations on 15 July 2016.

precipitation. RS rain rates and probabilities of precipitation, as well as their bin-averaged values, are projected on a double logarithmic scale in Fig. 6c. Note that the probability of precipitation cannot be averaged and therefore we take the mode of the probabilities within each bin to represent the probability of precipitation of a given bin. The probability of precipitation derived from RS is correlated well with the rain rate retrieval. This precipitation probability provides us with an uncertainty estimate associated with each measurement and it allows us to evaluate whether or not the observed scene is raining at all. Classical retrieval algorithms only provide rain rates.

After applying the created k-d trees to the MWHS-2 observation, we then gridded RS rain rates and NNS rain rates to compare them with those from DPR and GPCP. Hereafter, the analysis is based on the annual mean gridded rain rates with the unit of  $\text{mm d}^{-1}$ . Figures 7a and b show the scat-

terplots of annual RS rain rates against GPCP and DPR rain rates, respectively. The correlation coefficient between RS rain rates and GPCP rain rates is more than 0.96. It is worth noting that the scatter between the RS rain rates and GPCP rain rates is significantly smaller than the scatter between those from RS and DPR. This confirms the results observed in the MLR models. This again is mostly caused by the lower data density of DPR, which has only 25 independent beams per scan, as opposed to 88 for MWHS-2. The gridded rain rates are also stratified logarithmically based on RS rain rates. These averaged rain rates are also illustrated over the scatterplots in Fig. 7, in which the red dots and lines represent averages and standard deviations of rain rates of the y-axis. In other words, the average and standard deviation are of either GPCP rain rates or DPR rain rates. In both subplots, the yellow dots fall on the 1:1 lines with slight overestimation over the light precipitation range ( $< 0.4 \text{ mm d}^{-1}$ ),



**Fig. 7.** Scatterplots of annual mean  $2.5^\circ \times 2.5^\circ$  gridded rain rates from GPCP [y-axis, (a)] and DPR [y-axis, (b)] compared against rain rate retrievals by RS from MWHS-2 with units of  $\text{mm d}^{-1}$ , on logarithmic scale. (c) Comparison of rain rate retrievals between RS and NNS. Red dots and red lines are averages and standard deviations of either GPCP rain rates or DPR rain rates by subsetting RS rain rates logarithmically.

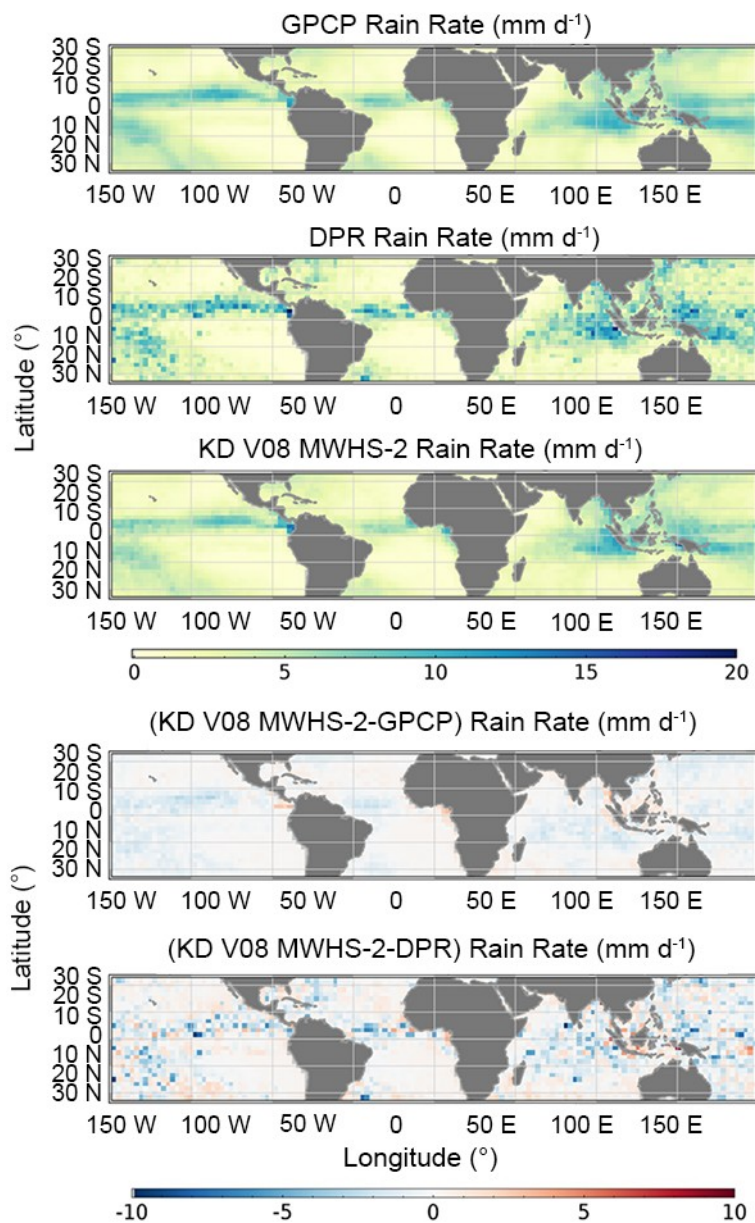
which means that the RS rain rate retrievals are in exceptional agreement with the rain rates from GPCP and DPR. The results of NNS are substantially equivalent to those of RS shown by Fig. 7c, where the correlation coefficient is 1, and both the bias and RMSE are extremely low. This demonstrates that these constructed k-d trees tend to be robust to noise and invariant to the spatial heterogeneity of rainfall. Moreover, this allows for the selection of NNS over RS in scenarios requiring lower computation costs and disregarding the probability of precipitation.

Because of the above analysis, we leave out the results of NNS and only show the spatial distributions of the rain rates from RS compared with those from GPCP and DPR in Fig. 8. Similar to that of MLR compared against GPCP and

DPR, the spatial distributions of RS rain rates also reflect the major areas of deep convection in the tropics well. However, the overestimation near Indonesia is less than that using the MLR method. Comparing the RS rain rates and DPR-derived rain rates, the latter are noisier than the former, which further confirms the previous inference that the lower DPR data density is at least partly responsible for the less congruent retrievals.

## 5. Conclusions

Three algorithms for rainfall retrieval estimation were developed in this paper by combining observed TBs from MWHS-2 and simulated TBs from the radiative transfer



**Fig. 8.** Spatial distribution of annual mean  $2.5^\circ \times 2.5^\circ$  gridded rain rates from MWHS-2 by RS compared against the rain rates from DPR and GPCP with units of  $\text{mm d}^{-1}$ .

model as well as matched rain rates derived directly from the GPM DPR. In order to avoid issues related to the simulation of surface emissivity over land, we limited this algorithm development to open oceans only. Using four tests with different combinations of channels, we found that adding 118 GHz to the other high frequencies (89, 150 and 183 GHz) improves retrieval performance. Results also suggest that channels peaking too high to be sensitive to ice scattering have contributed little to the retrieval process. Consequently, we selected channels 1, 5–15 (or channels 1, 6–15) as inputs to the three rain rate retrieval algorithms. The first algorithm is based on MLR that models the relationships between corresponding  $\Delta T_B$ s (independent variables, differences between observed and simulated MWHS-2  $T_B$ s) and rain rates (response variable). Both RS and NNS are based on the k-d tree data structure that is a 12-dimensional tree built using 70% of the matched  $\Delta T_B$ s of MWHS-2 at channels 1, 5–15. To eliminate the impact of slant path variations in the MWHS-2 observations, all three algorithms were performed for each individual scan position (MLR) or each individual group of scan positions (RS and NNS). The rain rate retrievals estimated from each algorithm are compared against the rain rates from GPCP and DPR, respectively, on an annual mean  $2.5^\circ$  latitude  $\times$   $2.5^\circ$  longitude gridded basis.

All three algorithms generally show good agreement with the GPCP rain rates, with correlation coefficients of 0.96 and maximum averaged bias less than 5%. While the averaged bias of the MLR rain rates is slightly lower than that of RS or NNS, a slight underestimation exists in MLR at light rainfall ( $< 0.3 \text{ mm d}^{-1}$ ). The results of RS and NNS based on k-d trees are in extremely high agreement in terms of the correlation coefficient, bias and RMSE, which makes them almost interchangeable. Yet, NNS can be carried out more efficiently than RS by using the tree properties to quickly eliminate large portions of the search space. In addition, RS is the only algorithm of the three that allows for the derivation of probability of precipitation, which is an important measure in weather forecasts.

Compared to the consistency between rain rates from the MWHS-2 observations and those from GPCP ( $R = 0.96$  for all three algorithms), the comparisons with DPR produce slightly lower correlation coefficients of 0.84 (for MLR) and 0.85 (for RRS and NSS). Their biases increase by up to  $\sim 3\%$ . The reason for this degraded correlation between monthly mean gridded FY-3C data and DPR likely lies in the relatively low data density of DPR, caused by its narrow swath. This causes the DPR statistics for each  $2.5^\circ \times 2.5^\circ$  box to be noisier simply because each box holds fewer individual measurements compared to either GPCP or our own gridded results from MWHS-2. This increased noise in DPR gridded estimates is readily apparent in the global maps shown in Fig. 8 (DPR being the second panel from the top).

These rainfall retrieval algorithms are developed based on scan positions of MWHS-2 observations, which helps eliminate the concern that different footprints of a cross-track

radiometer have different local zenith angles. In addition, a unique feature of RS is that it provides probability of precipitation apart from the instantaneous rain rate estimates. The probability of precipitation provides us with an uncertainty estimate associated with each rain rate retrieval, which allows us to evaluate whether or not the observed scene is raining at all. Classical rainfall retrieval algorithms do not provide this property.

The results from the validation process suggest that the high-frequency microwave channels between 89 GHz and 190 GHz can successfully be used for measurement of rainfall. The developed algorithms can be used in conjunction to improve upon what could be accomplished with only one method alone. Moreover, this study serves as proof of concept for developing rain rate estimation algorithms using satellite observations from FY-3C/MWHS-2, which for the first time carries 118 GHz channels. This holds significance for other future sensors carrying these bands, such as NASA's TROPICS mission and EUMETSAT's MWI, both of which will contribute to near-global high temporal and spatial resolution microwave measurements at various high frequencies, including 118 and 183 GHz.

**Acknowledgements.** This work was supported by a NASA grant (Grant No. NNX17AJ09G) to Vanderbilt University. The authors thank the NASA Precipitation Processing System (PPS) for providing the GPM 2BCMB data and the China Meteorological Administration/National Satellite Meteorological Center (CMA/NSMC) for providing the MWHS-2 L1B data for this research. The authors further acknowledge and support the help of EUMETSAT and the UK Met Office for providing and maintaining RTTOV.

## REFERENCES

- Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–Present). *Journal of Hydrometeorology*, **4**, 1147–1167, [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).
- Bauer, P., and A. Mugnai, 2003: Precipitation profile retrievals using temperature-sounding microwave observations. *J. Geophys. Res.*, **108**, 4730, <https://doi.org/10.1029/2003JD003572>.
- Bentley, J. L., 1980: Multidimensional divide-and-conquer. *Communications of the ACM*, **23**, 214–229, <https://doi.org/10.1145/358841.358850>.
- Blackwell, W. J., and Coauthors, 2018: An overview of the TROPICS NASA earth venture mission. *Quart. J. Roy. Meteor. Soc.*, **144**, 16–26, <https://doi.org/10.1002/qj.3290>.
- Boukabara, S.-A., and Coauthors, 2013: A physical approach for a simultaneous retrieval of sounding, surface, hydrometeor, and cryospheric parameters from SNPP/ATMS. *J. Geophys. Res.*, **118**, 12 600–12 619, <https://doi.org/10.1002/2013jd020448>.
- Chen, F. W., and D. H. Staelin, 2003: AIRS/AMSU/HSB precipitation estimates. *IEEE Trans. Geosci. Remote Sens.*, **41**, 410–417, <https://doi.org/10.1109/TGRS.2002.808322>.
- Chen, R. Y., and R. Bennartz, 2020: Sensitivity of 89–190-GHz



- microwave observations to ice particle scattering. *J. Appl. Meteorol. Climatol.*, **59**, 1195–1215, <https://doi.org/10.1175/jamc-d-19-0293.1>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Dong, C. H., and Coauthors, 2009: An overview of a new chinese weather satellite FY-3A. *Bull. Amer. Meteor. Soc.*, **90**, 1531–1544, <https://doi.org/10.1175/2009BAMS2798.1>.
- Ferraro, R. R., and G. F. Marks, 1995: The development of SSM/I rain-rate retrieval algorithms using ground-based radar measurements. *J. Atmos. Oceanic Technol.*, **12**, 755–770, [https://doi.org/10.1175/1520-0426\(1995\)012<0755:TDSRR>2.0.CO;2](https://doi.org/10.1175/1520-0426(1995)012<0755:TDSRR>2.0.CO;2).
- Ferraro, R. R., and Coauthors, 2005: NOAA operational hydrological products derived from the advanced microwave sounding unit. *IEEE Trans. Geosci. Remote Sens.*, **43**, 1036–1049, <https://doi.org/10.1109/TGRS.2004.843249>.
- Grody, N., J. Zhao, R. Ferraro, F. Z. Weng, and R. Boers, 2001: Determination of precipitable water and cloud liquid water over oceans from the NOAA 15 advanced microwave sounding unit. *J. Geophys. Res.*, **106**, 2943–2953, <https://doi.org/10.1029/2000jd900616>.
- He, J. Y., and H. N. Chen, 2019: Atmospheric retrievals and assessment for microwave observations from Chinese FY-3C satellite during hurricane matthew. *Remote Sensing*, **11**, 896, <https://doi.org/10.3390/rs11080896>.
- Hocking, J., P. Rayer, D. Rundle, and R. Saunders, 2019: RTTOV v13 Users Guide. [Available from <https://www.nwpsaf.eu/site/software/rttov/documentation/>]
- Holmlund, K., B. Bojkov, D. Klaes, and P. Schlüssel, 2017: The joint polar system: Towards the second generation eumetsat polar system. *Proc. 2017 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Texas, USA, IEEE, 2779–2782, <https://doi.org/10.1109/IGARSS.2017.8127574>.
- Hou, A. Y., and Coauthors, 2014: The global precipitation measurement mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, <https://doi.org/10.1175/bams-d-13-00164.1>.
- Kidd, C., T. Matsui, J. Chern, K. Mohr, C. Kummerow, and D. Randel, 2016: Global precipitation estimates from cross-track passive microwave observations using a physically based retrieval scheme. *Journal of Hydrometeorology*, **17**, 383–400, <https://doi.org/10.1175/jhm-d-15-0051.1>.
- Kummerow, C. D., D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. J. Munchak, and V. Petkovic, 2015: The evolution of the goddard profiling algorithm to a fully parametric scheme. *J. Atmos. Oceanic Technol.*, **32**, 2265–2280, <https://doi.org/10.1175/jtech-d-15-0039.1>.
- Laviola, S., and V. Levizzani, 2011: The 183-WSL fast rain rate retrieval algorithm: Part I: Retrieval design. *Atmospheric Research*, **99**, 443–461, <https://doi.org/10.1016/j.atmosres.2010.11.013>.
- Mattioli, V., C. Accadia, J. Ackermann, S. Di Michele, I. Hans, P. Schlüssel, P. Colucci, and A. Canestri, 2019: The EUMET-SAT Polar System - Second Generation (EPS-SG) passive microwave and sub-mm wave missions. *Proc. 2019 Photonics & Electromagnetics Research Symp. - Spring (PIERS-Spring)*, Rome, Italy, IEEE, 3926–3933, <https://doi.org/10.1109/PIERS-Spring46901.2019.9017822>.
- Rosenfeld, D., 1999: TRMM observed first direct evidence of smoke from forest fires inhibiting rainfall. *Geophys. Res. Lett.*, **26**, 3105–3108, <https://doi.org/10.1029/1999GL006066>.
- Sanò, P., G. Panegrossi, D. Casella, F. Di Paola, L. Milani, A. Mugnai, M. Petracca, and S. Dietrich, 2015: The Passive microwave Neural network Precipitation Retrieval (PNPR) algorithm for AMSU/MHS observations: Description and application to European case studies. *Atmospheric Measurement Techniques*, **8**, 837–857, <https://doi.org/10.5194/amt-8-837-2015>.
- Saunders, R., and Coauthors, 2007: A comparison of radiative transfer models for simulating Atmospheric Infrared Sounder (AIRS) radiances. *J. Geophys. Res.*, **112**, D01S90, <https://doi.org/10.1029/2006JD007088>.
- Saunders, R., and Coauthors, 2018: An update on the RTTOV fast radiative transfer model (currently at version 12). *Geoscientific Model Development*, **11**, 2717–2737, <https://doi.org/10.5194/gmd-11-2717-2018>.
- Spencer, R. W., H. M. Goodman, and R. E. Hood, 1989: Precipitation retrieval over land and ocean with the SSM/I: Identification and characteristics of the scattering signal. *J. Atmos. Oceanic Technol.*, **6**, 254–273, [https://doi.org/10.1175/1520-0426\(1989\)006<0254:PROLAO>2.0.CO;2](https://doi.org/10.1175/1520-0426(1989)006<0254:PROLAO>2.0.CO;2).
- Staelin, D. H., and F. W. Chen, 2000: Precipitation observations near 54 and 183 GHz using the NOAA-15 satellite. *IEEE Trans. Geosci. Remote Sens.*, **38**, 2322–2332, <https://doi.org/10.1109/36.868889>.
- Surussavadee, C., and D. H. Staelin, 2010: NPOESS precipitation retrievals using the ATMS passive microwave spectrometer. *IEEE Geoscience and Remote Sensing Letters*, **7**, 440–444, <https://doi.org/10.1109/LGRS.2009.2038614>.
- Vila, D., R. Ferraro, and R. Joyce, 2007: Evaluation and improvement of AMSU precipitation retrievals. *J. Geophys. Res.*, **112**, D20119, <https://doi.org/10.1029/2007JD008617>.
- Weng, F., L. M. Zhao, R. R. Ferraro, G. Poe, X. F. Li, and N. C. Grody, 2003: Advanced microwave sounding unit cloud and precipitation algorithms. *Radio Sci.*, **38**, 8068, <https://doi.org/10.1029/2002RS002679>.
- Wilheit, T. T., 1986: Some comments on passive microwave measurement of rain. *Bull. Amer. Meteor. Soc.*, **67**, 1226–1232, [https://doi.org/10.1175/1520-0477\(1986\)067<1226:SCOPMM>2.0.CO;2](https://doi.org/10.1175/1520-0477(1986)067<1226:SCOPMM>2.0.CO;2).
- Zhang, S. W., J. Li, Z. Z. Wang, H. J. Wang, M. H. Sun, J. S. Jiang, and J. Y. He, 2012: Design of the second generation microwave humidity sounder (MWHS-II) for Chinese meteorological satellite FY-3. *Proc. 2012 IEEE Int. Geoscience and Remote Sensing Symp.*, Munich, Germany, IEEE, 4672–4675, <https://doi.org/10.1109/IGARSS.2012.6350423>.