Advanced Rain/No-Rain Classification Methods for Microwave Radiometer Observations over Land

SHINTA SETO

Institute of Industrial Science, University of Tokyo, Tokyo, Japan

TAKUJI KUBOTA

Japan Aerospace and Exploration Agency, Tsukuba, Japan

NOBUHIRO TAKAHASHI AND TOSHIO IGUCHI

National Institute of Information and Communications Technology, Koganei, Japan

TAIKAN OKI

Institute of Industrial Science, University of Tokyo, Tokyo, Japan

(Manuscript received 11 October 2007, in final form 7 April 2008)

ABSTRACT

Seto et al. developed rain/no-rain classification (RNC) methods over land for the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI). In this study, the methods are modified for application to other microwave radiometers. The previous methods match TMI observations with TRMM precipitation radar (PR) observations, classify the TMI pixels into rain pixels and no-rain pixels, and then statistically summarize the observed brightness temperature at the no-rain pixels into a land surface brightness temperature database. In the modified methods, the probability distribution of brightness temperature under no-rain conditions is derived from unclassified TMI pixels without the use of PR. A test with the TMI shows that the modified (PR independent) methods are better than the RNC method developed for the Goddard profiling algorithm (GPROF; the standard algorithm for the TMI) while they are slightly poorer than corresponding previous (PR dependent) methods. M2d, one of the PR-independent methods, is applied to observations from the Advanced Microwave Scanning Radiometer for Earth Observing Satellite (AMSR-E), is evaluated for a matchup case with PR, and is evaluated for 1 yr with a rain gauge dataset in Japan. M2d is incorporated into a retrieval algorithm developed by the Global Satellite Mapping of Precipitation project to be applied for the AMSR-E. In latitudes above 30°N, the rain-rate retrieval is compared with a rain gauge dataset by the Global Precipitation Climatology Center. Without a snow mask, a large amount of false rainfall due to snow contamination occurs. Therefore, a simple snow mask using the 23.8-GHz channel is applied and the threshold of the mask is optimized. Between 30° and 60°N, the optimized snow mask forces the miss of an estimated 10% of the total rainfall.

1. Introduction

Rain-rate estimation algorithms for microwave radiometers usually consist of two processes: rain/no-rain classification (RNC) and rain-rate retrieval. The RNC process judges whether each pixel has rain or no rain and the retrieval process is applied only for the pixels judged as rain pixels. RNC has some difficulties over land surfaces, which are radiometrically warm and highly variable. Over snow-covered and desert areas, scattering signals caused by precipitating particles are often contaminated by those due to snow and sand particles on the surface.

Seto et al. (2005, hereinafter S05) developed RNC methods for the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) (Kummerow et al.
1998) over land using a statistical database. The database statistically summarizes the land surface brightness temperature of the TMI pixels judged to be no-rain pixels by simultaneous TRMM precipitation radar (PR) observations. Therefore, these methods, called PR-dependent methods, cannot be applied to other microwave radiometers not accompanied by spaceborne precipitation radar. Currently, no radiometers other than TMI regularly conduct simultaneous observations with spaceborne precipitation radar. In the near future, the number of microwave radiometers will increase as part of the Global Precipitation Measurement (GPM) mission (Smith et al. 2007), but only one sensor called the GPM Microwave Imager will be on the same platform as the dual-frequency precipitation radar.

The purpose of this study is to modify the RNC methods of S05 so that they can be applied to other microwave sensors. Section 2 presents a review of the PR-dependent methods and the RNC of the GPROF. In section 3, we describe modification of the methods to allow production of a land surface brightness temperature database without the use of spaceborne precipitation radar. These PR-independent methods are applied to TMI observations, and the results are compared with those of PR-dependent methods. In section 4, one of the PR-independent methods is applied to the Advanced Microwave Scanning Radiometer for Earth Observation Satellite (AMSR-E). Section 5 presents a summary of this study.

### 2. Brief review of the RNC methods in S05

S05 proposed two RNC methods over land for the TMI. Here, these two PR-dependent methods are briefly reviewed, as more details can be found in S05. The scattering index (SI) defined in Eq. (1) is supposed to indicate the strength of scattering induced by precipitation:

\[
SI = TB(85.5)V_e - TB(85.5)V_{obs},
\]

where \(TB(85.5)V\) indicates the brightness temperature of 85.5 GHz at vertical polarization, \(TB(85.5)V_e\) is the estimated \(TB(85.5)V\) if there is no rain, and \(TB(85.5)V_{obs}\) is the observed \(TB(85.5)V\).

In the first method of S05, called M1, the average (\(\mu\)) and standard deviation (\(\sigma\)) of \(TB(85.5)V\) at the no-rain pixels in a \(1° \times 1°\) latitude-longitude grid for one month are calculated. When all the PR pixels matched up with a TMI pixel are no-rain, the TMI pixel is regarded as a no-rain pixel. If the SI calculated by Eq. (1) by substituting \(\mu\) into \(TB(85.5)V_e\) is larger than \(k_0 \times \sigma\) \((k_0\) is a constant), the TMI pixel is judged to be a rain pixel; otherwise, it is judged to be a no-rain pixel.

In the second method of S05, called M2, the linear regression line between TB(21.3V) and TB(85.5V) under no-rain conditions [as shown by Eq. (2)] is calculated by the least mean-square error (LMSE) method:

\[
TB(85.5)V_{no-rain} \sim a + b \times TB(21.3)V_{no-rain}
\]

and

\[
TB(85.5)V_e = a + b \times TB(21.3)V_{obs}.
\]

If the SI calculated by Eq. (1) with TB(85.5)V_e given by Eq. (3) is larger than \(k_0 \times \sigma_e\), where \(\sigma_e\) is the standard deviation of the residuals of Eq. (2), the TMI pixel is judged to be a rain pixel. If \(k_0\) is set larger (smaller), the number of “rain” pixels decreases (increases). The best value of \(k_0\) should be different from the regions and seasons, but in this study, \(k_0\) is a global constant for the simplicity of the method.

In the RNC of the Goddard profiling algorithm (GPROF) (Kummerow et al. 2001), the standard algorithm for TMI, \(TB(85.5)V_e\), is set to be \(TB(21.3)V_{obs}\) and the threshold of the SI is fixed at 8 K. The estimation of \(TB(85.5)V_e\) by M1 and that by GPROF can be written in the same form as Eq. (3). In M1, \(a = \mu\) and \(b = 0\). In GPROF, \(a = 0\) and \(b = 1\). Table 1 summarizes the RNC methods used in this paper. As the term

<table>
<thead>
<tr>
<th>RNC methods</th>
<th>Parameters in Eq. (3)</th>
<th>Parameter estimation method</th>
<th>Dependence on PR</th>
<th>Snow–desert mask</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPROF</td>
<td>(0) (1)</td>
<td>None</td>
<td>Independent</td>
<td>Snow–desert</td>
</tr>
<tr>
<td>M1</td>
<td>Variable ((\mu))</td>
<td>Moment</td>
<td>Dependent</td>
<td>None</td>
</tr>
<tr>
<td>M1_m</td>
<td>Variable ((\mu))</td>
<td>Moment</td>
<td>Dependent</td>
<td>Snow–desert (simple)</td>
</tr>
<tr>
<td>M1a</td>
<td>Variable ((\mu))</td>
<td>Fitting</td>
<td>Dependent</td>
<td>None</td>
</tr>
<tr>
<td>M1b</td>
<td>Variable ((\mu))</td>
<td>Fitting</td>
<td>Independent</td>
<td>None</td>
</tr>
<tr>
<td>M1b_m</td>
<td>Variable ((\mu))</td>
<td>Fitting</td>
<td>Independent</td>
<td>Snow–desert (simple)</td>
</tr>
<tr>
<td>M2</td>
<td>Variable</td>
<td>LMSE</td>
<td>Dependent</td>
<td>None</td>
</tr>
<tr>
<td>M2c</td>
<td>Variable</td>
<td>LMAE</td>
<td>Dependent</td>
<td>None</td>
</tr>
<tr>
<td>M2d</td>
<td>Variable</td>
<td>LMAE</td>
<td>Independent</td>
<td>None</td>
</tr>
<tr>
<td>M2d_X</td>
<td>Variable</td>
<td>LMAE</td>
<td>Independent</td>
<td>Snow (simple)</td>
</tr>
</tbody>
</table>

Table 1. Summary of the RNC methods discussed in this paper.
TB(21.3)\textsubscript{obs} is not used \((b = 0)\) in M1, the diurnal variation of TB(85.5)e induced by the physical temperature of the land surface cannot be considered. In GPROF, differences in land surface characteristics by regions and by seasons are not considered. Therefore, masks to exclude snow-covered and desert areas are required. In M1 and M2, because variation of the parameters can represent the existence of scattering signals caused by snow and sand particles on the surface, masks are not required. Therefore, precipitation over a snow-covered or desert area can be detected to some degree. As a result, it can be concluded that M2 is better than M1 and the RNC of GPROF.

3. PR-independent RNC methods

a. Modification of M1

1) Method

Drawing the probability distribution of TB(85.5)V hints at the modification of M1. Figure 1 shows the cumulative distribution function (CDF) of TB(85.5)V for grids located in China (30°–31°N, 110°–111°E) and the Sahara Desert (15°–16°N, 15°–16°E) for July 1998. The vertical axis \(z\) is shown in Gaussian scale, which can be converted into the value of CDF \(\Phi(z)\) by the CDF of the normalized Gaussian distribution function, shown by Eq. (4):

\[
\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-x^2/2) \, dx,
\]

where \(x\) is a dummy variable. If a variable completely obeys a Gaussian distribution, its CDF should be a straight line on the \(z\)-TB(85.5)V plane. The three thick lines in each panel of Fig. 1 correspond to the CDF of no-rain data, rain data, and unclassified data. No-rain or rain is judged by PR observations matched up with TMI observations, as explained in the previous section. We assume that PR observations are not available for the rain/no-rain classification. In that case, both no-rain data and rain data are included in the unclassified data. Note that unclassified data do not indicate a mixture of no-rain and rain areas in a pixel or the data outside the PR swath.

In the case in Fig. 1a, which is one of the ideal cases, the CDF of no-rain data is very close to a straight line, indicating that the no-rain data closely follow a Gaussian distribution. The CDF of no rain and the CDF of unclassified data almost overlap in the higher half of TB(85.5)V where \(z\) is positive. As scattering by precipitation lowers TB(85.5)V, the higher half of the CDF of unclassified data is mostly composed of no-rain data.

If the following conditions i and ii are satisfied, the higher half of the CDF of the unclassified distribution can be extended to the lower half \((z < 0)\), and the extrapolated line will agree with the no-rain distribution:

(i) no-rain data completely follow a Gaussian distribution and
(ii) the higher half of the CDF of unclassified data is composed of no-rain data only.

In reality, these two conditions are never perfectly satisfied, but a straight line fitted to the higher half of the CDF of unclassified data can be a good estimate of the CDF of no-rain data if the conditions are satisfied to some extent, as in the example shown in Fig. 1a. On the other hand, in the case of Fig. 1b, the no-rain data do not obey a Gaussian distribution well as the CDF is skewed toward the lower temperature side.

When a fitted line is derived as \(z = \alpha + \beta \times \text{TB(85.5)V}\), the estimates of \(\mu\) and \(\sigma\) can be given as \(-\alpha/\beta\) and \(1/\beta\), respectively. This parameter estimation technique can be used for a PR-independent RNC method, which is referred to here as M1b. With the exception of the parameter estimation, M1b is the same

![Fig. 1. The CDFs of TB(85.5)V for no-rain data (dotted line), rain data (dashed line), and unclassified data (thick solid line) for the grids located at (a) (30°–31°N, 110°–111°E) and (b) (15°–16°N, 15°–16°E) for July 1998. The thin solid line represents the estimated CDF of no-rain data.](image-url)
as M1. For comparison, a PR-dependent method called M1a is also defined. In this method, the parameters $\mu$ and $\sigma$ are estimated by fitting a straight line to the higher half of the CDF of the no-rain data. If condition i is perfectly satisfied, M1a becomes the same as M1. If condition ii is perfectly satisfied, M1b becomes the same as M1a.

2) Evaluation

The above three methods (M1, M1a, and M1b) are applied to TMI observations of a narrow swath, where simultaneous observations with PR are also available, for the entire year of 1998. The parameters $\mu$ and $\sigma$ estimated by the three methods are compared for July 1998 (Fig. 2). The parameters estimated by M1 and M1a are slightly different. It is true particularly for $\sigma$. On the other hand, the estimates by M1a and M1b are very similar for both $\mu$ and $\sigma$. Figure 2 suggests that condition i is not satisfied very well, but condition ii is almost satisfied.

The RNC is evaluated in the same way as by S05. The RNC by PR is assumed to be perfect, and evaluation indices such as RTDA [the ratio of true detection with the weight of rain amount, defined as Eq. (5)] and RFAO [the ratio of false-alarm occurrence, defined as Eq. (6)] are calculated:

$$\text{RTDA} = \frac{\text{RR(PT)}}{\text{RR(PT)} + \text{RR(Pt)}}$$

and

$$\text{RFAO} = \frac{N(pT)}{N(pT) + N(pt)}$$

where $N$ is the number of occurrences, RR is total amount of surface rainfall derived by the standard algorithm of PR (Iguchi et al. 2000), and PT, Pt, pT, and pt indicate the combinations of RNC results by PR and TMI (defined in Table 2).

Figure 3 shows the evaluation indices for the entire year of 1998. The RTDA and RFAO for M1, M1a, and M1b are calculated with $k_0$ ranging from 2.0 to 5.0 with steps of 0.1. M1 is the best method among the three, and M1a is slightly better than M1b. When $k_0$ is set to 3.5 for M1 and M1b and 4.0 for M1a so that the three methods yield almost the same RTDA, RFAO becomes 0.003, 0.009, and 0.011 for M1, M1a, and M1b.
respectively. To investigate the RFAO difference between M1 and M1b, the spatial distribution of the RFAO is shown in Fig. 4. In M1b, high RFAO values are found for the Tibetan Plateau and the Sahara Desert while such high RFAO is not visible in M1.

If condition i is perfectly satisfied, the RFAO can be determined by \( k_0 \) alone. For example, if \( k_0 = 3.5 \), the RFAO should be 0.00023. In reality, with condition i not being completely satisfied, the RFAO is usually higher than the theoretical value. The correlation coefficient between TB(85.5V) and \( z/H \) is calculated for \( z/H \) with steps of 0.1. This value is called the Gaussian distribution index (GDI) later in this paper and is used to show how well condition i is satisfied. Figure 5 shows the relationship between the GDI and RFAO for M1 and M1b (\( k_0 = 3.5 \)) in January 1998. When the GDI is close to 1, the RFAO is small both in M1 and M1b. When the GDI decreases, the RFAO tends to increase in M1b. This tendency is not obvious in M1, where the so-called moment method is used to estimate \( a \) and \( b \). The estimated \( a \) can represent the variability of no-rain data even if the no-rain data do not obey a Gaussian distribution. In M1b, condition i is explicitly assumed to estimate \( a \) and \( b \). If the CDF of no-rain data is skewed to the lower temperature side, M1b yields large RFAO.

As shown in Fig. 1b, in desert and mountainous areas, the CDF of no-rain data is often skewed to the lower temperature side and the RFAO is high. The land surface brightness temperature affected by snow and sand particles on the surface may not obey a Gaussian distribution well.

3) SNOW AND DESERT MASKS

Snow and desert masks are useful to avoid the high RFAO occurring mainly over desert and mountainous areas, but applying snow and desert masks eliminates the possibility of detecting the true precipitation signal. The PR-dependent methods of S05 are applied without snow and desert masks so that precipitation over snow-covered and desert areas can be detected to some degree. As M1b shows higher RFAO mainly over desert and mountainous regions, the following snow and desert masks are tested for M1b:

1) snow mask: when TB(21.3V) is lower than 260 K, the pixel is judged to be a snow-covered (no rain) pixel;
2) desert mask: when TB(19.3P) – TB(19.3V) is larger than 20 K, the pixel is judged to be a desert (no rain) pixel.

These masks are simple as compared with those proposed in previous studies (e.g., Ferraro et al. 1998), but we consider them sufficient for the purpose of checking the necessity of snow and desert masks in our RNC methods.

M1b with the snow and desert masks (indicated as M1b_m) gives a slightly better result than M1b, as shown in Fig. 3. In M1b_m, the high RFAO values for the Tibetan Plateau and Sahara Desert become invisible as shown in Fig. 4c. Even if the GDI decreases, the average RFAO for M1b_m shown in Fig. 5b does not increase very rapidly. When snow and desert masks are applied to M1 (indicated as M1_m), RFAO improves slightly but RTDA degrades, so that M1_m as a whole cannot be considered better than M1 (Fig. 3). This is why S05 did not employ any snow or desert masks for the PR-dependent methods.

b. Modification of M2

1) METHOD

This section describes the PR-independent version of M2, or the method of estimating the parameters \( a \) and \( b \) in Eq. (2) without the rain/no-rain classification by PR. Figure 6 shows the scatterplot of TB(21.3V) vs TB(85.5V) for the grid located at (30°–31°N, 110°–111°E) for July 1998. Rain pixels occupy approximately 10% of all the pixels in this example. In M2, the LMSE
method is applied to no-rain data to obtain a regression line \[TB(85.5V) \sim 35.161 \times 0.874 \times TB(21.3V)\]. If we apply the LMSE method to unclassified data, the obtained regression line \[TB(85.5V) \sim -142.229 + 1.500 \times TB(21.3V)\] is quite different from that by M2. Although rain data occupy a small portion of all the data, they affect the results of regression analysis significantly. In the LMSE method, parameters \(a\) and \(b\) are determined so that \(S_1\) in Eq. (7) is minimized. In this method, the larger the deviation, the greater the emphasis on such data, since the weight is proportional to the square of the deviation from the regression line. To reflect no-rain data well on the regression line, deviated data should not be emphasized. Therefore, we attempted to minimize the mean of the absolute error \(S_2\) in Eq. (8) because the available solution to minimize \(S_2\) is relatively simple and easy (Press et al. 1996):

\[
S_1 = \frac{1}{N} \sum [y - (ax + b)]^2 \quad (7)
\]

and

\[
S_2 = \frac{1}{N} \sum |y - (ax + b)|. \quad (8)
\]

The method of selecting parameters \(a\) and \(b\), which minimize \(S_2\), is referred to here as the least mean absolute error (LMAE) method. When the LMAE method is applied to unclassified data in Fig. 6, the obtained regression line \[TB(85.5V) \sim 6.712 + 0.974 \times TB(21.3V)\] is relatively close to that by M2. The LMAE method can give the regression line reflecting no-rain data well. If the LMAE method is applied to no-rain data, the obtained regression line—\(TB(85.5V) \sim 33.790 + 0.879 \times TB(21.3V)\)—is almost the same as that by M2.
2) EVALUATION

The following three methods were tested for the entire year of 1998: in a PR-independent method called M2d, parameters $a$ and $b$ are estimated by applying the LMAE to unclassified data, while the parameters are estimated by applying the LMSE to no-rain data in M2. For reference, the parameters are estimated by applying the LMAE to no-rain data in a PR-dependent method called M2c. In M2c and M2d, $\sigma_r$ is calculated as the square root of the conditional average of $[y - (ax + b)]^2$ when $y - (ax + b)$ is positive.

Figure 7 shows the evaluation indices as well as Fig. 3. M2c is slightly worse than M2, giving higher RFAO than M2 if the same $k_0$ is used. M2d is inferior to the PR-dependent methods M2 and M2c but better than the other PR-independent methods M1b and the RNC of GPROF. As compared with M1b, M2d can use the term TB(21.3V) to explain the variation of the land surface physical temperature. As compared with the RNC of GPROF, M2d can change the parameters by regions and by months. This result is encouraging regarding the superiority of M2d.

4. Application of the PR-independent RNC methods to AMSR-E

a. RNC methods for AMSR-E

1) METHOD

M2d, developed in the previous section, is applied to AMSR-E observations in this section. AMSR-E is implemented aboard the Aqua satellite launched in May 2002 on a sun-synchronous orbit with a local observation time of 1330 for ascending orbits and 0130 for descending orbits. AMSR-E has advantages over the
TMI in that it has full global coverage including the polar region and it has many overlapping samples across the scan. AMSR-E employs six frequencies of 6.9, 10.7, 18.7, 23.8, 37.0, and 89.0 GHz with both vertical and horizontal polarization (Kawanishi et al. 2003). The combination of channels is similar to that of TMI except that the TMI does not have the channel of 6.9 GHz. However, due to the slight frequency differences between AMSR-E and TMI, 23.8 and 89.0 GHz of AMSR-E are used instead of 21.3 and 85.5 GHz of the TMI, respectively. The global constant parameter \( k_0 \) is empirically set to 3.5. A simple snow mask based on only TB(23.8V) is applied; if TB(23.8V) is lower than a given threshold, the pixel is judged to be a no-rain pixel. When the threshold is \( X \) (K), M2d with this snow mask is denoted as M2d_X. M2d_0 indicates that no mask is applied to M2d.

2) Evaluation by PR

As the number of matchups between PR and AMSR-E is limited, we conducted the evaluation of the RNC method of AMSR-E by PR as a case study. In 2003, there were 1654 cases where PR and AMSR-E observed the same region within a 10-min difference. Among the 1654 cases, 666 cases occurred over land. On average, 1.8 cases per day are available for the evaluation of the RNC method over land, but the number is not large enough for comprehensive evaluation as done in section 3 for TMI. Here, we show a matchup case that includes the largest rain area among the 666 overland cases in Fig. 8. This matchup occurred from northern Uruguay to southern Brazil, around 1730 UTC 25 October 2003. Figure 8a shows the RNC result done by AMSR-E with M2d_0 and Fig. 8b shows the RNC result by PR with its standard algorithm. Though there is a time difference of nearly 7 min between the observations by AMSR-E and PR, the rain area detected by two sensors corresponds well with the RTDA of 0.927 and RFAO of 0.178.

3) Evaluation by AMeDAS rain gauge dataset

Next, the evaluation of the RNC method for AMSR-E is done with the Automated Meteorological Data Acquisition System (AMeDAS) rain gauge dataset. AMeDAS is the ground observation network of meteorological variables including precipitation, surface air temperature, humidity, and wind speed and has been developed and managed by the Japan Meteorological Agency (JMA). Rain gauges of AMeDAS are distributed almost all over Japan with an average of 17-km intervals. The temporal resolution of the rain gauge data is 1 h. The rain gauge data are converted into the 0.1° gridded dataset for the entire year of 2003. The RNC results by M2d_0 and M2d_260 for AMSR-E are matched up with the AMeDAS rain gauge dataset and the evaluation indices are calculated.

The RTDA and RFAO are calculated according to Eqs. (5) and (6), but the RNC by AMeDAS is used instead of the RNC by PR. Figure 9 shows monthly and annual scores of RTDA and RFAO for all the area where AMeDAS rain gauge data are available and the overland AMSR-E algorithm is applicable. From May to November, the evaluation scores are relatively better and the difference between M2d_0 and M2d_260 is small because the effect of snow is rarely seen in the summer to autumn season. In winter and spring (from December to April), the evaluation scores are poor in M2d_0. In M2d_260, RFAO can be improved with the snow mask, while RTDA is slightly decreased. For the annual evaluation, the RTDA is 0.420 in M2d_0 and 0.414 in M2d_260, and the RFAO is 0.026 in M2d_0 and 0.019 in M2d_260.

Figure 10 shows the spatial distribution of the annual RTDA and RFAO in M2d_260. RTDA is smaller in the northern part of Japan, particularly for Hokkaido (northern island) and the area that faces the Japan Sea. Except for the region along the coast, RFAO is not very large. Along the coast, high RFAO probably resulted from the contamination of the ocean surface and land surface, so the overland algorithm should not be used for the area very near to the coast.

b. Application to GSMaP_AMSR-E retrieval algorithm

1) Method

Next, M2d_X is combined with a rain-rate estimation algorithm for AMSR-E called GSMaP_AMSR-E,
which is developed by the Global Satellite Mapping of Precipitation (GSMaP) project (Okamoto et al. 2005). In the GSMaP_AMSR-E version 4.5 algorithm utilized here, surface rainfall rates are retrieved from the observed polarization correction temperature at 89.0 GHz using a radiative transfer model (RTM) developed by Liu (1998), with statistically classified vertical profiles of precipitation observed by PR and background me-

FIG. 8. Comparison of rain/no-rain classification by (a) AMSR-E with RNC method M2d and (b) PR with its standard algorithm for the matchup case occurring in South America around 1730 UTC 25 Oct 2003. Dark blue indicates estimated rain areas and dark red indicates estimated no-rain areas. The results of AMSR-E outside the TRMM/PR swath are shown in light colors.
teorological variables from a global analysis dataset provided by JMA. The sensitivity of the brightness temperature to precipitation and surface parameters is discussed by S05 with the RTM by Liu (1998). Further descriptions of the latest development of the algorithm by the GSMaP project are given in Kubota et al. (2007). In the present paper, the GSMaP_AMSR-E algorithm is equipped with the M2d method instead of the original RNC method.

GSMaP_AMSR-E with M2d_X was applied for all the AMSR-E observations over land between 60°S and 60°N and for the entire year of 2003. This result is referred to as GA_M2d_X.

2) COMPARISON WITH GPCC

We used the monitoring product of the Global Precipitation Climatology Center (GPCC) rain gauge analysis dataset (denoted as GPCC_Gauge) for evaluation of the estimates in latitudes higher than 30°N where the surface can be covered by snow. The GPCC_Gauge has a temporal resolution of 1 month and spatial resolution of 1° × 1° latitude–longitude [the latest information about this dataset was available online at the time of writing at http://gpcc.dwd.de/]. Although GPCC_Gauge suffers from problems, such as undercollection of snowfall and low sampling number, especially in cold regions, we consider it the best currently available, global, gauge-based precipitation dataset. The grid data of GA_M2d_X are prepared with the same resolution.

The maps in Fig. 11 show the annual average of the estimated rain amount, with supplemental figures presenting zonal and seasonal variations.

According to the GPCC_Gauge (Fig. 11a), rain amounts >200 mm month\(^{-1}\) can be seen in limited areas, such as the southern part of Japan and the Pacific Ocean side of Canada and Alaska. In other areas, the rain amount generally does not exceed 100 mm month\(^{-1}\). Rain amount is relatively large from June to October in areas higher than 45°N.
In GA_M2d_0, severe overestimation can be seen around the Tibetan Plateau and the western side of the Caspian Sea (Fig. 11b), while the apparent overestimation almost disappears in GA_M2d_260 (Fig. 11c). In January, the overestimation is severe around 40°N. In April and May, severe overestimation is found at higher latitudes. Severe overestimation in GA_M2d_0 occurs in the boundary region between snow-covered areas and snow-free areas rather than in fully snow-covered areas.

Figure 12 shows scatterplots of TB(23.8V) and TB(89.0V) for the grid located at (54°–55°N, 49°–50°E) for January and April. This grid is selected just as an example of the area where snow cover changes by seasons. According to the GPCC_Gauge, the monthly rainfall is not very large in either month; therefore, a large part of the data should be composed of no-rain pixels. In April, the scatterplot consists of two parts: the main part along the line TB(89.0V) = TB(23.8V) and the flagging part along the line TB(23.8V) = 260 K. It is impossible to draw a straight line reflecting the two parts. The flagging part is probably affected by scattering by snow particles on the surface, not by scattering by precipitation. The regression line drawn by M2d reflects the main part, and the flagging part is consequently judged to be rain data, leading to severe overestimation of rainfall. Therefore, M2d requires a snow mask, as does the RNC method of GPROF. On the other hand, in January, the scatterplot mostly consists of data affected by scattering by snow particles on the surface. It is easy to draw the regression line for M2d, and overestimation does not occur.

3) MASK OPTIMIZATION

The above analysis showed that a snow mask is necessary when applying M2d at high latitudes. Next, we optimize the snow mask threshold to remove false alarms and to minimize the missing of true rainfall by the snow mask.

Figure 13 shows the estimated average rain amount over all land between 30° and 60°N for different thresholds. The horizontal axis in Fig. 13 is shown as the target ratio (TR), which is calculated as the ratio of mask-free pixels to total pixels. When no mask is applied (X = 0), TR should be 1. When the threshold becomes high, TR decreases and the estimated rain amount decreases. When TR is 0 or all the pixels are masked, the estimated rain amount is of course zero. In Fig. 13, we drew lines for the GPCC_Gauge as well as for GA_M2d_X. The original value by the GPCC_Gauge is set at TR = 1 (this value is denoted GG_0) and the zero is set at TR = 0, and the estimates are linearly interpolated between TR = 0 and 1 so that GPCC_Gauge estimates at TR are calculated as GG_0 × TR. GG_X indicates the GPCC_Gauge estimates masked with the threshold of X (K), and thus calculated as GG_0 multiplied by the TR corresponding with X (K). Note that the linear interpolation assumes that the rain rate is independent of TB(23.8V).
Without the mask, GA_M2d_0 yields larger estimates than GG_0. When the threshold is 260 K, GA_M2d_260 gives 39.7 mm month\(^{-1}\), which is less than the estimate by GG_0 (46.6 mm month\(^{-1}\)) but larger than that by GG_260 (35.9 mm month\(^{-1}\)). The GA_M2d_260 value is apparently underestimated by 6.9 (46.6 – 39.7) mm month\(^{-1}\) but rainfall of 10.7 (46.6 – 35.9) mm month\(^{-1}\) is also missed by the snow mask. For the mask-free pixels, GA_M2d_260 overestimates against GPCC_Gauge by 3.8 (39.7 – 35.9) mm month\(^{-1}\). To eliminate the difference between GA_M2d_X and GG_X, the threshold \(X\) should be set to approximately 262 K. GA_M2d_262 and GG_262 are almost the same (34.4 mm month\(^{-1}\)). In this case, the rainfall missed by the snow mask is 12.2 (46.6 – 34.4) mm month\(^{-1}\). When the estimate of GA_M2d_X is equal to that of GG_X, the value of \(X\) is said to be the optimized threshold for M2d.

Next, the optimized threshold is determined separately for each grid and month. If the estimate by AMSR-E without the snow mask is less than the original GPCC estimates, the optimized threshold is regarded as 0 K and the missed rainfall is considered to be zero. The missed rainfall by the optimized snow mask is calculated. Figure 14 shows the annual average of missed rainfall. The global average is 4.7 mm month\(^{-1}\), but rainfall of 10.7 (46.6 – 35.9) mm month\(^{-1}\) is also missed by the snow mask. For the mask-free pixels, GA_M2d_260 overestimates against GPCC_Gauge by 3.8 (39.7 – 35.9) mm month\(^{-1}\). To eliminate the difference between GA_M2d_X and GG_X, the threshold \(X\) should be set to approximately 262 K. GA_M2d_262 and GG_262 are almost the same (34.4 mm month\(^{-1}\)). In this case, the rainfall missed by the snow mask is 12.2 (46.6 – 34.4) mm month\(^{-1}\). When the estimate of GA_M2d_X is equal to that of GG_X, the value of \(X\) is said to be the optimized threshold for M2d.

The optimized threshold is determined separately for each grid and month. If the estimate by AMSR-E without the snow mask is less than the original GPCC estimates, the optimized threshold is regarded as 0 K and the missed rainfall is considered to be zero. The missed rainfall by the optimized snow mask is calculated. Figure 14 shows the annual average of missed rainfall. The global average is 4.7 mm month\(^{-1}\), but rainfall of 10.7 (46.6 – 35.9) mm month\(^{-1}\) is also missed by the snow mask. For the mask-free pixels, GA_M2d_260 overestimates against GPCC_Gauge by 3.8 (39.7 – 35.9) mm month\(^{-1}\). To eliminate the difference between GA_M2d_X and GG_X, the threshold \(X\) should be set to approximately 262 K. GA_M2d_262 and GG_262 are almost the same (34.4 mm month\(^{-1}\)). In this case, the rainfall missed by the snow mask is 12.2 (46.6 – 34.4) mm month\(^{-1}\). When the estimate of GA_M2d_X is equal to that of GG_X, the value of \(X\) is said to be the optimized threshold for M2d.

Next, the optimized threshold is determined separately for each grid and month. If the estimate by AMSR-E without the snow mask is less than the original GPCC estimates, the optimized threshold is regarded as 0 K and the missed rainfall is considered to be zero. The missed rainfall by the optimized snow mask is calculated. Figure 14 shows the annual average of missed rainfall. The global average is 4.7 mm month\(^{-1}\), but rainfall of 10.7 (46.6 – 35.9) mm month\(^{-1}\) is also missed by the snow mask. For the mask-free pixels, GA_M2d_260 overestimates against GPCC_Gauge by 3.8 (39.7 – 35.9) mm month\(^{-1}\). To eliminate the difference between GA_M2d_X and GG_X, the threshold \(X\) should be set to approximately 262 K. GA_M2d_262 and GG_262 are almost the same (34.4 mm month\(^{-1}\)). In this case, the rainfall missed by the snow mask is 12.2 (46.6 – 34.4) mm month\(^{-1}\). When the estimate of GA_M2d_X is equal to that of GG_X, the value of \(X\) is said to be the optimized threshold for M2d.

The optimized threshold is determined separately for each grid and month. If the estimate by AMSR-E without the snow mask is less than the original GPCC estimates, the optimized threshold is regarded as 0 K and the missed rainfall is considered to be zero. The missed rainfall by the optimized snow mask is calculated. Figure 14 shows the annual average of missed rainfall. The global average is 4.7 mm month\(^{-1}\), but rainfall of 10.7 (46.6 – 35.9) mm month\(^{-1}\) is also missed by the snow mask. For the mask-free pixels, GA_M2d_260 overestimates against GPCC_Gauge by 3.8 (39.7 – 35.9) mm month\(^{-1}\). To eliminate the difference between GA_M2d_X and GG_X, the threshold \(X\) should be set to approximately 262 K. GA_M2d_262 and GG_262 are almost the same (34.4 mm month\(^{-1}\)). In this case, the rainfall missed by the snow mask is 12.2 (46.6 – 34.4) mm month\(^{-1}\). When the estimate of GA_M2d_X is equal to that of GG_X, the value of \(X\) is said to be the optimized threshold for M2d.

5. Summary

In this paper, the RNC methods developed by S05 for the TMI over land were modified to be independent of PR so that they can be applied for other microwave radiometers. M1b, which is a PR-independent version of M1, estimates the parameters \(\mu\) and \(\sigma\) by fitting a Gaussian distribution to the higher temperature half of the CDF of unclassified TB(85.5V). M2d, which is a
PR-independent version of M2, obtains the regression line between TB(21.3V) and TB(85.5V) by the LMAE method. The accuracies of M1b and M2d are slightly worse than those of M1 and M2 when applied for TMI observations. Among the PR-independent methods, M2d has the best accuracy. M2d is better than M1b as M2d uses the term TB(21.3V); M2d is also better than the RNC of GPROF as M2d changes the parameters by regions and by seasons. M2d is applied to AMSR-E and is evaluated by using the PR and AMeDAS rain gauge dataset. M2d is combined with the retrieval algorithm of GSMaP_AMSR-E and applied for the entire year of 2003. In latitudes higher than 30°N, the estimated monthly rainfall is compared with GPCC_Gauge. Without the snow mask, GA_M2d yields overestimation against GPCC_Gauge, particularly where snow-covered and snow-free pixels coexist. It shows that a snow mask is unavoidable when M2d is applied in mid–high latitudes. The snow mask tested in this study is simple because the purpose was to check the necessity of the snow mask. In future applications, a more sophisticated snow mask using texture information as the RNC of GPROF should be used. The optimization of the threshold shown in section 4b(3) can improve the snow mask. The use of a rain gauge dataset to determine the threshold, although this technique is not available for real-time processing, can be considered a satellite–gauge merged dataset.

The PR-independent methods developed in this study are simple enough to be applied to other microwave radiometers. We are planning to apply this method to GSMaP retrieval algorithms for various microwave radiometers. The database for the GSMaP algorithm for the TMI can be modified. For previous PR-dependent methods, database values are lacking for latitudes between 37° and 39°, where TMI can observe but PR cannot. PR-independent methods can fill this gap. Despite extensive effort, limitations remain if we use only the 80–90-GHz channel to detect precipitation over land. We would like to also apply the PR-independent methods for microwave sounders, which employ different frequencies, such as 50–60 and 180–190 GHz.

Acknowledgments. This study is part of a larger study called Production of a High-Precision, High-Resolution Global Precipitation Map Using Satellite Data (also called GSMaP), led by Prof. Ken’ichi Okamoto (Osaka Prefecture University, Japan) under the Core Research for Evolutional Science and Technology (CREST) program of the Japan Science and Technology Agency (JST). The AMSR-E brightness temperature data were provided by the Japan Aerospace Exploration Agency (JAXA). For matchup data between AMSR-E and PR, Dr. Keiji Imaoka at JAXA kindly gave us valuable information. We really appreciate their cooperation.

REFERENCES


