Bayesian Estimation of Precipitation from Satellite Passive Microwave Observations Using Combined Radar–Radiometer Retrievals

MIRCEA GRECU

Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, and NASA Goddard Space Flight Center, Greenbelt, Maryland

WILLIAM S. OLSON

Joint Center for Earth Systems Technology, University of Maryland, Baltimore County, Baltimore, and NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 26 April 2005, in final form 14 October 2005)

ABSTRACT

Precipitation estimation from satellite passive microwave radiometer observations is a problem that does not have a unique solution that is insensitive to errors in the input data. Traditionally, to make this problem well posed, a priori information derived from physical models or independent, high-quality observations is incorporated into the solution. In the present study, a database of precipitation profiles and associated brightness temperatures is constructed to serve as a priori information in a passive microwave radiometer algorithm. The precipitation profiles are derived from a Tropical Rainfall Measuring Mission (TRMM) combined radar–radiometer algorithm, and the brightness temperatures are TRMM Microwave Imager (TMI) observed. Because the observed brightness temperatures are consistent with those derived from a radiative transfer model embedded in the combined algorithm, the precipitation–brightness temperature database is considered to be physically consistent. The database examined here is derived from the analysis of a month-long record of TRMM data that yields more than a million profiles of precipitation and associated brightness temperatures. These profiles are clustered into a tractable number of classes based on the local sea surface temperature, a radiometer-based estimate of the echo-top height (the height beyond which the reflectivity drops below 17 dBZ), and brightness temperature principal components. For each class, the mean precipitation profile, brightness temperature principal components, and probability of occurrence are determined. The precipitation–brightness temperature database supports a radiometer-only algorithm that incorporates a Bayesian estimation methodology. In the Bayesian framework, precipitation estimates are weighted averages of the mean precipitation values corresponding to the classes in the database, with the weights being determined according to the similarity between the observed brightness temperature principal components and the brightness temperature principal components of the classes. Because the classes are stratified by the sea surface temperature and the echo-top-height estimator, the number of classes that are considered for retrieval is significantly smaller than the total number of classes, making the algorithm computationally efficient. The radiometer-only algorithm is applied to TMI observations, and precipitation estimates are compared with combined TRMM precipitation radar (PR)–TMI reference estimates. The TMI-only algorithm, supported by the empirically derived database, produces estimates that are more consistent with the reference values than the precipitation estimates from the version-6 TRMM facility TMI algorithm. Cloud-resolving model simulations are used to assign a latent heating profile to each precipitation profile in the empirically derived database, making it possible to estimate latent heating using the radiometer-only algorithm. Although the evaluation of latent heating estimates in this study is preliminary, because realistic conditional probability distribution functions are attached to latent heating structures in the algorithm's database, a generally positive impact on latent heating estimation from passive microwave observations is expected.

Corresponding author address: Dr. Mircea Grecu, GEST/UMBC, NASA Goddard Space Flight Center, Code 613.1, Greenbelt, MD 20771.
E-mail: grecu@agnes.gsfc.nasa.gov

© 2006 American Meteorological Society

1. Introduction

Precipitation estimation from satellite passive microwave radiometer data is a mathematically ill-posed problem. That is, the problem does not have a unique solution that is insensitive to errors in the input data. Traditionally, to make the problem well posed, a priori information derived from physical models or independent, high-quality observations is incorporated into the solution. For example, the algorithm used to estimate precipitation from observations provided by the microwave radiometer aboard the Tropical Rainfall Measuring Mission (TRMM) satellite employs a database derived from cloud-resolving model (CRM) simulations (Kummerow et al. 2001). This database consists of CRM-simulated precipitation profiles and associated brightness temperatures derived from radiative transfer calculations. For each multispectral set of observed brightness temperatures, the estimation procedure assigns to each precipitation profile in the database the probability that it is the "observed" profile. The final estimate is a probabilistic combination of the precipitation profiles in the database. Although other approaches are possible, for example, the direct maximization of the posteriori probability distribution of brightness temperatures (Evans et al. 1995), precipitation databases or parameterizations (which can be viewed as a compact way of representing the information in a database) are necessary in any passive microwave estimation algorithm.

In many instances, CRMs represent a useful tool for constructing the databases or the parameterizations needed in passive microwave retrievals. They are physically based, flexible, and can shed light on phenomena difficult to understand and investigate from direct observations. However, there are also drawbacks associated with the use of CRMs. First, the precipitation distribution in nature may be different from that simulated by CRMs. This is because CRMs are initialized using a relatively small set of large-scale environmental conditions that may not be statistically representative of the distribution of the large-scale environments in nature. Second, CRMs might be deficient in handling ice processes, resulting in distributions of simulated 85-GHz brightness temperatures different from those observed in nature (Bauer 2001). These drawbacks may be responsible for the differences between precipitation estimates from the version-5 TRMM precipitation radar (PR) and TRMM Microwave Imager (TMI) algorithms (Kummerow et al. 2001). Although these differences are reduced in the version-6 algorithms, some discrepancies still exist between the TRMM PR and TMI estimates (these will be illustrated in section 3). From this perspective, but also for the benefit of future precipitation satellite missions, it is desirable to construct a precipitation–brightness temperature database free from the weaknesses of databases derived from current CRM simulations. Such a database can be constructed from a large set of precipitation profiles and associated brightness temperatures derived directly from observations.

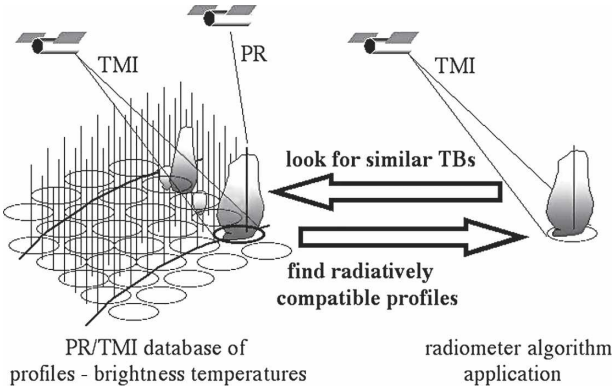

FIG. 1. Schematic of the radiometer-only precipitation profile estimation method. (left) TRMM PR and TMI observations are combined to retrieve precipitation profiles (vertical lines) in the overlap swath. These retrieved profiles form the precipitation–brightness temperature database for the TMI-only algorithm. (right) TMI brightness temperatures (TBs) are used to find radiatively compatible precipitation profiles in the TRMM PR–TMI database. The compatible profiles are combined to form the solution profile.

Our approach to the radiometer-only algorithm is illustrated in Fig. 1. First, we construct a database of coincident precipitation profiles and associated brightness temperatures using TRMM combined active and passive microwave observations. The combined algorithm developed by Grecu et al. (2004) is used to retrieve precipitation profiles from the TRMM PR and TMI observations. The application of the combined algorithm to 1 month of TRMM observations over ocean regions yields a database of more than one million retrieved profiles and coincident TMI brightness temperatures. The database is organized to facilitate the efficient application of a Bayesian algorithm to estimate precipitation profiles from radiometer-only observations. In essence, the database of precipitation profiles is searched to find profiles that are compatible with the TMI observations and ancillary data; the compatible profiles are then combined to form a solution profile.

Radiometer algorithm databases derived from TRMM observations have been considered before by Shin and Kummerow (2003). They illustrated the potential of such empirically based Bayesian estimation

algorithms but provided only a partial evaluation of their algorithm, because only simulation-based experiments were considered in their study. In their experiments, brightness temperatures were synthesized from combined retrievals and used for precipitation retrieval from radiometer observations only. The radiometer-only retrievals were then compared with the combined retrievals used to synthesize the passive observations. Evaluations of this kind using simulated radiances are not entirely representative, because applications to real observations may be affected by errors in the radiative transfer calculations (and consequently the algorithm may perform worse than in the synthetic experiments). Also, because of computational restrictions, only the profiles within two latitude–longitude boxes with dimensions of 15° × 30° over the Pacific Ocean during December 1999 were used to construct and test the databases in the Shin and Kummerow (2003) study. A follow-up study by Masunaga and Kummerow (2005) refined the combined retrieval methodology used to construct the database, but no evaluation of its impact on radiometer-only retrievals was considered.

In the current study, a more comprehensive evaluation of a Bayesian algorithm supported by combined TRMM retrievals is performed. Specifically, we compare estimates from the Bayesian algorithm formulated in this study with version-6 TRMM facility TMI algorithm estimates, and to estimates from the combined algorithm of Grecu et al. (2004). The passive microwave observations are provided by the TMI, and the comparison is made for monthly estimates at 0.5° × 0.5° resolution. This approach provides direct evidence for whether or not some of the deficiencies in Bayesian algorithms using CRM databases can be eliminated and more consistency with radar–radiometer-based estimates can be achieved using empirically derived databases. Another notable difference from the formulation of Shin and Kummerow (2003) is the use of a radiometer-based estimator of the echo-top height (ET; the height beyond which the reflectivity drops below 17 dBZ, i.e., storm height in the terminology of the TRMM PR products). This estimator facilitates an efficient exploration of the database in the Bayesian algorithm, but in addition, it helps to establish the connection between precipitation and latent heating (LH) profiles in the database. Although estimation of precipitation profiles is the primary focus of this investigation, latent heating estimation is another important algorithm application that is demonstrated.

In section 2 a description of the database construction and the Bayesian estimation algorithm is provided. Results from the application of the newly derived Bayesian algorithm are discussed in section 3. In section 4, the Bayesian algorithm for LH estimation is described and demonstrated. Conclusions and recommendations for future work are provided in section 6.

2. Formulation

a. Construction of a combined PR–TMI database

In this study, the combined radar–radiometer algorithm of Grecu et al. (2004) is applied to coincident TRMM PR and TMI observations over ocean regions to construct a large database of coincident precipitation profiles and associated brightness temperatures. It should be mentioned here that because the TRMM observations are limited to latitudes between 37°S and 37°N, an empirical database constructed from TRMM data is only appropriate for algorithm applications in the Tropics and subtropics. However, as already mentioned, there is a benefit in considering alternatives to CRM databases even in these regions. Furthermore, the methodology can be applied to data collected in future satellite precipitation missions to extend the database derived from TRMM. Although the algorithm is applicable to both overland and overocean radiometer data, provided that both overland and overocean combined estimates are included in the database, only the overocean data are considered in the current investigation. This is because the rain signature in radiometer observations is less specific over land, and often it is difficult to distinguish rain signatures from variations in surface emission. For this reason, in the past, passive microwave remote sensing of precipitation over land has been considered separately from the overocean estimation problem (Petty and Krajewski 1996). Here, the focus is exclusively on precipitation estimation over ocean surfaces.

The combined radar–radiometer algorithm (Grecu et al. 2004) is based on physical models that simulate high-resolution brightness temperatures as functions of observed reflectivity profiles and two additional free variables associated with each profile. The two additional variables are the intercept in a normalized gamma precipitation particle size distribution and the mean density of the ice phase precipitation. These variables are determined in an optimal estimation framework that minimizes the differences between simulated and observed normalized polarizations, as defined by Petty (1994) (see Grecu et al. 2004). The second variable was not considered in the original formulation of Grecu et al. (2004) because it does not significantly affect the estimation of surface precipitation, but it is important for the accurate estimation of ice phase precipitation, and therefore it is considered in the current combined retrievals used for database construction. The repeated

application of the combined algorithm to TRMM data yields a large database of precipitation profiles and associated TMI brightness temperatures. (This database will be referenced henceforth as the combined TRMM PR–TMI database.) In the current study, because the database is further used for precipitation estimation from TMI observations, the observed brightness temperatures as well as those simulated by the algorithm procedure are stored in the database. The reason for storing both the simulated and observed brightness temperatures is that there are small but nonnegligible random differences between the simulated and observed brightness temperatures, especially at high frequencies. To evaluate the impact of these differences on the retrievals, one can perform retrievals using both the simulated and observed brightness temperatures from the database and compare the results. The use of observed TMI brightness temperatures from the database is expected to lead to smaller errors because it ensures the statistical consistency of brightness temperatures and combined TRMM PR–TMI-estimated profiles, and all of the errors in the TMI-only retrievals can therefore be attributed to ambiguities in the database. Henceforth, the Bayesian TMI-only algorithm supported by the combined TRMM PR–TMI database will be called the TRMM PR–TMI-based TMI algorithm. In the future, if the Bayesian radiometer algorithm is to be applied to other sensors [such as the Special Sensor Microwave Imager (SSM/I)], then brightness temperatures must be simulated at the sensors' resolutions and frequencies from the combined TRMM PR–TMI retrievals and included in the precipitation profile–brightness temperature database.

The spatial resolution of the TRMM PR–TMI-combined estimates of Grecu et al. (2004) is that of TRMM PR observations, that is, approximately 4.3 km. The precipitation profiles saved in the database are obtained by averaging 3×3 arrays of combined TRMM PR–TMI precipitation profiles. Consequently, they have a resolution of approximately 13 km. We chose to use the averaged lower-resolution profiles instead of the initial profiles for two reasons—first, to make our retrievals more consistent with the Goddard profiling algorithm (GPROF) retrievals (the GPROF retrievals have a similar resolution), and second, to reduce the random errors in retrievals (we noted, based on various numerical experiments that we carried out, that the retrieval performance increases when the estimate resolution decreases, which makes the 13-km resolution a good compromise between accuracy and resolution).

A large dataset of precipitation profiles and associated brightness temperatures is necessary to ensure the correct joint probability distribution function of precipitation and brightness temperatures. In this study, we consider 1 month of TRMM observations over ocean to construct the dataset and perform a sensitivity analysis to determine whether this amount of data is sufficient to obtain stable estimates. One month of TRMM data yields more than one million coincident (close to subsatellite point) overocean TRMM PR and TMI observations, including nonprecipitating TRMM PR–TMI profiles (zero rainfall rate at all heights). Our TMI-only rain algorithm involves two steps. In the first step, a discrimination procedure is used to determine the potentially raining pixels. This procedure is the same as the one used in Kummerow et al. (2001). Additional information on the discrimination may be found in Ferraro et al. (1998). In the second step, a Bayesian algorithm is applied only to the potentially raining pixels while the other pixels are assigned zero rain. To ensure the consistency between the retrieval procedure and the database construction procedure, all combined TRMM PR–TMI profiles that are classified as potentially raining by the TMI-only rain/no-rain discrimination procedure are included in the database. The combined TRMM PR–TMI raining pixels wrongly classified as nonraining by the TMI-only procedure contribute less than 2% of the total precipitation, and therefore errors in the TMI-only rain/no-rain discrimination do not have a large impact on the final results.

The amount of data in the combined TRMM PR–TMI database is prohibitively large for a Bayesian estimation algorithm. To reduce the computational burden the following three operations are performed:

1) The empirical orthogonal functions of all observed brightness temperatures over ocean during 1 month are determined and the principal components are evaluated,

$$\mathbf{PC}_{\mathrm{TB}} = \mathbf{T}_B^O\,\mathbf{EOF}, \tag{1}$$

where $\mathbf{PC}_{\mathrm{TB}}$ is a matrix of $n \times m$ elements (n is the number of profiles and $m = 9$ is the number of TMI channels) containing the principal components, **EOF** is an $m \times m$ matrix of orthogonal functions that makes the transformation from the brightness temperature space to the principal component space, and $\mathbf{T}_B^O$ is a matrix of $n \times m$ elements containing the observed brightness temperatures in the combined TRMM PR–TMI database. Equation (1) provides a more compact representation of the TMI's nine brightness temperatures in a smaller number of variables, because a linear combination of only a few empirical orthogonal functions (**EOF** columns) represents most of the variance of the brightness temperature observations. Therefore, the

projection of the brightness temperatures onto the empirical orthogonal functions results in a matrix $\mathbf{PC}_{TB}$ with only a few significant columns.

2) The precipitation profiles and the associated brightness temperatures are binned by a climatologic estimate of the sea surface temperature (SST) and a TMI-based statistical estimator of the ET. The bin size is 3 K for SST and 1 km for ET.
3) For each (SST, ET) bin, the precipitation profiles with similar principal components ($\mathbf{PC}_{TB}$) are clustered into classes containing approximately 40 profiles, and all of the profiles in the same class are replaced by their average. Only the first five principal components are used in the clustering procedure. These principal components are subsequently used in the estimation procedure as well. The first five principal components explain 97% of the brightness temperature variance. The principal components corresponding to the average profile in a class are determined by averaging the principal components of the profiles in that class. It is found that using more than five principal components does not improve the estimation performance, and therefore only five principal components are used in the study. As already mentioned, the clustering operation is justified by the large number of profiles in the database. Because at the retrieval time only brightness temperature or $\mathbf{PC}_{TB}$ information is available, the $\mathbf{PC}_{TB}$-based clustering appears more natural than clustering based on the precipitation profile similarity. This is because similar rain profiles may have significantly different $\mathbf{PC}_{TB}$ vectors, and by averaging in the $\mathbf{PC}_{TB}$ space when clustering is based on rain profile similarity, one may get an average that is significantly different from any single $\mathbf{PC}_{TB}$ vector in a given class. This results in a rain distribution conditioned on average $\mathbf{PC}_{TB}$ that may be completely different from the initial (before clustering) rain distribution conditioned on $\mathbf{PC}_{TB}$. On the other hand, the clustering based upon $\mathbf{PC}_{TB}$ (used in the current study) does not change the conditional rain distribution. It just replaces conditional rain distributions by their mean values, but this operation does not affect the final estimate, which is a probabilistic average of the means of these distributions.

The clustering is achieved using an iterative algorithm. First, the number of classes is determined as

$$\mathrm{nc} = \mathrm{int}\left[\frac{n(\mathrm{SST}, \mathrm{ET})}{40}\right] + 1, \tag{2}$$

where n(SST, ET) is the number of profiles within bin (SST, ET) and int[x] is the greatest integer less than or equal to x. Second, the profiles are randomly assigned to the nc classes. Third, the average principal components are determined for each class. Fourth, the profiles are reassigned to the class with the closest average principal components. Steps three and four are repeated until convergence is achieved. Additional details on this clustering procedure and an example thereof can be found in Jain et al. (1999). The approximate number of profiles in a class, that is, 40, is determined by trial and error. A smaller number of profiles per class does not improve the retrieval performance and just increases the computational effort, while a larger number of profiles per class makes the distribution of estimates look discrete for high-intensity rain situations.

These operations significantly reduce the number of profiles in the database and make the Bayesian estimation more efficient. They also provide a mechanism to evaluate the uncertainties in the retrieval: in addition to the mean, the standard deviation of precipitation corresponding to the profiles in a given class is stored in the database, and this standard deviation represents a measure of the uncertainty of an estimate for that specific class. A formal estimate of precipitation uncertainty for a given set of observations is provided in section 2b.

Step 2 of the database aggregation procedure requires the estimation of the echo-top height from TMI-only observations. This is accomplished using a nonlinear regression (neural network) to predict the ET as a function of the first five principal components ($\mathbf{PC}_{TB}$) and the climatological estimate of the SST. Mathematically, this is equivalent to solving, in the least squares sense, the following equation:

$$\mathrm{ET} = f(\mathbf{PC}_{TB}, \mathrm{SST}, \mathbf{w}), \tag{3}$$

where ET is the echo-top height of a given TRMM PR profile, $\mathbf{PC}_{TB}$ and SST are the associated principal components and sea surface temperatures, respectively, and $\mathbf{w}$ is a set of parameters to be determined. Here, the TRMM PR ET is not the actual cloud-top height but the height where the reflectivity drops below the value that the radar can reliably detect. Nevertheless, the TRMM PR ET is an important physical variable in characterizing the physical processes associated with the precipitation depth. The function f is a two-layer feed-forward neural network, that is, a nonlinear function that graphically resembles a network of neurons. The determination of $\mathbf{w}$ is done using a gradient-based minimization algorithm. Simpler functional forms may be used for f, for example, just a linear function, but the ET estimation performance increases significantly with the complexity of f. For example, the 85-GHz polarization-corrected temperature (PCT) of Spencer et al.

(1989) is a reasonably good predictor of the ET, that is, the correlation coefficient between the 85-GHz PCT and the ET is 0.7. However, the neural network provides a more accurate and still computationally efficient ET estimator. Thus, the correlation coefficient between the neural network ET estimate and the TRMM PR ET is 0.78, and the relative root-mean-square is about 30% smaller than that of the 85-GHz PCT-based estimator. In the top panel of Fig. 2, a two-dimensional histogram of the TRMM PR-observed ET versus the predicted ET is shown. One may note that the most frequent counts occur around the one-to-one line, which indicate a good performance of the estimator. Histograms of the TRMM PR and predicted ETs are shown in the bottom panel of Fig. 2. The agreement between the observed and predicted histograms also indicates good performance.

Once the combined TRMM PR–TMI database is constructed and organized as described above, a Bayesian algorithm can be formulated to estimate precipitation from TMI-only observations, as described in section 2b.

b. *TMI-only Bayesian algorithm for precipitation estimation*

The algorithm for precipitation estimation from TMI-only observations developed in this study differs to some extent from other Bayesian estimation algorithms [e.g., the algorithm developed by Kummerow et al. (2001)]. The differences stem from the necessity of making efficient use of the large a priori database. As previously described, the database constructed from the combined TRMM PR–TMI retrievals does not consist of individual profiles, but averages over small sets of profiles clustered by their similarity in the space of brightness temperature principal components. Therefore, given a set of principal components and estimates of SST and ET, a precipitation profile can be estimated as

$$\hat{E}(\mathbf{R}|\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET}) = \sum_{i=1}^{\mathrm{nc}} \mathbf{R}_i \cdot P(C_i|\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET}), \tag{4}$$

where nc is the number of classes within the (SST, ET) bin, $\mathbf{R}_i$ is the average precipitation profile of class C_i, and $P(C_i|\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET})$ is the probability that the observed profile lies in class C_i. Probability $P(C_i|\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET})$ can be estimated using Bayesian classification techniques (Duda et al. 2000). Thus, using Bayes's theorem,

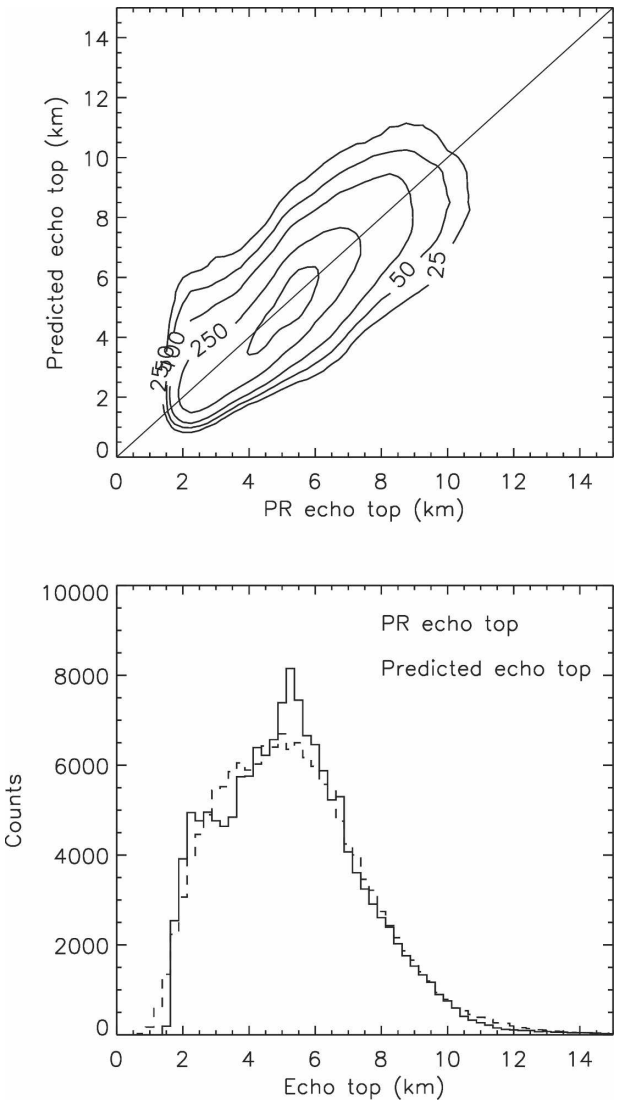

FIG. 2. (top) Joint and (bottom) unconditional distributions of TMI-predicted and TRMM PR-observed echo-top heights.

$$P(C_i|\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET}) = \frac{P(\mathbf{PC}_{\mathrm{TB}}|C_i)P(C_i)}{P(\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET})}. \tag{5}$$

Variables SST, ET do not explicitly appear in $P(\mathbf{PC}_{\mathrm{TB}}|C_i)$ and $P(C_i)$ because class C_i implicitly corresponds to bin (SST, ET). Assuming that the principal components are normally distributed within the classes, probability $P(\mathbf{PC}_{\mathrm{TB}}|C_i)$ can be estimated from

$$P(\mathbf{PC}_{\mathrm{TB}}|C_i) = \frac{\exp[-0.5(\mathbf{PC}_{\mathrm{TB}}^{O} - \mathbf{PC}_{\mathrm{TB}}^{i})^{\mathrm{T}}\, \mathbf{S}_i^{-1}(\mathbf{PC}_{\mathrm{TB}}^{O} - \mathbf{PC}_{\mathrm{TB}}^{i})]}{\sqrt{(2\pi)^{\mathrm{npc}}|\mathbf{S}_i|}}, \tag{6}$$

where $\mathbf{PC}_{\mathrm{TB}}^{O}$ is the observed principal component vector and $\mathbf{PC}_{\mathrm{TB}}^{i}$ is the mean principal component vector corresponding to class C_i. Here, npc is the number of components considered in retrievals (five in this study) and $\mathbf{S}_i$ is the covariance matrix of the principal components corresponding to the profiles that constitute class C_i. Equations (4), (5), and (6) can be combined, yielding

$$\hat{E}(\mathbf{R}|\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET}) = \frac{1}{\Gamma}\sum_{i=1}^{\mathrm{nc}} \mathbf{R}_i \frac{\exp[-0.5(\mathbf{PC}_{\mathrm{TB}}^{O} - \mathbf{PC}_{\mathrm{TB}}^{i})^{\mathrm{T}}\mathbf{S}_i^{-1}(\mathbf{PC}_{\mathrm{TB}}^{O} - \mathbf{PC}_{\mathrm{TB}}^{i})]}{\sqrt{(2\pi)^{\mathrm{npc}}|\mathbf{S}_i|}} P(C_i), \tag{7}$$

where Γ is the normalization factor,

$$\Gamma = \sum_{i=1}^{\mathrm{nc}} \frac{\exp[-0.5(\mathbf{PC}_{\mathrm{TB}}^{O} - \mathbf{PC}_{\mathrm{TB}}^{i})^{\mathrm{T}}\mathbf{S}_i^{-1}(\mathbf{PC}_{\mathrm{TB}}^{O} - \mathbf{PC}_{\mathrm{TB}}^{i})]}{\sqrt{(2\pi)^{\mathrm{npc}}|\mathbf{S}_i|}} P(C_i), \tag{8}$$

and $P(C_i)$ is determined from the number of profiles within class C_i. Consequently, (7) and (8) can be used with the combined TRMM PR–TMI database to retrieve precipitation profiles from TMI-only observations. The error variance of the precipitation estimate is

$$\hat{V}(\mathbf{R}|\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET}) = \frac{1}{\Gamma^2}\sum_{i=1}^{\mathrm{nc}} V(\mathbf{R}_i)\left\{\frac{\exp[-0.5(\mathbf{PC}_{\mathrm{TB}}^{O} - \mathbf{PC}_{\mathrm{TB}}^{i})^{\mathrm{T}}\mathbf{S}_i^{-1}(\mathbf{PC}_{\mathrm{TB}}^{O} - \mathbf{PC}_{\mathrm{TB}}^{i})]}{\sqrt{(2\pi)^{\mathrm{npc}}|\mathbf{S}_i|}} P(C_i)\right\}^2, \tag{9}$$

where $\hat{V}(\mathbf{R}_i)$ is the variance of precipitation profiles in class C_i.

3. Results

There are various alternatives for constructing the combined TRMM PR–TMI for the TMI-only algorithm. For example, multiple databases can be constructed as a function of region and season and applied as a function of region and season. This alternative, although possible, is potentially problematic, however. As the region for which a database is constructed becomes smaller, the number of TRMM orbits needed to create a representative database increases, possibly requiring (given the seasonal dependence) a few years of data. Also, at the retrieval time the manipulation of many large databases could be cumbersome. Another possibility is to include additional geophysical parameters that would explain some of the regional and seasonal differences in profile subpopulations within a global database. The SST is such a parameter, and given that the database construction explicitly accounts for its variability, the database is applicable globally.

In this study, 1 month of TRMM PR–TMI data is used to construct the database and investigate the seasonal variation of the algorithm's performance. Although in principle an algorithm retrained every month would have merit, in application it is convenient to use an algorithm that does not require retraining. In constructing the combined TRMM PR–TMI database, we use all TRMM orbits from July 2000. The SST estimates employed to organize the database as previously explained are derived from TMI observations using the formulation of Wentz (1997). Although, in principle one can estimate the SSTs from TMI observations simultaneously with the precipitation and use short-period (on the order of days) SST averages in the precipitation retrieval, the climatological TMI-based SST estimates are used in this study. An obvious evaluation of the algorithm performance is the comparison of algorithm estimates with estimates from the combined TRMM PR–TMI method used in the database construction. Different variants of the TRMM PR–TMI-based TMI algorithm can be derived as a function of way in which the information in the database is used. As described earlier, for TMI retrievals one can use the observed brightness temperatures and their associated principal components in the database or the simulated brightness temperatures and their associated principal components. Moreover, because the ET estimate is used only to improve the algorithm efficiency, one can derive a solution by considering all of the profiles in the database irrespective of their ET estimates. Similarly, one can derive an algorithm that retrieves precipitation by exploring the whole database without discriminating the profiles as a function of SST. A concise description of the different alternative formulations considered in this paper is given in Table 1. Given in Table 1 is also the performance of the algorithm in terms of the correlation coefficient and the relative root-mean-square

TABLE 1. Performance of various TRMM PR–TMI-based TMI formulations for instantaneous retrievals at the database spatial resolution.

Formulation	Characteristics	Correlation coef	Rmse (%)
1	Observed TMI brightness temperatures in the database, ET and SST stratification	0.74	0.52
2	Simulated TMI brightness temperatures in the database, ET and SST stratification	0.68	0.57
3	Observed TMI brightness temperatures in the database, no ET stratification	0.74	0.52
4	Observed TMI brightness temperatures in the database, no SST stratification	0.69	0.56

error (rmse). The correlation coefficient is calculated between the instantaneous, high-resolution TMI retrievals and the TRMM PR–TMI-combined reference estimates. The relative rmse is calculated as the ratio of the root-mean-square difference between the TMI retrievals and TRMM PR–TMI-combined reference estimates to the standard deviation of the TRMM PR–TMI-combined reference estimates. One may note from Table 1 that the random differences between the observed brightness temperatures and the simulated brightness temperatures in the combined retrievals have a small negative impact on the retrievals. The third formulation exhibits the same performance as the first, which indicates that neither improvement nor deterioration in the performance is obtained through the ET stratification. However, the ET estimates are an important ingredient in the algorithm because they significantly reduce the computational effort. The fourth formulation shows performance deterioration relative to the first, but the SST does not seem to have a large impact on the retrievals. This is most likely because other kinds of uncertainties dominate, and also because a large proportion of precipitation originates in regions with relatively constant (and high) SSTs. When the retrievals are aggregated monthly within $0.5° \times 0.5°$ grid boxes, the differences among the four formulations become indistinguishable. In the remainder of the study, only retrievals based on the first formulation are analyzed. Also, monthly rather than instantaneous retrievals are considered.

The precipitation monthly products are determined by averaging the instantaneous footprint-scale estimates within $0.5° \times 0.5°$ boxes. An additional criterion for the TMI monthly estimates is that they are based exclusively on TMI observations within the narrower TRMM PR swath. Because the TRMM PR and the TMI have different swaths, and consequently, different sampling, including TMI estimates from outside the TRMM PR swath could result in additional differences between the TMI and TRMM PR-based monthly estimates.

In Fig. 3, the zonal mean rainfall estimates for the same month that was used for database construction (July 2000) from the TRMM PR–TMI-based TMI algorithm and the combined TRMM PR–TMI method are shown. The observed TMI brightness temperatures stored in the database are used in the TRMM PR–TMI-based TMI algorithm for this particular retrieval, and unless explicitly specified otherwise, the observed brightness temperatures in the database are used in the TMI-only retrievals. The estimates from the two algorithms agree well. For reference, estimates from the version-6 TRMM facility TMI algorithm (GPROF) are also shown. The GPROF estimates also agree well with both the TRMM PR–TMI-based TMI and the combined TRMM PR–TMI estimates, however, some differences are apparent in the intertropical convergence zone (ITCZ). Although not shown in Fig. 3, the version-6 TRMM facility TRMM PR algorithm yields estimates that are very similar to the combined TRMM PR–TMI at the monthly level. The disagreement between the version-5 TRMM facility TMI and TRMM PR algorithm estimates was evident even at the monthly level, and systematic differences as large as 23% existed over oceans (Kummerow et al. 2001). Surface rain-rate estimates from the version-6 algorithms are in much better agreement (as seen in Fig. 3), but notable differences between the TMI and TRMM PR

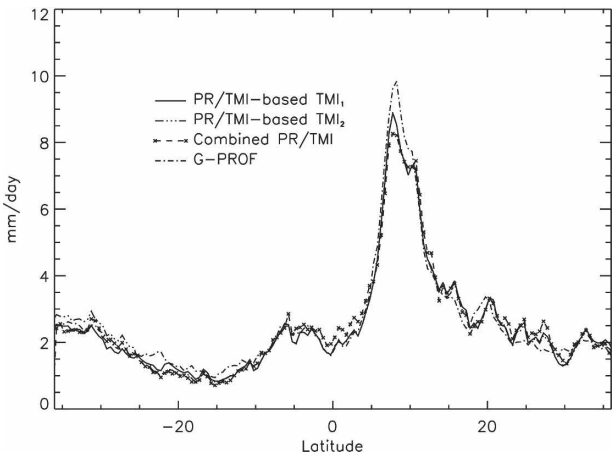

FIG. 3. Zonal over-ocean surface rainfall estimates for July 2000 from the TRMM PR–TMI-based TMI algorithm, GPROF, and the combined TRMM PR–TMI algorithm.

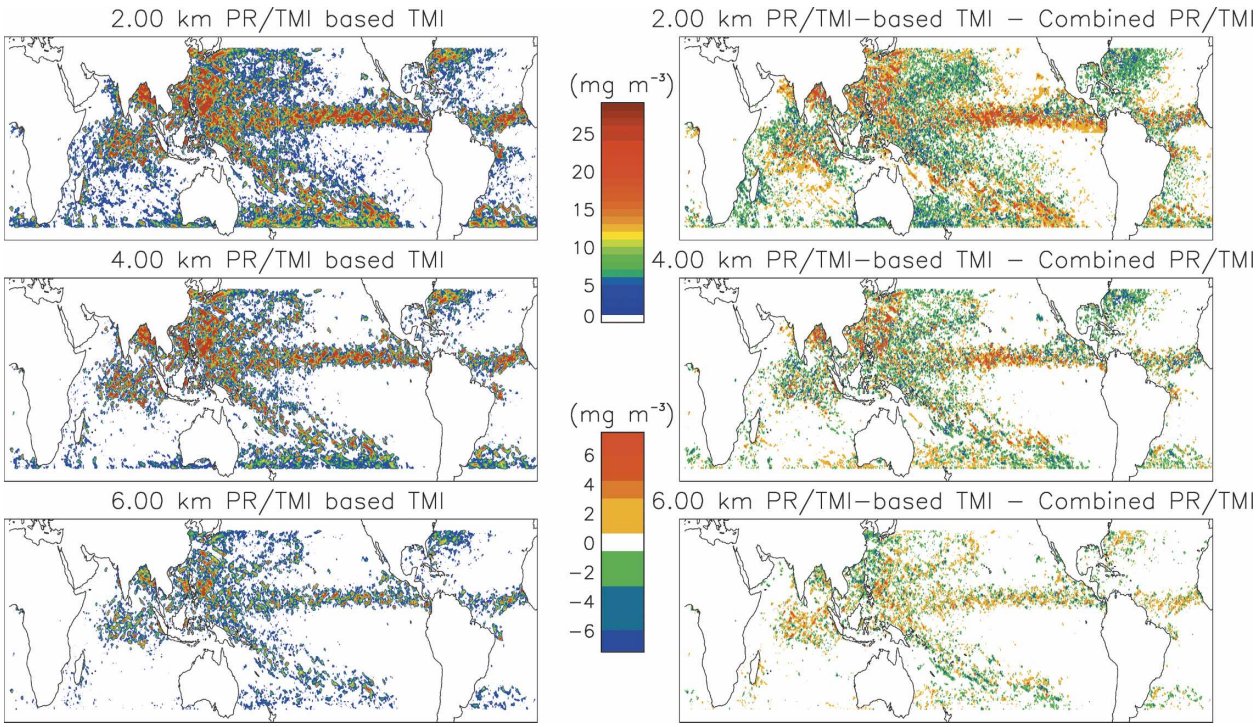

FIG. 4. Monthly 0.5° × 0.5° estimates of precipitation water contents from the TRMM PR–TMI-based TMI algorithm at three different altitudes: (left) the actual estimates and (right) the differences between the TRMM PR–TMI-based TMI and the combined TRMM PR–TMI estimates.

estimates still exist. As discussed below, more serious discrepancies between the algorithms occur at levels above the surface.

Presented on the left-hand side of Fig. 4 are global maps of TRMM PR–TMI-based TMI 0.5° × 0.5° monthly estimates of precipitation water content at three different altitudes, while on the right-hand side are differences between the TRMM PR–TMI-based TMI and combined TRMM PR–TMI 0.5° × 0.5° monthly estimates. It is apparent from Fig. 4 that the TRMM PR–TMI-based TMI algorithm performance is not uniform. In some regions, for example, the central-eastern part of the ITCZ, the algorithm tends to overestimate, while in other regions [e.g., tropical western Pacific (TWP)] the algorithm generally underestimates precipitation. These biases are caused by differences between eastern and western Pacific rain systems (Berg et al. 2002) and are an indication that the SST and ET estimators along with the brightness temperature principal components at the pixel level are insufficient to detect subtle differences in the rain systems. Both the magnitudes of precipitation estimates and the differences between the two algorithms decrease with altitude.

In Fig. 5, the GPROF monthly estimates and their differences relative to the combined TRMM PR–TMI estimates are shown. The 2-km estimates are similar to the combined TRMM PR–TMI-based estimates, although differences are somewhat larger relative to the TRMM PR–TMI-based TMI estimates near the southern boundary of the observing domain. On the other hand, at 4 and 6 km the differences between the GPROF and the TRMM PR–TMI-based estimates increase significantly. At 4 km the largest differences are preponderantly located in the South Pacific, along the edge of the midlatitude storm track in the Southern Hemisphere. This region is characterized by freezing levels generally lower than 3 km, and, most likely, the systematic differences between precipitation distributions above the freezing level in CRM simulations and those from the combined TRMM PR–TMI retrievals are responsible for the differences between the GPROF and the combined TRMM PR–TMI estimates at the monthly level. At 6 km, the differences between the GPROF and combined estimates are even greater than at 4 km and are evident almost everywhere where significant precipitation exists. Because of errors in describing the vertical transition from liquid to ice phase precipitation, and uncertainties in quantifying the electromagnetic properties of the ice phase, the combined TRMM PR–TMI estimates of mixed and ice phase precipitation have greater uncertainties than liquid phase

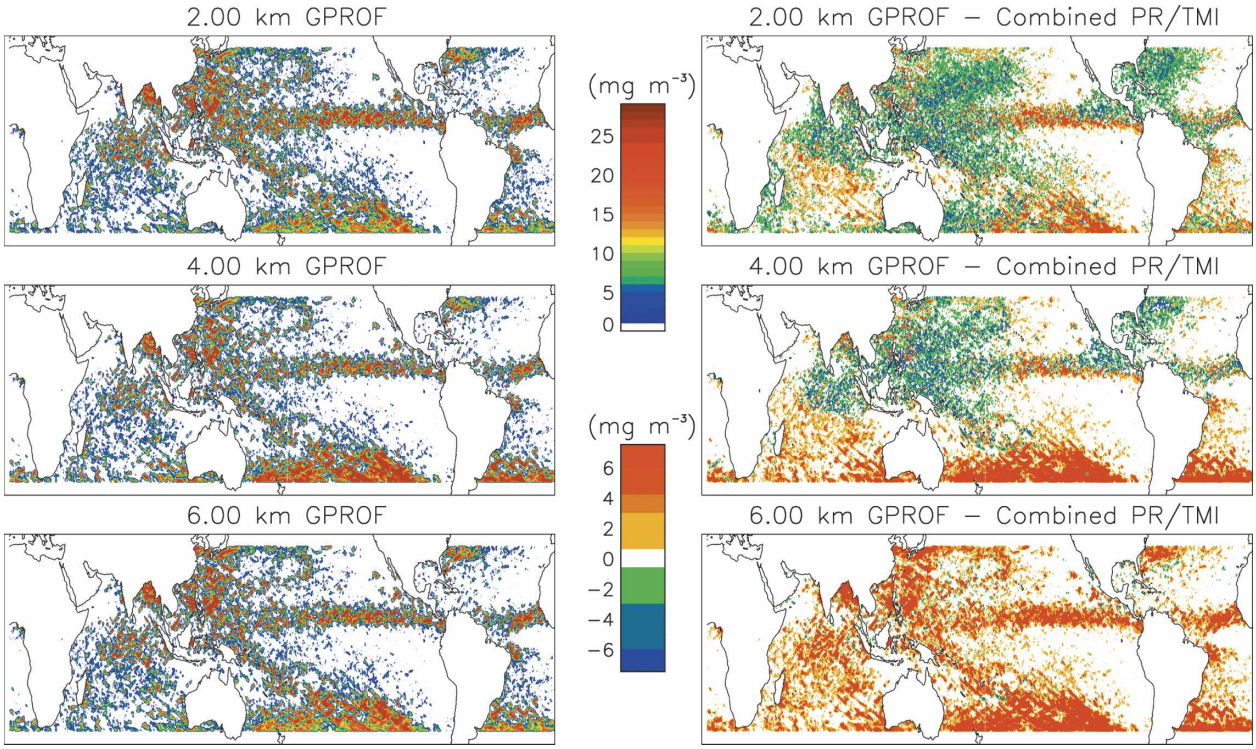

FIG. 5. Same as Fig. 4, but for GPROF instead of the TRMM PR–TMI-based TMI algorithm.

precipitation estimates do. However, despite these uncertainties, the combined TRMM PR–TMI estimates are more reliable than precipitation distributions based on CRM simulations.

To investigate the nature of the differences between GPROF and the combined TRMM PR–TMI estimates, the following experiment is performed. Here, GPROF is extended to retrieve not only precipitation profiles but also reflectivity profiles. This is achieved by calculating the reflectivity from the precipitation water contents in GPROF's supporting database. Then the reflectivity profiles are retrieved from TMI observations using GPROF, and the retrieved reflectivity distributions at 6 km are compared with the reflectivity distributions from the TRMM PR observations. The 4-km level is not considered in this comparison because many observations at that level are within the radar bright band, and, given the uncertainties in modeling the bright band from CRM hydrometeor profiles, the differences are difficult to interpret. Thus, only reflectivity observations greater than 17 dBZ (the TRMM PR detectability threshold) at the 6-km level are considered in the comparison. The result of this comparison is that over the month of July 2000 the average retrieved reflectivity is about 6–7 dB greater than average observed reflectivity, which suggests that the CRM profiles in the GPROF database contain generally larger amounts of ice phase precipitation than in nature. This is consistent with the findings of Bauer (2001), who hypothesized that CRMs produce too much ice phase precipitation, resulting in simulated high-frequency brightness temperatures generally lower than those observed by TRMM.

Figure 6 contains scatterplots of TRMM PR–TMI-based TMI (and GPROF) 0.5° × 0.5° monthly estimates versus combined TRMM PR–TMI estimates at the same space and time resolution. The scatter of estimates in this figure is consistent with the difference maps shown in Figs. 4 and 5. That is, there is good agreement between the TRMM PR–TMI-based TMI and the combined TRMM PR–TMI estimates, while the agreement between GPROF and the combined TRMM PR–TMI estimates is good at 2 km, deteriorates at 4 km, and is poor at 6 km. It is important to note, though, that at 6 km the correlation between GPROF and the combined estimates is still very high (about 0.9), but the multiplicative bias is extremely large (the GPROF estimates are larger by a factor of about 5). As already mentioned, this startling difference is likely caused by excessive amounts of ice-phase hydrometers in the CRM simulations supporting GPROF.

Previous studies have indicated that GPROF performance varies regionally and seasonally. For example,

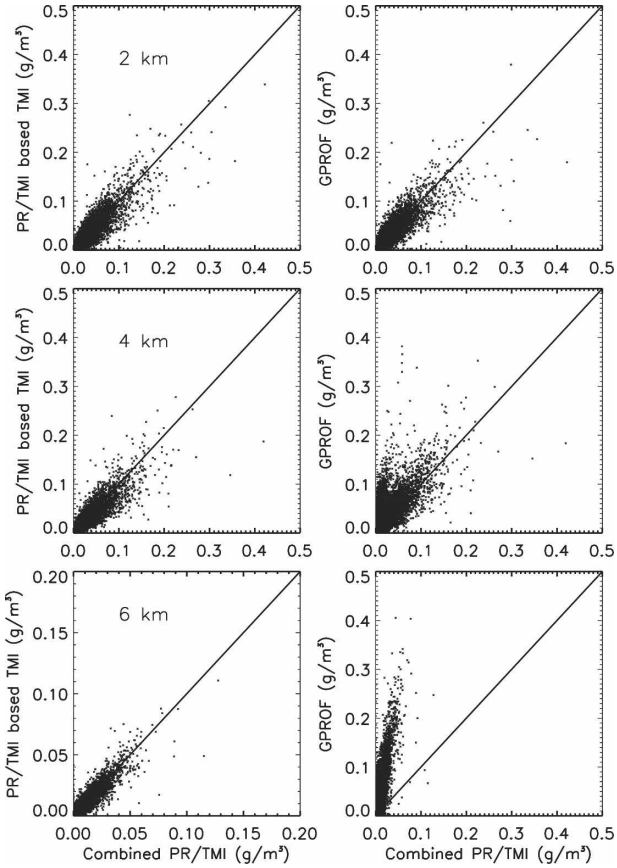

FIG. 6. Scatterplots of monthly 0.5° × 0.5° TRMM PR–TMI-based (left) TMI precipitation water content estimates and (right) GPROF estimates against combined TRMM PR–TMI estimates for (top) 2, (middle) 4, and (bottom) 6 km.

Ikai and Nakamura (2003) compared the version-5 TRMM facility TMI (GPROF 5) and PR algorithms and found regional and seasonal dependence in the agreement. The two estimates were usually highly correlated, but systematic differences were found. The authors attributed the differences to errors in the GPROF estimation of the freezing level and inadequate attenuation–reflectivity, and reflectivity–precipitation relationships. Although the factors pointed out by Ikai and Nakamura (2003) indeed affect the agreement between TMI and TRMM PR retrievals of precipitation, a factor that can more seriously impact the agreement between TMI and TRMM PR precipitation estimates is the representativeness of the database used in the TMI algorithm. Even if perfect estimation of the freezing level and perfect TRMM PR retrievals of precipitation are achieved, the TMI-only estimates would still disagree with the TRMM PR estimates if the database used in the TMI-only retrievals does not represent the statistical distribution of the precipitation profiles derived from the TRMM PR. To investigate the differences between TMI and TRMM PR estimates, Ikai and Nakamura (2003) selected six geographic domains, as defined in Table 2. We use the same domains to compare the TRMM PR–TMI-based TMI and the combined TRMM PR–TMI monthly 0.5° × 0.5° estimates at 2.0-km altitude. Scatterplots based on these geographic domains are presented in Fig. 7.

It may be noted from Fig. 7 that estimates from the TRMM PR–TMI-based TMI and the combined TRMM PR–TMI generally do not exhibit large systematic differences. The largest difference, about 20% relative bias, occurs for domain tropical central Pacific (TCP). All of the other domains exhibit a slight negative bias of less than 5%. Ikai and Nakamura (2003) found larger differences (up to 110% in terms of surface rainfall rate in TCP), which reinforces our hypothesis that the database used in the version-5 TRMM facility TMI algorithm is partly responsible for the differences between the version-5 TMI and TRMM PR estimates and that a database consistent with TRMM PR observations promotes agreement between TMI-only retrievals and TRMM PR-based precipitation estimates. A similar analysis, that is, the comparison of the monthly 0.5° × 0.5° combined TRMM PR–TMI and GPROF estimates, yields comparable results. The bias is slightly higher, that is, 25%, than that corresponding to the TRMM PR–TMI-based TMI algorithm in TCP, and there is a negative bias of about −10% in TWP. In all of the other regions, the absolute bias is below 5%. This analysis indicates significantly better agreement of GPROF version-6 estimates with TRMM PR-based estimates at the surface. However, above the freezing levels large discrepancies still exist as apparent in the previous results.

Additional issues of practical importance are whether or not the TRMM PR–TMI-based TMI algorithm

TABLE 2. Domains selected to investigate the regional variation of the performance of the TRMM PR–TMI-based TMI algorithm. These domains are the same as those selected by Ikai and Nakamura (2003).

Regions	Lat	Lon
Midlatitude North Atlantic (MNA)	30°–37°N	45°–75°W
Midlatitude western North Pacific (MWNP)	30°–3°7N	140°–170°E
Midlatitude central North Pacific (MCNP)	30°–37°N	150°W–180°
Midlatitude western South Pacific (MWSP)	30°–37°S	150°W–180°
Tropical western Pacific (TWP)	0°–10°N	130°–160°E
Tropical central Pacific (TCP)	0°–10°N	120°–150°W

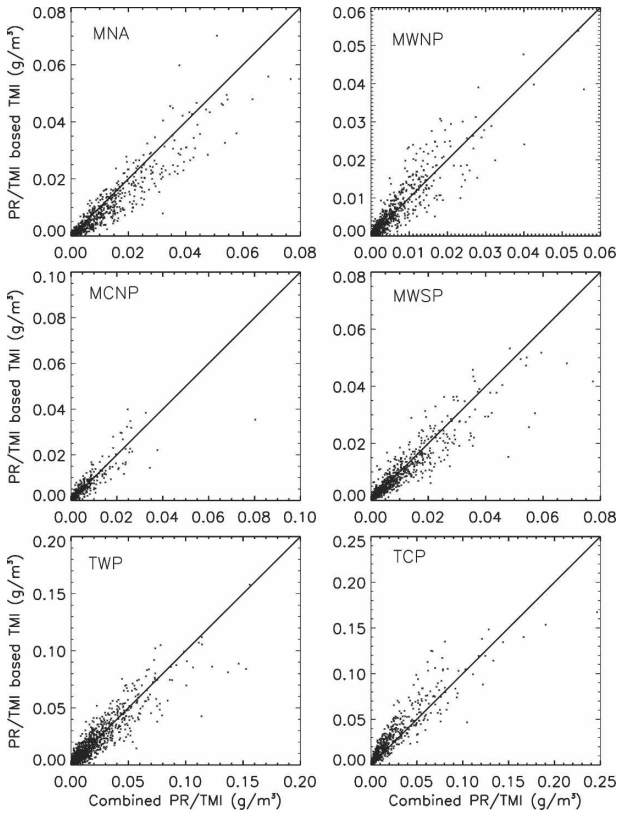

FIG. 7. Scatterplots of monthly 0.5° × 0.5° TRMM PR–TMI-based TMI precipitation water contents estimates against combined TRMM PR–TMI estimates for the six different regions defined in Table 2.

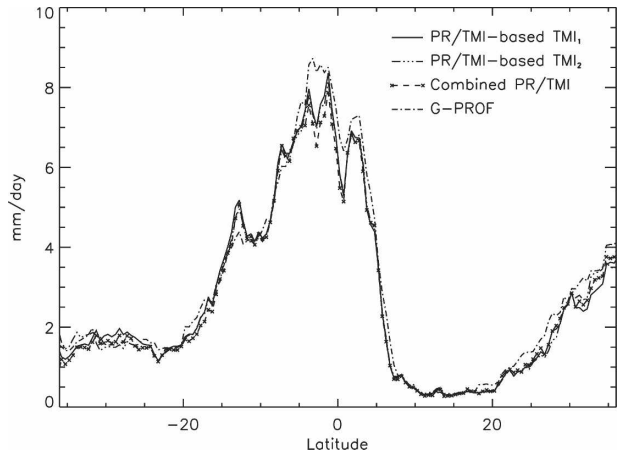

FIG. 8. Zonal over-ocean surface rainfall estimates for January 1998 from the TRMM PR–TMI-based TMI algorithm, GPROF, and the combined TRMM PR–TMI algorithm. Two different supporting databases, that is, derived using observations from July 2000, and January 1998, respectively, are employed in the TRMM PR–TMI-based TMI algorithm.

derived using just 1 month of data is general, and whether its performance deteriorates when applied to TMI observations from a different season. To investigate these issues two databases are constructed. The first is the database used in the previous examples, that is, the database constructed using July 2000 combined TRMM PR–TMI retrievals, and the second is constructed using January 1998 combined TRMM PR–TMI retrievals. The Bayesian algorithm is applied to TMI data from January 1998 using the two databases in turn. The zonal mean rainfall estimates derived from the two databases are shown in Fig. 8. For reference, the version-6 TRMM facility TMI algorithm (GPROF) estimates and the combined TRMM PR–TMI estimates are also shown. One may note that there is essentially no dependence of the zonal means on the database used in the TRMM PR–TMI-based TMI algorithm. Neither do monthly 0.5° × 0.5° rain estimate maps (not shown) show significant sensitivity to the database utilized. Moreover, estimates from the TRMM PR–TMI-based TMI algorithm agree well with the combined TRMM PR–TMI estimates, while GPROF estimates are somewhat higher in the ITCZ. Version-6 TRMM facility PR algorithm estimates (not shown) are somewhat lower than the TRMM PR–TMI-based TMI estimates in the ITCZ. Although this investigation does not validate either GPROF or the TRMM PR-only algorithms, it does show that if unbiased estimates are derived from a month of combined TMI and TRMM PR observations, these estimates can be employed to obtain largely unbiased estimates from TMI-only observations for other months. At the instantaneous and database resolution level, the TRMM PR–TMI-based TMI algorithm's performance for January 1998 is similar to that reported in Table 1.

4. Latent heating application

The Bayesian algorithm formulated in section 2 can be readily extended to estimate LH (i.e., the heating associated with the phase transformation of water vapor, cloud, and precipitation particles) from TMI-only observations. Knowledge of the spatial and temporal distribution of heating in the Tropics is required for various applications (see Simpson et al. 1988). These applications range from long-term climate prediction to storm-scale diagnostics and forecasting. Further discussion may be found in Olson et al. (1999). The large spectrum of applications justifies the development of algorithms for latent heating estimation from observations provided by spaceborne instruments. Examples of such algorithms are those derived by Olson et al. (1999) for passive microwave observations and by Tao et al. (1993) and Yang and Smith (1999) for both passive and

active microwave observations. The Olson et al. (1999) algorithm is based on a Bayesian procedure that assigns, for a given set of observations, probabilities to a set of latent heating profiles derived from CRM simulations. Tao et al. (1993) use only the surface precipitation, a convective–stratiform classification, and a table of normalized latent heating profiles (derived also from CRM simulations) to reconstruct vertical heating profiles, while Yang and Smith (1999) assume that the observed precipitation distribution is steady state and use hydrometeor conservation equations to estimate the latent heating. Given the lack of information regarding the dynamical context of the observed precipitation, only approximate algorithms may be derived, and various approaches are possible.

In the present study, a procedure similar to that of Shige et al. (2004) is used to assign a heating profile physically consistent with each precipitation profile derived from the combined TRMM PR–TMI algorithm. First, cloud-resolving model simulations are used to create lookup tables of the mean latent heating vertical structure and surface precipitation rate. The model simulations are produced using the Goddard Cumulus Ensemble (GCE) model over the South China Sea (Tao et al. 2003). The model heating vertical profiles and surface precipitation rates are sorted by convective–stratiform classification and the radar echo top (17 dBZ is the minimum detectable signal of the TRMM PR). The stratiform profiles are further sorted by the ratio of the difference between the precipitation rate at the bottom of the melting layer and the surface precipitation rate to the precipitation rate at the bottom of the melting layer. The mean model heating profile and surface rain rate are calculated and tabulated as functions of convective–stratiform class, radar echo top, and stratiform precipitation ratio (when applicable).

Given the convective–stratiform classification, echo-top height, and stratiform precipitation ratio of a combined TRMM PR/TMI profile estimate in the database, the corresponding tabulated heating profile is extracted and rescaled by the combined surface precipitation rate estimate (assuming that net heating in the profile is approximately equal to the surface rain rate times the latent heat of condensation L_v, at least in the space–time mean). The reconstructed heating profile, indexed by its associated brightness temperature principal components SST and ET is entered into the algorithm-supporting database in essentially the same manner as the given combined TRMM PR–TMI-retrieved precipitation profile.

It should be noted that the method of Shige et al. (2004), as well as other methods for estimating heating from precipitation data, yields latent heating estimates with considerable random error. This error is suppressed by profile clustering in the database and probabilistic averaging using a formula similar to (4),

$$\hat{E}(\mathbf{LH}|\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET}) = \sum_{i=1}^{\mathrm{nc}} \mathbf{LH}_i \cdot P(C_i|\mathbf{PC}_{\mathrm{TB}}, \mathrm{SST}, \mathrm{ET}), \tag{10}$$

where $\mathbf{LH}_i$ is the average latent heating profile associated with the profiles in class C_i based on the methodology described in section 2a. Equation (10) can be extended to estimate $Q_1 - Q_R$, where Q_R is the heating/cooling rate associated with radiative processes) if the eddy heat flux convergence term is added to the LH (see Yanai et al. 1973). The eddy heat flux term can be readily calculated in the CRM simulations so that each precipitation and latent heating profile in the aforementioned lookup table can be associated with an eddy flux term. Therefore, the $Q_1 - Q_R$ profile (hereinafter called $Q_{1\mathrm{R}}$) can be estimated in a manner similar to the LH profile.

The accuracy of TRMM PR–TMI-based TMI latent heating estimates can be evaluated by comparison with available rawinsonde-based analyses of diabatic heating. Rawinsondes are vertical profiling instruments that measure the pressure, temperature, humidity, and wind at a given geographical location. By analyzing the data provided by a network of rawinsondes, one may estimate the large-scale average vertical distribution of heating sources and sinks within the convex region circumscribed by the network. Rawinsonde networks have been deployed in various field experiments to gain an understanding of large-scale heating and moistening processes. One such experiment is the South China Sea Monsoon Experiment (SCSMEX) for which heating profile estimates have been derived over the Northern Enhanced Sounding Array (NESA) by Johnson and Ciesielski (2002). The rawinsonde analysis provides estimates of the apparent heat source Q_1 (Yanai et al. 1973), while the TMI-only algorithm provides only $Q_{1\mathrm{R}}$. By adding a domain- and time-invariant Q_R component to the TMI-based $Q_{1\mathrm{R}}$ estimates, one can approximate Q_1 from the TMI estimates. Here, the climatological radiative heating profiles of Dopplick (1979) provide Q_R.

As mentioned earlier, the procedure used to assign $Q_{1\mathrm{R}}$ profiles to the combined TRMM PR–TMI precipitation profiles in the TMI algorithm's supporting database represents by itself a complete, stand-alone heating estimation algorithm from TRMM PR-only or combined TRMM PR–TMI observations. The estimates from this algorithm are potentially more accurate than the TMI-only algorithm at the instantaneous level be-

cause the precipitation profiles are better determined, but they are, most likely, less accurate in terms of large-scale daily averages because of the less frequent temporal sampling by the TRMM PR. For reference, Q_1 estimates from the TRMM PR–TMI observations are included in the comparison of the TMI and rawinsonde estimates. Estimates of Q_1 from the TMI and combined TRMM PR–TMI algorithms are averaged over the NESA region to match the rawinsonde estimates. In addition to estimates of Q_1, the rawinsonde analyses provide estimates of the average surface precipitation rate over the analysis domain. These estimates are derived from a budget equation relating the vertical integral of the apparent moisture sink Q_2, which can be readily determined from rawinsonde analyses, to the surface precipitation and evaporation. Evaporation rate estimates from the Japan Meteorological Agency (JMA) reanalysis, adjusted by shipboard flux measurements, were used by Johnson and Ciesielski (2002) to estimate the surface precipitation from Q_2, and their estimates of surface precipitation rate appear here for the purpose of intercomparison.

Time series of surface precipitation and Q_1 from TMI and the combined TRMM PR–TMI algorithms, as well as those from the rawinsonde analyses, are shown in Fig. 9. It may be noted from the figure that the agreement among the three estimates is fairly good. The rawinsonde surface precipitation estimates are better correlated with the TMI surface precipitation estimates (correlation coefficient 0.87) than with the TRMM PR surface precipitation estimates (correlation 0.78). Although all of the estimates are affected by sampling errors, the TRMM PR surface precipitation estimates are probably subject to the largest sampling errors in spite of being the most accurate at the footprint instantaneous level, leading to a discrepancy with the other estimates. Sampling errors affect not only the precipitation estimates but also heating estimates, and therefore it is expected that for periods when sampling errors in the precipitation estimates are large, heating estimates will also be subject to large sampling errors. Figure 9 confirms this hypothesis. It may be noted that whenever the three precipitation estimates agree, for example, 6 June, the heating estimates agree as well, while the largest differences in heating occur when the largest differences in precipitation rate occur. The ability of the TRMM PR–TMI-based TMI algorithm to estimate the echo-top altitude appears to be effective in capturing the variation in the depth of the layer in which heating is confined. The heating appears to extend closer to the ground in the rawinsonde estimates than in the satellite estimates, which indicates a deficiency in the CRM simulations or in the procedure that associates heating profiles with stratiform precipitation profiles. It was demonstrated by Li et al. (2005, manuscript submitted to *J. Atmos. Sci.*) that evaporation may be excessive in GCE model simulations utilizing the three ice bulk microphysics scheme, which may explain the more intense low-level cooling in the TMI-only heating estimates.

Time-averaged profiles of Q_1 estimates from the combined TRMM PR–TMI and TRMM PR–TMI-based TMI algorithms, and rawinsondes are shown in Fig. 10. Also, shown in Fig. 10 is the Q_R profile used in this study. One may note in the figure magnitude differences among the estimates. These differences are caused by differences in the average rain from the two algorithms and rawinsondes. The TRMM PR–TMI-based TMI rain estimates are larger than the rawinsonde rain estimates by about 25% and are larger than the combined TRMM PR–TMI rain estimates by about 70% (it should be mentioned that the large differences between the TRMM PR–TMI-based TMI and combined TRMM PR–TMI estimates are mainly caused by sampling). The Q_1 profiles in Fig. 10 are consistent with the rain estimates. Other differences consist of the above-mentioned pronounced cooling in the combined TRMM PR–TMI estimates and a somewhat lower peak relative to the rawinsonde estimates in the TRMM PR–TMI-based TMI Q_1. The lower peak in the TRMM PR–TMI-based TMI Q_1 may be a consequence of inconsistencies between the CRM convective–stratiform classification and the combined TRMM PR–TMI classification. This is because the TRMM PR–TMI-based TMI latent heating profiles are probabilistic averages of CRM latent heating stratiform and convective profiles. Assuming that the TRMM PR–TMI convective–stratiform classification is realistic, a lower peak in the Q_1 indicates a slight overestimation of the convective CRM Q_1 profile, which may occur if weak convective profiles are classified as stratiform profiles. Lang et al. (2003) present a detailed analysis on the uncertainty in latent heating estimation because of convective–stratiform classification.

An additional intercomparison (not shown in the paper) reveals that the TRMM PR–TMI-based TMI estimates of Q_1 are similar to those from the technique of Olson et al. (1999) for the SCSMEX NESA region. Nonetheless, important differences between the TRMM PR–TMI-based TMI estimates of Q_1 and those from the technique of Olson et al. (1999) are expected in general. This is because the database upon which the GPROF and Olson et al. (1999) technique are based favors the retrieval of deep structures, even in regions where the TRMM PR indicates shallow structures, for example, the midlatitude storm track of the South Pa-

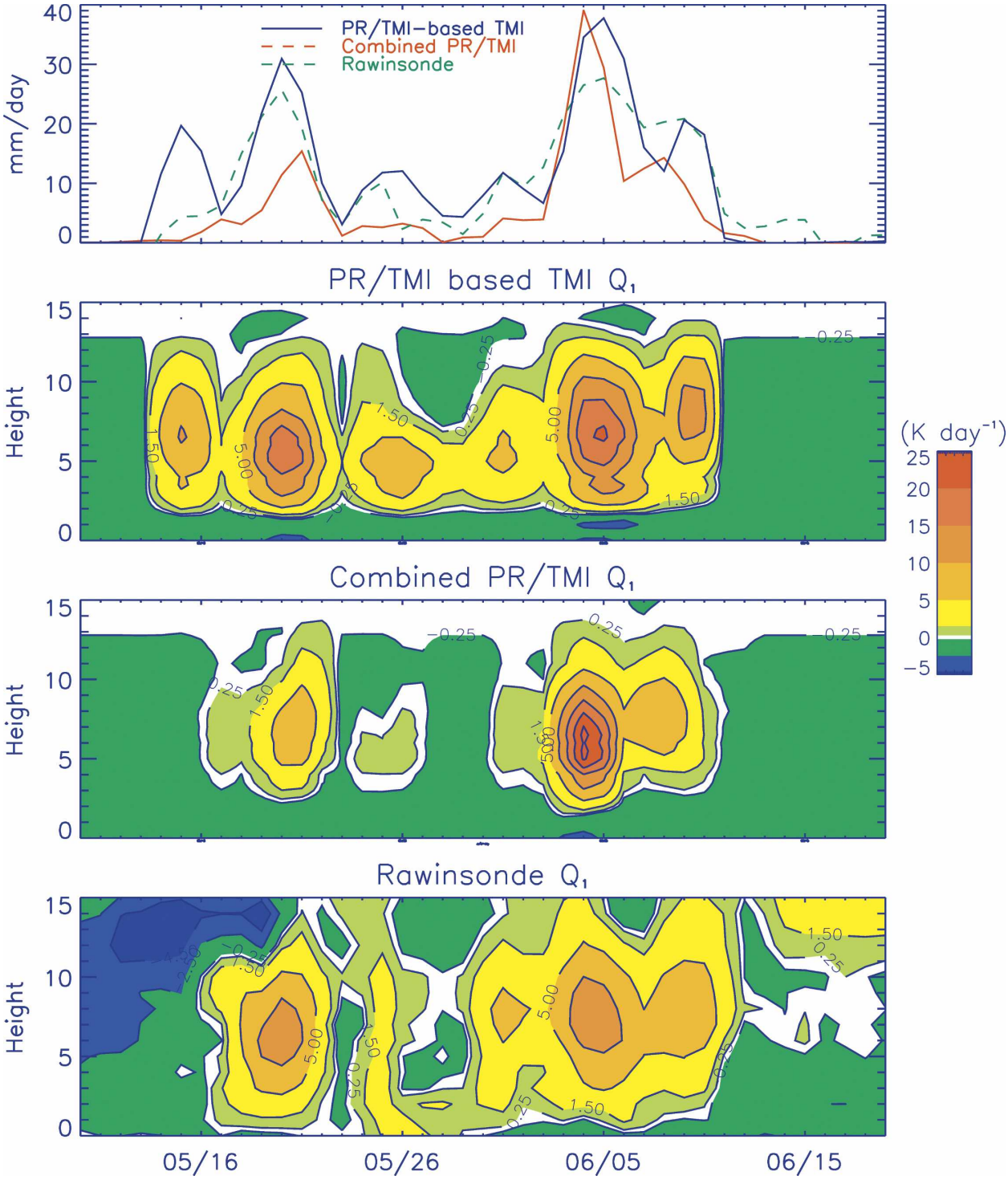

FIG. 9. Time series of surface precipitation rate and apparent heat source Q_1 for the SCSMEX NESA region. Shown from top to bottom are precipitation estimates from different methods, TRMM PR–TMI-based TMI estimates of Q_1, combined TRMM PR–TMI estimates of Q_1, and rawinsonde estimates of Q_1.

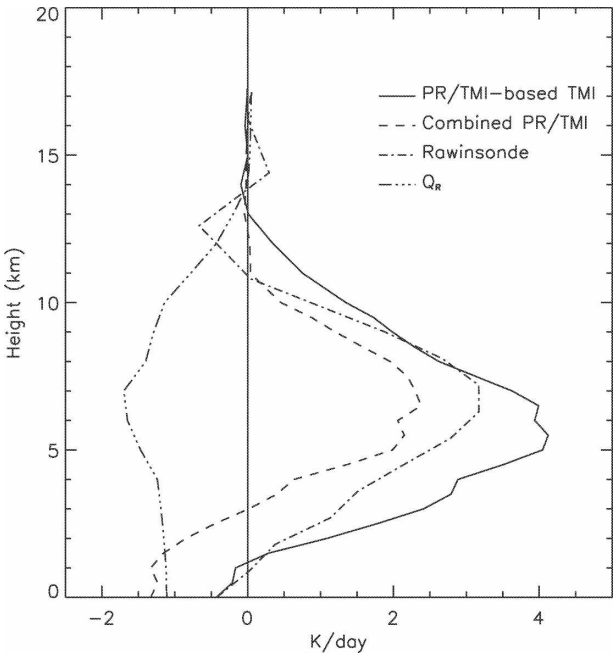

FIG. 10. Time-averaged profiles of Q_1 from the TRMM PR–TMI-based TMI and the combined TRMM PR–TMI algorithms, and from rawinsondes for the SCSMEX NESA region.

cific. However, the distribution of LH is a strong function of height (Shige et al. 2004), and consequently, the differences resulting from the database constituency between the GPROF and the TRMM PR–TMI-based TMI algorithm will manifest themselves not only in terms of precipitation but also in terms of LH.

Ultimately, the procedure developed in this section can be used to estimate latent heating globally. However, realistic estimates will require expanded heating lookup tables based upon a large variety of CRM simulations. The current heating tables do not realistically represent the distribution of low-level, shallow precipitating clouds. Consequently, in this initial test the TMI heating estimation algorithm is applied only to regions for which we have representative CRM simulations, such as SCSMEX, and not globally.

5. Conclusions

In this study, an algorithm for estimating precipitation from passive microwave-only observations is derived using a large record of precipitation retrievals from combined active and passive microwave observations. The combined active and passive retrievals are obtained from TRMM PR and TMI observations over the period of 1 month. The combined active–passive algorithm is physically based and may be used to derive a passive microwave-only precipitation estimation method for any combination of microwave frequencies and resolutions. A retrieval algorithm applicable to only TMI observations is investigated in this paper, however.

The combined TRMM PR–TMI-estimated vertical precipitation profiles and associated brightness temperatures are stored in a database. In the database, the profiles are sorted by the local SST and a TMI estimate of the ET. The profiles are further clustered based on the similarity of the associated brightness temperatures using principal components. The database organization facilitates the use of a computationally efficient Bayesian procedure to estimate the precipitation profile from TMI-only observations.

The TRMM PR–TMI-based TMI-only algorithm yields precipitation estimates similar to those from the combined TRMM PR–TMI algorithm. Although some regional biases exist even in monthly $0.5° \times 0.5°$, the algorithm appears to be effective in capturing the regional differences in vertical precipitation structure. In contrast, the TRMM facility TMI algorithm (GPROF) generally estimates significantly larger amounts of ice phase precipitation above the freezing level. Even though the combined TRMM PR–TMI algorithm might be biased above the freezing level because of uncertainties in the modeling of electromagnetic properties of the ice-phase precipitation particles, the TMI-only algorithm derived from it is more stable than GPROF because the uncertainties in modeling the properties of the ice phase are easier to understand, quantify, and mitigate than the potential systematic errors in the CRM simulations that support GPROF. However, the database constructed directly from retrievals is meant to be only a practical alternative for databases constructed from CRMs, and the effort to improve the realism, diversity, and overall number of CRM simulations for database construction should continue. This would not only facilitate a better understanding of the retrieved precipitation structures, but it would also allow for the estimation of variables (such as the vertical heating profile) that are related to precipitation but are not directly observed. In addition, it is anticipated that CRMs will form the basis of improved superparameterizations in general circulation models, and therefore precipitation simulations must be improved if assimilation of microwave data in these models is to be viable.

Results of this study suggest little sensitivity of the TRMM PR–TMI-based TMI-only estimates to the specific data used in the database creation, assuming that at least one month of TRMM PR–TMI observations is utilized. Here, combined TRMM PR–TMI observations from July 2000 are used to create the database, and the retrieval algorithm is applied to TMI observa-

tions from January 1998. The monthly 0.5° × 0.5° estimates from the TRMM PR–TMI-based TMI-only algorithm, and estimates from the same algorithm but with a database created using January 1998 combined TRMM PR–TMI retrievals, are almost indistinguishable. This suggests that, given the SST and ET stratification of the database, 1 month of combined TRMM PR–TMI retrievals might be sufficient to provide precipitation–brightness temperatures relationships that do not exhibit significant seasonal and regional variations.

Using CRM simulations, the database of retrieved profiles, precipitation profiles, and associated brightness temperatures is augmented with physically consistent latent heating profiles. That is, a latent heating profile is associated with each precipitation profile in the database, making it possible to estimate latent heating with the TRMM PR–TMI-based TMI-only algorithm. The TMI latent heating algorithm is applied to TRMM observations made within the SCSMEX NESA region during the field campaign's intensive observing period. An intercomparison reveals fairly good agreement between the TMI and rawinsonde network estimates of Q_1. The heating estimation algorithm can be applied to other regions as well, provided that the precipitation database is augmented with heating profiles from realistic CRM simulations representative of those regions.

In conclusion, combined TRMM PR–TMI precipitation retrievals provide valuable information that can be used to support passive microwave precipitation estimation algorithms. Such algorithms may be used to overcome deficiencies in current precipitation algorithms that are supported by cloud-resolving model simulations. At the same time, cloud-resolving model simulations may still be used to augment the information inferred from TRMM PR–TMI precipitation retrievals and enhance the algorithms they support.

Acknowledgments. We thank Drs. W.-K. Tao and Chung-Lin Shie of NASA Goddard Space Flight Center, who provided the GCE model simulation data for this study. We are also grateful to Professor Richard H. Johnson and Dr. Paul Ciesielski of Colorado State University for providing the rawinsonde estimates of Q_1 and surface precipitation. This research was funded by the NASA Precipitation Measurement Missions project.

REFERENCES

Bauer, P., 2001: Over-ocean rainfall retrieval from multisensor data of the Tropical Rainfall Measuring Mission. Part I: Design and evaluation of inversion databases. *J. Atmos. Oceanic Technol.,* **18,** 1315–1330.

Berg, W., C. Kummerow, and C. A. Morales, 2002: Differences between east and west Pacific rainfall systems. *J. Climate,* **15,** 3659–3672.

Dopplick, T. G., 1979: Radiative heating of the global atmosphere—Corrigendum. *J. Atmos. Sci.,* **36,** 1812–1817.

Duda, R. O., P. E. Hart, and D. G. Stork, 2000: *Pattern Classification*. Wiley Interscience, 680 pp.

Evans, F., J. Turk, T. Wong, and G. L. Stephens, 1995: A Bayesian approach to microwave precipitation profile retrieval. *J. Appl. Meteor.,* **34,** 260–279.

Ferraro, R. F., E. A. Smith, W. Berg, and G. J. Huffman, 1998: A screening methodology for passive microwave precipitation retrieval algorithms. *J. Atmos. Sci.,* **55,** 1583–1600.

Grecu, M., W. S. Olson, and E. N. Anagnostou, 2004: Retrieval of precipitation profiles from multiresolution, multifrequency active and passive microwave observations. *J. Appl. Meteor.,* **43,** 562–575.

Ikai, J., and K. Nakamura, 2003: Comparison of rain rates over the ocean derived from TRMM Microwave Imager and precipitation radar. *J. Atmos. Oceanic Technol.,* **20,** 1709–1726.

Jain, A. K., M. N. Murthy, and P. J. Flynn, 1999: Data clustering: A review. *ACM Comput. Surv.,* **31,** 264–323.

Johnson, R. H., and P. E. Ciesielski, 2002: Characteristics of the 1998 summer monsoon onset over the northern South China Sea. *J. Meteor. Soc. Japan,* **80,** 561–578.

Kummerow, C., and Coauthors, 2001: The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteor.,* **40,** 1801–1820.

Lang, S., W.-K. Tao, J. Simpson, and B. Ferrier, 2003: Modeling of convective–stratiform precipitation processes: Sensitivity to partitioning methods. *J. Appl. Meteor.,* **42,** 505–527.

Masunaga, H., and C. D. Kummerow, 2005: Combined radar and radiometer analysis of precipitation profiles for a parametric retrieval algorithm. *J. Atmos. Oceanic Technol.,* **22,** 909–929.

Olson, W. S., C. D. Kummerow, Y. Hong, and W. K. Tao, 1999: Atmospheric latent heating distributions in the Tropics derived from satellite passive microwave radiometer measurements. *J. Appl. Meteor.,* **38,** 633–664.

Petty, G. W., 1994: Physical retrievals of over-ocean rain rate from multichannel microwave imagery. 2. Theoretical characteristics of normalized polarization and scattering indices. *Meteor. Atmos. Phys.,* **54,** 79–99.

——, and W. F. Krajewski, 1996: Satellite estimation of precipitation over land. *Hydrol. Sci. J.,* **41,** 433–451.

Shige, S., Y. N. Takayabu, W. K. Tao, and D. E. Johnson, 2004: Spectral retrieval of latent heating profiles from TRMM PR data. Part I: Development of a model-based algorithm. *J. Appl. Meteor.,* **43,** 1095–1113.

Shin, D.-B., and C. Kummerow, 2003: Parametric rainfall retrieval algorithm for passive microwave radiometers. *J. Appl. Meteor.,* **42,** 1480–1496.

Simpson, J., R. F. Adler, and G. R. North, 1988: A proposed Tropical Rainfall Measuring Mission (TRMM) satellite. *Bull. Amer. Meteor. Soc.,* **69,** 278–295.

Spencer, R. W., H. M. Goodman, and R. E. Hood, 1989: Precipitation retrieval over land and ocean with SSM/I: Identification and characteristics of the scattering signal. *J. Atmos. Oceanic Technol.,* **6,** 254–273.

Tao, W. K., S. Lang, J. Simpson, and R. Adler, 1993: Retrieval algorithms for estimating the vertical profiles of latent heat release: Their applications for TRMM. *J. Meteor. Soc. Japan,* **71,** 685–700.

——, C.-L. Shie, J. Simpson, S. Braun, R. H. Johnson, and P. E. Ciesielski, 2003: Convective systems over the South China Sea: Cloud-resolving model simulations. *J. Atmos. Sci.,* **60,** 2929–2956.

Wentz, F. J., 1997: A well-calibrated ocean algorithm for Special Sensor Microwave/Imager. *J. Geophys. Res.,* **102** (C4), 8703–8718.

Yanai, M., S. Esbensen, and J.-H. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.,* **30,** 611–627.

Yang, S., and E. A. Smith, 1999: Moisture budget analysis of TOGA COARE area using SSM/I-retrieved latent heating and large scale Q_2 estimates. *J. Atmos. Oceanic Technol.,* **16,** 633–655.