An Observationally Generated A Priori Database for Microwave Rainfall Retrievals

CHRISTIAN D. KUMMEROW, SARAH RINGERUD, JODY CROOK, DAVID RANDEL, AND WESLEY BERG

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

(Manuscript received 17 March 2010, in final form 8 August 2010)

ABSTRACT

The combination of active and passive microwave sensors on board the Tropical Rainfall Measuring Mission (TRMM) satellite have been used to construct observationally constrained databases of precipitation profiles for use in passive microwave rainfall retrieval algorithms over oceans. The method uses a very conservative approach that begins with the operational TRMM precipitation radar algorithm and adjusts its solution only as necessary to simultaneously match the radiometer observations. Where the TRMM precipitation radar (PR) indicates no rain, an optimal estimation procedure using TRMM Microwave Imager (TMI) radiances is used to retrieve nonraining parameters. The optimal estimation methodology ensures that the geophysical parameters are fully consistent with the observed radiances. Within raining fields of view, cloud-resolving model outputs are matched to the liquid and frozen hydrometeor profiles retrieved by the TRMM PR. The profiles constructed in this manner are subsequently used to compute brightness temperatures that are immediately compared to coincident observations from TMI. Adjustments are made to the rainwater and ice concentrations derived by PR in order to achieve agreement at 19 and 85 GHz, vertically polarized brightness temperatures at monthly time scales. The database is generated only in the central 11 pixels of the PR radar scan, and the rain adjustment is performed independently for distinct sea surface temperature (SST) and total precipitable water (TPW) values. Overall, the procedure increases PR rainfall by 4.2%, but the adjustment is not uniform across all SST and TPW regimes. Rainfall differences range from a minimum of $-57\%$ for SST of 293 K and TPW of 13 mm to a maximum of $+53\%$ for SST of 293 K and TPW of 45 mm. These biases are generally reproduced by a TMI retrieval algorithm that uses the observationally generated database. The algorithm increases rainfall by 5.0% over the PR solution with a minimum of $-99\%$ for SST of 293 K and TPW of 14 mm to a maximum of $+11.8\%$ for an SST of 294 K and TPW of 50 mm. Some differences are expected because of the algorithm mechanics.

1. Introduction

Remote sensing of rainfall has a long history dating back to early cloud indexing schemes using visible and infrared radiation such as those developed by Barrett (1970), Kilonsky and Ramage (1976), or Arkin (1979). Some methods used both visible and infrared radiation (Lovejoy and Austin 1979) to eliminate cold but non-raining cirrus clouds, while Griffith et al. (1978) used the temporal evolution of the infrared signal to refine rainfall estimates from geostationary images. With the introduction of the Special Sensor Microwave Imager (SSM/I) in 1987, and a study by Ebert et al. (1996), showing that microwave methods were generally superior to infrared methods for instantaneous rainfall estimates, the emphasis in rainfall retrievals shifted to these microwave sensors. The emphasis was further focused on the microwave area with the launch of the Tropical Rainfall Measuring Mission (TRMM) in 1997 that flies both active and passive microwave sensors (Kummerow et al. 2000). Systematic research prompted by this mission further advanced passive microwave retrieval schemes as well as introduced new space-borne radar algorithms. While the evaluation of products from the radar, radiometer, and their differences is ongoing (e.g., Wolff et al. 2005; Robertson et al. 2003; Berg et al. 2006) it is clear that both have strengths and weaknesses that may be exploited by combining the two sensors. Infrared methods, in the meantime, have undergone a transformation to concentrate primarily on high-resolution space and time estimates that are anchored by the passive microwave estimates. Sapiano and Arkin (2009) provide a comprehensive evaluation of these methods.

Radiometer methods themselves have a significant history but have largely converged upon physical methods over oceans, while remaining mostly statistical over land.
The physical methods employed over oceans depend either on conceptual precipitation models to derive relationships between rainwater and brightness temperatures (e.g., Wilheit et al. 1991; Petty 2001; Hilburn and Wentz 2008) or Bayesian schemes (Bauer et al. 2001; Kummerow et al. 2001; Marzano et al. 1999), which employ cloud-resolving model simulations to establish a priori databases of potentially precipitating clouds and their associated brightness temperatures. While not perfect, these cloud-resolving models do provide a physically complete description of the rainfall structures and thus offer a convenient alternative to somewhat ad hoc cloud structures used in the conceptual models.

Bayesian schemes that rely on cloud-resolving models benefit from the complete description of clouds but are not free of errors. Both the correctness as well as the completeness of these databases can lead to retrieval errors (Kummerow et al. 1999). The correctness problem was identified in that study to relate to errors in the cloud hydrometeor profiles. Particularly important errors were those errors related to the vertical rainfall profiles that could lead to more or less surface rainfall for the same column water content. A typical problem is illustrated in Fig. 1 showing two hydrometeor profiles with similar water contents and brightness temperatures but different surface rainfall rates (23 vs 29 mm h^{-1}). The differences in this case are caused by differences in both the vertical profiles of rainwater as well as the distinct ratios of rainwater and cloud water. This multiplicity of solutions can lead to biases if the cloud-resolving model database does not contain the proper distribution of clouds.

Without additional constraints, retrievals are thus subject to systematic biases that might be introduced by a cloud-resolving model that consistently favors one type of profile over another irrespective of the observed rain profiles. When the vertical profiles of rainfall were constrained by the TRMM precipitation radar (PR) as was done in Kummerow et al. (2006), the potential biases introduced by the cloud-resolving model are reduced significantly. There remain uncertainties due to unknown drop size distributions (DSD), but these were shown to be at a 5% level for rainfall accumulations. This represented only one-fourth of

![Fig. 1. Vertical hydrometeor profiles generated by cloud-resolving models. The two profiles have nearly indistinguishable Tb at TMI frequencies, but different surface rainfall rates.](image)
the 20% error that unknown DSD can introduce into the PR rainfall estimate. The reduction comes about because the radiometer measurements contain information on the integrated liquid water content.

The second major source of errors, discussed by Kummerow et al. (2006), was representativeness errors. While these were characterized in that study using PR-derived rain statistics, they were historically difficult to diagnose and correct for in Bayesian schemes when only cloud-resolving models were used to generate the a priori database. Incorrect ratios of convective, stratiform, and shallow rainfall in the simulations often lead to errors in the retrieved fraction and rainfall rates associated with these rain types. Very much related to this problem is also the tendency to use only raining pixels from the cloud-resolving models. This forces retrieval algorithms to classify pixels as either raining or not raining prior to applying the Bayesian methodology. Papers such as Ferraro et al. (1998) dealt solely with the errors introduced by distinct rainfall detection schemes used at that time. While the problem has been well recognized, it is not a simple problem to mitigate. Light rainfall or drizzle does indeed become virtually indistinguishable from nonprecipitating clouds in passive microwave signatures as illustrated in Berg et al. (2006) who looked at aerosol effects upon light precipitation.

The TRMM satellite offers the opportunity to use the radar as a means of constructing an observationally based a priori database. A number of potential methods exist. The simplest of these is to assume the TRMM PR represents the truth and to combine its surface rainfall with the coincident brightness temperatures from the TMI instrument. The method requires only some basic beam averaging to bring the radar footprints to the appropriate TRMM Microwave Imager (TMI) resolution. This solution will be referred to here as the empirical database in that it does not require physical consistency between the PR rain profiles and the TMI brightness temperatures. Thus, the geophysical parameters derived from the radar are treated as truth. In many cases, unfortunately, they do not yield the observed brightness temperatures when used as input to a radiative transfer model. The empirical database therefore cannot be transformed to work with other radiometers that do not share the same channels, view angles, and fields of view of the TRMM TMI sensor. This severely limits the utility of this database.

In contrast to empirical databases, physical approaches can be used to derive geophysical parameters that are fully consistent with both the PR and TMI measurements. While physical consistency with both sensors does not, by itself, guarantee the correct solution, it does better constrain the solution than either sensor alone. It also provides a solution that should approximate brightness temperatures of sensors such as SSM/I or the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) that are similar but not identical to TMI. Physical methods can thus exploit the radar–radiometer combination on the TRMM satellite to build databases for any similar radiometer, and forms the basis for radiometer retrievals for the Global Precipitation Mission (GPM) as described by Hou et al. (2009, manuscript submitted to Bull. Amer. Meteor. Soc.).

There exist combined algorithms employing physical principles in the literature. Olson et al. (1996) combined aircraft radar and radiometer measurements in a Bayesian framework. Because relatively little radar–radiometer data was available at that time, the study focused only on a single squall line. A broader approach designed to work jointly with the TRMM radar and radiometer data was described by Haddad et al. (1997). This algorithm used the TMI 10.7-GHz brightness temperature to infer the path integrated attenuation and thus constrains the PR’s assumed drop size distribution. While the method does not ensure physical consistency between the PR and all TMI channels, it does accomplish this to first order since the 10.7-GHz channels represents the bulk emission from the hydrometer column. Two more recent approaches that focus on oceanic backgrounds are Grecu and Olson (2006) and Masunaga and Kummerow (2005). The Grecu and Olson work used an optimal estimation algorithm to find drop size distributions to simultaneously match radar reflectivities and radiometer observations at 10.7, 19.35, and 37.0 GHz. Masunaga and Kummerow developed a similar framework but used the TRMM PR product as a starting point and varied the drop size distribution only for lightly raining pixels for which the radar could not determine its own drop size distribution from attenuation measurements. They also added the 85 GHz of TMI to the matching procedure. While the two latter studies represent distinct approaches, both studies needed to constrain the range of DSD in order to keep the solutions well behaved.

The scheme developed here retrieves geophysical parameters everywhere over ocean—both raining and nonraining—in order to avoid the completeness issues that have previously led to the need for rainfall screening procedures. It begins with the TRMM PR solution and modifies it only when necessary to achieve consistency with TMI observation. Unlike previous methods that focused solely on DSD adjustment, the current method also adds precipitation below the PR detection threshold. Increasing the areal extent of light rainfall is not only consistent with recent CloudSat observations (Ellis et al. 2009), but also very effective at forcing agreement between radar and radiometer in lightly raining scenes. The solution still makes use of cloud-resolving model (CRM) simulations to maintain a complete description of all relevant
geophysical parameters—not simply the ones that can be observed from the TRMM radar. Cloud water, water vapor, distinct ice habits and latent heating are thus still provided by the cloud-resolving model, albeit now severely constrained by the radar reflectivity profiles and the observed Tb.

Details of the database construction are presented in section 2. The discussion begins with the initial guess provided by the TRMM PR and the resulting differences between the Tb computed from this first guess and the observations. This is followed by details of the adjustment procedure. Section 3 then deals with practical issues related to the database construction and examines the properties of the database, while Section 4 briefly examines the impact of these complete databases upon a Bayesian rainfall retrieval algorithm using only radiometer data.

2. The physically consistent database

To be radiometrically consistent, the radar-derived geophysical parameters must, when used together with a radiative transfer computation, lead to the observed brightness temperatures. Unfortunately, microwave radiances are sensitive to additional parameters such as the surface emissivity that are not measured by the radar. Nonraining parameters including the surface wind speed, water vapor, and cloud water outside precipitating clouds must therefore be obtained from the radiometer itself. Radiometer footprints where the PR detects no rainfall are analyzed first.

An optimal estimation procedure based upon (Rogers 2000) and described for the present application by Elsaesser and Kummerow (2008) is used to retrieve surface wind (wind), total precipitable water (TPW), and cloud liquid water path (LWP) from the TMI observations. The sea surface temperature (SST) is specified from the Reynolds weekly climatology (Reynolds et al. 2002). The climatology was used primarily to ensure that the database SST could be made consistent with the SST used by the retrieval algorithm for sensors such as SSM/I that do not have the low-frequency channels required to derive SST independently.

Water vapor is retrieved assuming the specific humidity is distributed exponentially with a fixed scale height according to

\[ q(z) = q_0 e^{-z/H}, \]

where, \( H \), the scale height, is set to 2.3 km. The air temperature in this formulation is assumed to be equal to the SST with a zonal mean lapse rate determined from the European Centre for Medium-Range Weather Forecasts (ECMWF) and given by

\[ \Gamma = -1.631 e^{-0.5(\phi-4.464/24.178)^2} - 4.464 + 0.00075\phi - 0.00020\phi^2, \]

where \( \Gamma \) is the lapse rate (in K km\(^{-1}\)) and \( \phi \) is the absolute value of the latitude. Surface emissivity is computed from a specular emissivity model based on Deblonde and English (2001) and a rough sea surface model based on Liebe et al. (1991) model. Slight modifications were made to the model for better agreement with the Millimeter-Wave Propagation Model (MPM93) of Liebe et al. (1993).

Raining areas are screened from the algorithm using the TRMM PR rain flag.

In tailoring the optimal estimation (OE) retrieval described by Elsaesser and Kummerow (2008) to the present study’s purposes, all nine TMI channels are utilized. Near-surface wind speed and LWP are retrieved directly while the TPW is inferred from the retrieved specific humidity \( q_0 \) and vertical humidity profile as given in Eq. (1). With nine channels available from TMI, the above solution is well constrained and could alternatively be cast as a minimum variance problem without any a priori constraints. In this particular application there is little difference between the two approaches. The OE approach is retained primarily because it is more robust when dealing with outliers.

Key to reproducing OE results is the uncertainties that are specified both for the a priori values as well as the brightness temperatures. For this application, a priori values of 8.0 m s\(^{-1}\) (wind), 24.7 kg m\(^{-2}\) (TPW), and 0.07 kg m\(^{-2}\) (LWP) are based on climatology. The uncertainties, expressed as \( S_q \) in the OE framework, are assumed to be large, 15.1 kg m\(^{-2}\), and 0.19 kg m\(^{-2}\), respectively, resulting in little dependence from the a priori guess in the final solution. No covariance among the parameters is assumed. The brightness temperature uncertainties \( S_q \) represent the measurement, forward model, and assumption errors. They follow the values set in Elsaesser and Kummerow (2008) of 1.03, 1.39, 1.23, 1.83, 1.21, 1.28, 2.32, 1.89, and 3.49 K for the 10-V, 10-H, 19-V, 19-H, 21-V, 37-V, 37-H, 85-V, and 85-H channels, respectively.
not realistic for the modeling component of these uncertainties, it does represent a worst-case scenario.

When compared to radiosonde (TPW) and buoy (wind) measurements, the OE results have a bias of $-0.04$ mm and $-0.82$ m s$^{-1}$, respectively with root-mean-square errors of 0.79 mm for TPW and 1.50 m s$^{-1}$ for wind. These results are consistent with other published retrievals of these parameters (e.g., Deblonde 2001; Mears et al. 2001). More important for the present application, however, is the match between observed and computed brightness temperatures (Tb). Comparisons are shown in Fig. 2. While there is some scatter, the majority of nonraining points clearly lie very close to the 1:1 line that indicates a good fit between observed and simulated Tb. Outliers are retained in the files, but are eventually rejected when the final database is constructed. The rejection is based upon excessive differences between simulated and observed Tb that cannot be explained by uncertainty arguments. Rejection of these outliers was found to bias rainfall in the database slightly—implying that the rejected pixels are not random. The bias level was at the 1%–2% level depending upon the SST and TPW bin and thus represents a fundamental uncertainty.

**a. The default rain structure**

Nonraining parameters are retrieved by TMI when the PR indicates that no rain is present in the TMI field of view. When precipitation is present in the PR data, the TRMM PR operational rainfall profiles (TRMM 2A25, V6) are used as a first guess. Only the middle 11 pixels of the PR are used to avoid using PR pixel whose surface clutter contamination extends too high above the earth’s surface. Drop size distributions as well as ice properties that are not provided with the output were taken from Iguchi et al. (2000). The brightband characteristics described by Masunaga and Kummerow (2005) were used. Still, the PR radar does not provide all the information needed to compute brightness temperatures. The SST, wind speed, water vapor, cloud water, and cloud ice, as well as the nature of precipitating ice (i.e., snow, graupel, or hail) must be
obtained by separate means. The SST and wind speed is obtained through interpolation of the nonraining fields outside the precipitation. Cloud water, water vapor, and the distribution of ice species are obtained by matching reflectivity profiles from the PR to a cloud-resolving model database. The database is the same used in previous version of the Goddard profiling algorithm (GPROF) and is described in detail in Kummerow et al. (2001). In this case, only the cloud-resolving model that best fits the observations is taken, with no weighting as is done in Bayesian schemes. The matching criteria are the vertical distributions of rainfall rate, ice content, and freezing level reported by the PR. Once the best match is identified, the complete hydrometeor profile, including water vapor and latent heating structure from the cloud-resolving model is carried forward. This includes the rainfall at the surface where PR does not have data but interpolates values instead. In rare occasions, no suitable cloud-resolving model profiles could be found. In these cases, an artificial cloud model profile was generated based upon the PR profile. These were generally light rain profiles that exceeded 17 dBZ in a few layers, but did not have continuous reflectivity profiles. These pixels were replaced by homogeneous vertical rain columns.

At this point in the process, all nonraining TMI pixels have an assigned TPW, LWP, and wind, while all raining PR pixels are filled with hydrometeor information as well as cloud water and water vapor from the cloud-resolving model. A final interpolation is necessary to map the nonraining TMI pixels into PR pixel space. This includes some nonraining PR pixels for which no TMI retrievals were performed because they were located within raining TMI footprints. Because TMI pixels overlap and PR pixels are contained within more than one TMI field of view (FOV), each PR pixel is filled with a weighted average of nearest neighbors. In this manner, each PR pixel is assigned surface properties and a raining or nonraining hydrometeor profile. Radiative transfer computations are used and the TMI beam is reconstructed to derive TMI brightness temperatures from this high-resolution (~4 km) hydrometeor field. Figure 3a shows PR rainfall rates for a small orbit segment. Figure 3b shows the computed 19-GHz vertically polarized (19V) Tb while Fig. 3c shows the observed 19V Tb for the same orbit section. Figure 3d shows the difference between the simulated and observed Tb. If the difference is negative, it implies too little emission in the simulated scene (i.e., not enough water) and the PR rain should be increased to match observed Tb. The opposite is true if the simulated Tb is warmer than the observed Tb. To first order, there appears to be relatively good agreement between computed and observed Tb suggesting that the relatively good agreement between PR and TMI rainfall in V6 of the products is indeed physically consistent. Figure 4 is similar to Fig. 2 but for all pixels rather than simply nonraining ones. It again shows the scatter between computed and observed 19V Tb for July 1999. While the agreement is generally good, Fig. 4 shows that there are small but systematic biases between these Tb that must be removed for the database to be consistent with both radar and radiometer observations.

If mean biases between observed and simulated Tb are viewed on a global map, there is a clear trend of biases that follows the seasonal warming and cooling as well as more subtle differences related to meteorological regimes. Figure 5 shows these biases for December-January-February (DJF) and June-July-August (JJA), respectively. From a database perspective, the difference between these figures is problematic since it requires an adjustment that is regionally and temporally variant. Such an adjustment is undesirable because it is not only unphysical, but it can also make simple shifts in the rainfall location appear as changes in the overall rainfall.

Instead of using time and regionally varying characteristics of precipitation, this study uses SST and TPW to characterize biases between observed and simulated Tb. When viewed in a coordinate system of SST and TPW as shown in Fig. 6, only vestiges of the seasonal differences are evident. In all four seasons (DJF, MAM, JJA, SON) shown in Fig. 6, the general patterns are extremely consistent. At the very warmest SST and TPW, the PR solution leads to brightness temperatures at 19V that are slightly warmer than the observed ones. The same is true for cold SST with very low water vapor amounts. Over a wide swath of the plot, however, the PR leads to brightness temperatures that are somewhat colder than the observations. The largest differences, except for some edge pixels, which have poor sampling, appear to be in a diagonal strip extending from an SST of 289 and TPW of 25 mm to an SST of 295 and TPW of 45 mm. This corresponds roughly to midlatitude winter hemisphere precipitation.

The brightness temperature differences shown in Fig. 6 are also similar to rainfall biases among V6 of the operational PR and TMI products that were first discussed by Berg et al. 2006. The top panel of Fig. 7 shows the rainfall biases from version 6 of the operational TRMM PR and TMI products as a function of SST and TPW for 12 months corresponding to June 1999 to May 2000. The same 12 months (corresponding to the four seasons shown in Fig. 6) are aggregated in the bottom panel of Fig. 7. The similarities between the two panels argue for a physical basis to the differences between TRMM radar and radiometer rainfall estimates rather than simple algorithm errors. If one focuses on the regions of Fig. 7 where the bulk of the rain occurs (heavy contours), it is clear that this corresponds in the tropics to areas where there is relatively good agreement between PR and TMI rainfall (as well as computed
and observed Tb) while the extratropics are regions where the PR had less rain than TMI and the simulated Tb tend to be colder than the observations. The latter implies that the rain from the radar needs to be increased in order to be consistent with TMI.

b. The adjustment procedure

Because the database is intended for an operational radiometer algorithm, the adjustment procedure was made purposely conservative. The PR operational rainfall product, that is, was adjusted as little as possible while still achieving consistency between observed and computed brightness temperatures. A full year (June 1999–May 2000) of combined radar–radiometer retrievals over the middle 11 PR pixels as shown in Fig. 3 are used as a baseline. Nonraining areas generally agree quite well ($\Delta Tb < 1$ K). This is expected as the OE approach works by minimizing differences between observed and simulated Tb. Larger discrepancies exist in raining areas. Where computed Tb are too low, it was found that Tb could be raised most easily by adding rainfall to PR pixels that is below the detection threshold of the PR. Compared to

Fig. 3. (a) Observed rainfall from TRMM PR, (b) computed brightness temperatures from PR hydrometeor retrievals, (c) observed brightness temperatures, and (d) differences between observed and computed brightness temperatures.
modifying DSD, the addition of subthreshold rainfall was found to be more effective at increasing Tb for a given increase in rainfall. This additional rainfall is allowed in any PR pixel that borders a pixel defined as “rain certain” by the PR. The maximum rainfall allowed in these pixels is limited to 0.25 mm h$^{-1}$ and thus remains fully consistent with the PR observations that cannot detect rainfall at these low rain rates. The procedure adds significant raining pixels to be more consistent with the more sensitive CloudSat results that show a much greater probability of precipitation than does TRMM (Ellis et al. 2009).

Mechanically, the adjustment was implemented by adding the maximum 0.25 mm h$^{-1}$ to every PR pixel surrounding rain-certain pixels. A complete year of the database was rerun and differences between simulated and observed Tb were compared to the “default” run for the entire year as shown in Fig. 6 or the bottom panel of Fig. 7. If the adjustment added too much water such that the new bias was now positive in a given SST and TPW bin, the bin was adjusted with an interpolated rain amount designed to eliminate the residual bias between observed and simulated 19V Tb for that SST and TPW bin.

The percentage increase in rainfall for each SST and TPW bin from the above procedure is shown in Fig. 8a with the corresponding decrease in Tb differences between simulated and observed Tb shown in Fig. 8b. Differences can be contrasted with those shown in the bottom panel of Fig. 7. While the rainfall increase from this procedure is quite modest, there is a diagonal swath in the left-hand side of Fig. 8a that shows significant increases in rainfall. This occurs for SST and TPW bins that contain light and relatively little rainfall. Adding even 0.25 mm h$^{-1}$ to significant areas causes a large overall increase in rain accumulation in these regions.

The above procedure has the effect of increasing the rainwater content and thus increasing the computed Tb everywhere that the simulated Tb was originally too low. Regions shown in Fig. 7 where the computed Tb were already too warm cannot be modified by adding rainfall as this can only increase Tb further. There are also regions

Fig. 4. Comparisons between observed and computed Tb (K) from “default” hydrometeor profile. TRMM data over oceans from July 1999 is used.
shown in Fig. 7 where adding the maximum 0.25 mm h$^{-1}$ is not sufficient to remove the Tb bias. For these pixels, the rainfall DSD is modified—either increased or decreased in order to remove the Tb bias. Assuming smaller drops than the operational PR algorithm effectively lowers the PR reflectivity for the same water content and thus requires more water in the atmospheric column to produce the observed reflectivity. The opposite is true if larger drops are assumed. The DSD is modified only for pixels for which the PR has no independent DSD information coming from the observed path integrated attenuation (PIA). Following Masunaga and Kummerow (2005), the median mass-weighted diameter $D_0$ is modified in the cloud-resolving models from the default PR assumption to a value of $\pm 40\%$ larger–smaller than the default assumption. The $40\%$ threshold is necessary to avoid occasional numerical instabilities, but is not needed often in the current procedure. As before, the PR observed reflectivities are then used to find the best fit cloud-resolving model profile and the new solution is constructed. The optimal adjustment is once again found by interpolation between the reference simulation (now with sub-PR threshold rain added) and the $\pm 40\%$ $D_0$ simulation. At this point, all differences between simulated and observed Tb at 19-GHz vertical polarization.

In addition to adjusting the rainfall, the procedure simultaneously adjusts ice densities in order to achieve reasonable agreement between simulated and observed 85-GHz Tb. This is achieved strictly by converting snow ($\rho = 0.1$ g m$^{-3}$) to graupel ($\rho = 0.4$ g m$^{-3}$) when the simulated Tb at 85 GHz are too warm and graupel turns into snow when the simulated Tb are too cold. Increasing the density of ice increases the scattering and consequently lowers computed Tb at 85 GHz.

Table 1 shows the final Tb biases and root-mean-square differences between observed and simulated Tbs for all pixels for the full year of data. Biases can be seen to be uniformly small across all TMI channels. Even though the rainfall and ice adjustments are performed to minimize the 19- and 85-V channels, the other channels are also reacting to the liquid and ice in the column. Standard deviations are not insignificant. Reducing these, as well as the cross correlation of errors shown in Table 2 is something that a more sophisticated retrievals scheme should strive for.

3. The database construction

The above procedure created roughly $6.2 \times 10^7$ raining and nonraining profiles corresponding to a full year of analyzed TRMM data from 1 June 1999 to 31 May 2000. Unfortunately, the database is not complete as TRMM never samples very cold SST.

a. Extension of the database to colder SST

A shortcoming of the database that is generated from the TRMM radar and radiometer is that the satellite does
not pass over temperate or polar regions. While the sampled SST and TPW covers a somewhat broader area than would be achieved if the database were constructed strictly for individual latitude–longitude boundaries, it does not extend over all latitudes. Figure 10 shows the latitudinal extent covered by the database for a boreal summer and winter. For this figure, an arbitrary number of 100 database entries in the appropriate SST and TPW bin were chosen to signify that a valid retrieval was possible. If the database had fewer than 100 entries, no retrieval is made. The figure therefore shows the percentage of times that these criteria were met for a 1-month period.

Figure 10 clearly shows that the tropics are well covered by TRMM. The figure also shows that the first SST and TPW that are not covered by TRMM appear around 45°N and 45°S, and that by 60°, the a priori database has less than 50% coverage irrespective of the season.

Because the core satellite of the GPM will have an inclination of 65°, it will, on occasion, sample SST right up to the ice edge and thus fill the entire SST/TPW space. As an interim solution, the colder SST and TPW bins needed for polar-orbiting satellites such as SSM/I or AMSR-E are currently filled by removing the lower atmospheric layers from existing profiles. Starting with
SST of 284, 283, and 82 K, the lowest 500 m of the atmosphere are removed by simultaneously lowering the SST by 3 K. Surface precipitation is computed for each new profile based upon the liquid and ice water content in the lowest remaining atmospheric layer. While this was viewed as a practical method to fill in some colder regimes, it clearly should not be viewed with as much confidence as the observed portion of the database and is used primarily as a placeholder for the GPM mission scheduled for launch in 2013 (Hou et al. 2009, manuscript submitted to Bull. Amer. Meteor. Soc.). Even colder SSTs (SST < 278 K) are not produced as it was thought to represent too large a departure from the observed structures. Retrieval algorithms can easily keep...
track of database entries that were added or nonexistent as is the case for very cold SST.

b. Clustering of the database profiles

A second item of concern was the sheer number of profiles generated by the TRMM satellite, even if applied only to the inner 11 pixels of the PR. One year of data generated approximately 62 million database entries and 10 times as many could be generated over the 10 yr of mission life. A cluster analysis is therefore used to reduce the database to 1000 raining and 200 nonraining entries within each SST and TPW bin. This is done strictly to speed up the retrieval algorithm. The number of entries to retain was arrived at by insisting that the clustering not change the final retrieval solution by more than 0.01%. Cluster analysis is an empirical method designed to find similarities in multidimensional data and group points together in “clusters.” The goal of the technique is to define groups that are distinct from one another and group data points that are highly associated together. This is achieved using an iterative process in which central points, or centroids, are chosen (first at random). The Euclidean distance from each data point to each centroid is computed, after which each individual point is assigned to the nearest

![Fig. 8](image-url) (a) Percentage increase in rainfall from the addition of rainfall below the threshold of PR-detection capabilities ($R < 0.25$ mm h$^{-1}$). (b) The residual biases between observed and simulated Tb (K) for the year of analysis (June 1999–May 2000).

![Fig. 9](image-url) (a) Change from reference PR rainfall (mm) after sub-PR threshold rainfall addition and DSD adjustment procedures. (b) The residual biases between observed and simulated 19-GHz Vpol Tb (K) for the year of database (June 1999–May 2000).
centroid. The center of each resulting cluster is then recalculated, and the distances computed again. For this analysis the number of iterations of this process is set at 10. The specific clustering technique used here is the k-means technique as described by Anderberg (1973). As input, the nine TMI brightness temperatures as well as the surface rainfall are used. Mean Tb and surface rainfall representative of each cluster are computed and stored in the final database, along with the number of individual pixels that make up each cluster. The number of individual entries is necessary to keep track of the relative frequency of occurrence of each original profile.

c. Database rainfall

The observationally generated database serves as the a priori information for a Bayesian rainfall retrieval algorithm. The database described here serves specifically as the a priori information for the operational TRMM TMI 2A12 product, version 7. As such, it is important to understand not only how the database was created, but its properties and relationship to other products. Figure 11 compares the database mean rainfall to the existing operational (V6) TRMM algorithms. The new database is somewhat lower than the PR in the ITCZ region between the equator and 10°N. This is a direct result of the reduction in rainfall for very high SST and TPW values apparent in Fig. 9 also. Specifically, the domain in Fig. 9 that covers 302–303 K and 65 mm of TPW was too warm based on the PR first guess and thus has a reduced rainfall content after the adjustment. The same phenomenon is also active between

![Figure 10](image-url)

**Table 1.** Biases and root-mean-square differences between observed and simulated Tb corresponding to a full year of database entries for the different TMI channels.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Bias</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10V</td>
<td>-0.13</td>
<td>2.73</td>
</tr>
<tr>
<td>10H</td>
<td>1.80</td>
<td>4.59</td>
</tr>
<tr>
<td>19V</td>
<td>-1.72</td>
<td>4.82</td>
</tr>
<tr>
<td>19H</td>
<td>-1.50</td>
<td>7.00</td>
</tr>
<tr>
<td>21V</td>
<td>0.20</td>
<td>3.40</td>
</tr>
<tr>
<td>37V</td>
<td>1.50</td>
<td>6.48</td>
</tr>
<tr>
<td>37H</td>
<td>-0.87</td>
<td>11.84</td>
</tr>
<tr>
<td>85V</td>
<td>1.72</td>
<td>8.56</td>
</tr>
<tr>
<td>85H</td>
<td>-1.20</td>
<td>12.08</td>
</tr>
</tbody>
</table>

**Table 2.** Cross correlation of errors among TMI channels corresponding to differences between observed and simulated Tb for a full year of database entries.

<table>
<thead>
<tr>
<th>Channel</th>
<th>10V</th>
<th>10H</th>
<th>19V</th>
<th>19H</th>
<th>21V</th>
<th>37V</th>
<th>37H</th>
<th>85V</th>
<th>85H</th>
</tr>
</thead>
<tbody>
<tr>
<td>10V</td>
<td>1.0</td>
<td>0.59</td>
<td>0.53</td>
<td>0.31</td>
<td>0.24</td>
<td>0.29</td>
<td>0.19</td>
<td>-0.09</td>
<td>-0.11</td>
</tr>
<tr>
<td>10H</td>
<td>0.59</td>
<td>1.0</td>
<td>0.54</td>
<td>0.58</td>
<td>0.40</td>
<td>0.31</td>
<td>0.31</td>
<td>0.007</td>
<td>0.02</td>
</tr>
<tr>
<td>19V</td>
<td>0.53</td>
<td>0.54</td>
<td>1.0</td>
<td>0.82</td>
<td>0.62</td>
<td>0.68</td>
<td>0.63</td>
<td>0.08</td>
<td>0.14</td>
</tr>
<tr>
<td>19H</td>
<td>0.31</td>
<td>0.58</td>
<td>0.82</td>
<td>1.0</td>
<td>0.72</td>
<td>0.63</td>
<td>0.69</td>
<td>0.07</td>
<td>0.16</td>
</tr>
<tr>
<td>21V</td>
<td>0.24</td>
<td>0.40</td>
<td>0.62</td>
<td>0.72</td>
<td>1.0</td>
<td>0.53</td>
<td>0.54</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>37V</td>
<td>0.29</td>
<td>0.31</td>
<td>0.68</td>
<td>0.63</td>
<td>0.53</td>
<td>1.0</td>
<td>0.91</td>
<td>0.26</td>
<td>0.42</td>
</tr>
<tr>
<td>37H</td>
<td>0.19</td>
<td>0.31</td>
<td>0.63</td>
<td>0.69</td>
<td>0.54</td>
<td>0.91</td>
<td>1.0</td>
<td>0.18</td>
<td>0.46</td>
</tr>
<tr>
<td>85V</td>
<td>-0.09</td>
<td>0.007</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.26</td>
<td>0.18</td>
<td>1.0</td>
<td>0.78</td>
</tr>
<tr>
<td>85H</td>
<td>-0.11</td>
<td>0.02</td>
<td>0.14</td>
<td>0.16</td>
<td>0.12</td>
<td>0.42</td>
<td>0.46</td>
<td>0.78</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Figure 10.** Zonal maps showing fraction of retrievals that can be made in June–September–December 1999 and December 2000 from database constructed as a function for SST and TPW.
the equator and 10°S where the South Pacific convergence zone (SPCZ) dominates the rainfall. The new database is also lower than either the TMI (2A12) or the combined (2B31) products in this portion of the globe. This reduction is primarily a function of the new database starting out with the PR product as a base and adjusting this base only if needed. The operational radiometer and combined algorithms do not have such a constraint. As such, the new database is closer to the PR value than the other products.

Figure 9 also predicts the increase in extratropical rainfall seen in Fig. 11. The new database is now larger than any of the TRMM operational products because Tb had to be increased substantially in order to make PR and TMI consistent. Compared to previous versions of TMI, the new database adds significant rain in the midlatitude winter hemisphere. This was also evident, although not as obvious in the Tb biases between the older version of TMI and PR in regions of low SST and water vapor as shown in Fig. 7.

d. Rain frequency in the database

Unlike databases that are constructed from CRMs of raining scenes, the current database includes both raining and nonraining entries. The probability of rainfall is considered first. Because of the database construction, rain is defined whenever the PR-observed rainfall (defined as rain certain in the PR rainfall product). This is convolved with the 37-GHz footprint to produce a database with TMI resolution. Because of the convolution with a Gaussian antenna patterns, the probability of rain in the database becomes somewhat arbitrary—depending primarily upon

the limits of the Gaussian integral. Table 3 shows the rainfall that is discarded if the very light rainfall amounts in the database are set to zero in order to add realism to the rain–no-rain discrimination. The current database uses a somewhat arbitrary value of 0.01 mm h⁻¹ as the threshold for precipitation. Entries below this threshold are set to zero.

The above discussion illustrates the general problem of defining the probability of rain with a measurement system that does not define rainfall in a binary manner. While it is straightforward to compute a pixel average rainfall of 0.0001 mm h⁻¹ based upon a PR pixel that is barely raining and a Gaussian convolution function, it is not at all straightforward to interpret this rainfall rate. It is well below any reasonable detection threshold of the TMI, the PR, or even any ground-based measurement device. Assuming a homogeneous rain column over a 4-km depth, a rainfall rate of 0.01 mm h⁻¹ amount would only increase Tb at 10 GHz by 0.01 K, while increasing the 19-GHz Tb by 0.02 K. This is consistent with the precision with which

---

### Table 3. Probability of rain (%) as a function of the rain threshold used and the fractional rain accumulation (%) captured above the given threshold.

<table>
<thead>
<tr>
<th>Rain threshold (mm h⁻¹)</th>
<th>Rainfall reduction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000 001</td>
<td>0.000</td>
</tr>
<tr>
<td>0.000 01</td>
<td>0.054</td>
</tr>
<tr>
<td>0.0001</td>
<td>0.060</td>
</tr>
<tr>
<td>0.001</td>
<td>0.071</td>
</tr>
<tr>
<td>0.01</td>
<td>0.230</td>
</tr>
</tbody>
</table>
Tb are provided and already much smaller than the noise value of 0.5 K at these channels. The threshold of 0.01 mm h\(^{-1}\) was thus retained in the database primarily to keep some connection with rainfall rates that are potentially detectable.

The distribution of rainfall rates is shown in Fig. 12. For comparison purposes, the distribution that would have been obtained from the “empirical” database—namely, the PR solution averaged to the TMI resolution is also shown, as is the distribution of the PR rainfall rates at their native resolution. The differences between native PR and the database illustrate some of the difficulties in directly interpreting rainfall probabilities or rainfall histograms when sensors of different spatial resolution are used.

e. Cloud differences as a function of SST and TPW

Using SST and TPW to construct the database was motivated originally by the work of Berg et al. (2006) who showed consistent biases between radar and radiometer as a function of TPW. The SST is needed as it, together with the water vapor, describes the freezing level in raining conditions as first described by Wilheit et al. (1991). Biases between radar and radiometer, however, could not be directly attributed to the total precipitable water because it was implicitly accounted for in the retrieval scheme. Instead, that paper speculated that cloud physics might be responsible for creating systematic differences between cloud properties and generally static retrieval assumptions.

A separate study by Masunaga and Kummerow (2006) examined cloud properties based upon their infrared cloud-top properties and radar echo. This study merges the IR and PR cloud description with a TPW characterization. Figure 13 shows the vertical structures for clouds in different TPW regimes. Low TPW values (<15 mm) are characterized by shallow precipitating systems. The most frequently observed structure is one with a radar echo top near 2 km and a cloud top near 280 K. While it is possible to get deeper clouds even at low TPW values, echo tops rarely exceed 4 km. As water vapor increases, the shallow clouds persist but can be observed to exist at slightly warmer temperatures. For precipitating scenes, this means only that the TPW also contains some information of the

![Figure 13](image-url)
atmospheric temperature. Shallow cloud tops therefore become warmer in warmer environments. Beyond the shallow cumulus, a deeper cloud appears when TPW exceeds 15 mm. This cloud deepens with increasing water vapor and is seen to split somewhat into a shallow and a congestus part, which transforms into a shallow and a deep convective cloud up to 60 mm in TPW. Above this threshold, the shallow cloud disappears and the histogram is dominated by extremely deep convective systems.

While the cloud structures shown in Fig. 13 do not immediately translate into any discrete microphysical properties that need to be accounted for in either radar or radiometer algorithms, they do point to the changing morphology of clouds in different water vapor environments. These distinct shapes, furthermore, are not nearly as evident if sea surface temperature is used to classify clouds.

4. The rainfall retrieval—An example

This paper focused upon details of how a database of observed precipitating and nonprecipitating profiles was created. Here, the impact of these comprehensive databases upon a retrieval algorithm is investigated. The purpose is not to investigate the retrieval algorithm as much as it is simply to demonstrate the effect of these databases upon a Bayesian scheme as has been implemented for the TRMM operational algorithm. Because the database contains the observed proportions of raining and nonraining pixels, no a priori decision is made regarding the rain status of a given pixel. Brightness temperatures are simply compared to the a priori database of cloud structures and their associated Tb. The uncertainty is assigned based upon channel noise figures plus the difference between observed Tb and the simulated Tb during the database construction. Cross correlation of errors based upon Table 2 are used in the retrieval.

Three databases are used. Each consists of 3 months (DJF 1999–2000). All retrievals are performed for December 1999. The first database is the empirical database described in the introduction. Here, the observed PR rainfall is merged with the observed TMI Tb, and in this case, the PR rainfall can be considered truth. The second database is termed the “ideal” database. This database uses the final adjusted PR rainfall but the observed TMI Tb. It is termed ideal because it presumes that rainfall could be adjusted in such a way that each of the TMI Tb exactly matched the observed ones. This is the goal of the adjustment procedure but one that is met only rarely. The third database is the “actual” database that contains the adjusted rainfall rates and the Tb that are computed from the hydrometeor profiles. This is the only database that can be extended to other sensors and the one that is used in the TRMM TMI operational algorithm starting with version 7. Results for the three retrievals, along with the TRMM PR zonal accumulations are shown in Fig. 14.

As can be seen from the figure, the empirical database leads to nearly the exact PR solution. This is an important conclusion in that it affirms that the Bayesian retrieval scheme faithfully reproduces the mean rainfall accumulations found in the a priori database. While the retrieval adds scatter to individual points, one can be confident that the rainfall accumulations derived from the radar–radiometer combination can be reproduced by the radiometer only algorithm. The ideal and actual algorithms more closely follow the database and the changes it forces in the PR solution. The ideal solution can be seen to add a little rainfall to the subtropics where the PR is apparently a bit dry. The actual database adds even more rainfall to the subtropics. This might be an indication that the PR–TMI solution in the database did not add enough rainfall to fully match the observed Tb in the regions of isolated convection that are often found in the subtropical subsidence zones. As with Fig. 11, however, all changes appear small and well behaved based upon the initial disagreement between TRMM PR and TMI observations.
Acknowledgments. This project was funded under NASA Grant 5-39527 as part of the ongoing algorithm development effort for TRMM and GPM.

REFERENCES


