Optimization of Cloud-Radiation Databases for Passive Microwave Precipitation Retrievals over Ocean

EUN-KYOUNG SEO, SUNG-DAE YANG, MIRCEA GRECU, GEUN-HYEOK RYU, GUOSHENG LIU, SVETLA HRISTOVA-VELEVA, YOO-JEONG NOH, ZIAD HADDAD, AND JINHO SHIN

a Department of Earth Science Education, Kongju National University, Chungnam, South Korea
b Applied Meteorological Research Division, National Institute of Meteorological Sciences, Jeju, South Korea
c Goddard Earth Sciences Technology and Research Center, Morgan State University, Baltimore, Maryland, and Laboratory for Atmospheres, NASA Goddard Space Flight Center, Greenbelt, Maryland
d National Meteorological Satellite Center, Korea Meteorological Administration, Jincheon-gun, South Korea
e Florida State University, Tallahassee, Florida
f Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, Los Angeles, California
g Colorado State University, Fort Collins, Colorado

(Manuscript received 29 September 2015, in final form 27 May 2016)

ABSTRACT

Using Tropical Rainfall Measuring Mission (TRMM) observations from storms collected over the oceans surrounding East Asia, during summer, a method of creating physically consistent cloud-radiation databases to support satellite radiometer retrievals is introduced. In this method, vertical profiles of numerical model-simulated cloud and precipitation fields are optimized against TRMM radar and radiometer observations using a hybrid empirical orthogonal function (EOF)–one-dimensional variational (1DVAR) approach. The optimization is based on comparing simulated to observed radar reflectivity profiles and the corresponding passive microwave observations at the frequencies of the TRMM Microwave Imager (TMI) instrument. To minimize the discrepancies between the actual and the synthetic observations, the simulated cloud and precipitation profiles are optimized by adjusting the contents of the hydrometers. To reduce the dimension of the hydrometeor content profiles in the optimization, multivariate relations among hydrometeor species are used.

After applying the optimization method to modify the simulated clouds, the optimized cloud-radiation database has a joint distribution of reflectivity and associated brightness temperatures that is considerably closer to that observed by TRMM PR and TMI, especially at 85 GHz. This implies that the EOF–1DVAR approach can generate profiles with realistic distributions of frozen hydrometeors, such as snow and graupel. This approach may be similarly adapted to operate with the variety and capabilities of the passive microwave radiometers that compose the Global Precipitation Measurement (GPM) constellation. Furthermore, it can be extended to other oceanic regions and seasons.

1. Introduction

Two-thirds of the global rainfall amount falls within the tropics, which makes the measurement of tropical rainfall very important in understanding the weather and climate (e.g., Simpson et al. 1988). The launch of the Tropical Rainfall Measuring Mission (TRMM) satellite in 1997 marked a new era in precipitation retrievals from space. TRMM carries two major instruments dedicated to rain retrievals: the multichannel microwave radiometer [TRMM Microwave Imager (TMI)] and the first spaceborne precipitation radar (PR). In particular, PR provides valuable measurements to validate and improve the rainfall amount estimated from TMI (e.g., Iguchi et al. 2000; Olson et al. 1999; Kummerow et al. 2001, 2011; Berg et al. 2006; Shige et al. 2006, 2008; Seo et al. 2007, 2015; Gopalan et al. 2010).

Corresponding author address: Eun-Kyoung Seo, Department of Earth Science Education, Kongju National University, 56 GongjuDaehak-Ro, Kongju, Chungnam 32588, South Korea. E-mail: ekseo@kongju.ac.kr

Copyright 2016 American Meteorological Society
The official TMI precipitation algorithm, the Goddard Profiling algorithm (GPROF), is based on the derivation and use of an a priori cloud-radiation database. The database is composed of a large set of cloud profiles and their corresponding brightness temperatures (TBs). The GPROF utilizes this predefined database to determine the most likely precipitation profiles associated with a set of observed TMI TBs using Bayes’s theorem. The cloud profiles in the database can be produced by physical cloud-resolving models (Smith et al. 1992; Mugnai et al. 1993; Olson et al. 1996; Kummerow et al. 2001), or constructed by using simple conceptual models (Wilheit et al. 1977; Kummerow et al. 1991), or, as is the case most recently, constructed from radar-based retrievals (Aonashi et al. 2009; Kummerow et al. 2011, 2015). In the vein of improving the cloud profiles in the database, the latest GPROF uses an observationally generated a priori database for microwave rainfall retrievals (Kummerow et al. 2011, 2015). Regardless of the approach in constructing the cloud-radiation database, the degree of realism of the cloud profiles still remains a key to the successful performance of the retrieval algorithms (Kummerow et al. 2006, 2011; Biggerstaff and Seo 2010). Hence, the representativeness of the simulated cloud profiles and their local environmental and surface conditions become one of the critical components in Bayesian algorithms.

Previous studies have shown that the vertical and horizontal distributions of simulated hydrometers depend significantly upon the microphysical scheme used in their simulations (Ferrier 1994; Panegrossi et al. 1998; Hristova-Veleva 2000; Biggerstaff et al. 2006; Seo and Biggerstaff 2006; Grecu and Olson 2006; Kim et al. 2013). It was shown that 37- and 85-GHz scattering signatures of cloud model simulations are, in general, stronger than those of TMI observations due to the overproduction of precipitation-sized ice particles in the simulations or their inaccurate representation in the forward models (e.g., Biggerstaff et al. 2006; Lang et al. 2007; Biggerstaff and Seo 2010; Han et al. 2010; Seo and Biggerstaff 2012). Consequently, rain retrievals are dependent upon the underlying microphysical parameterizations and assumptions used in the cloud-resolving model generation of the precipitation radiation database. Therefore, a critical step toward producing realistic retrievals is ensuring that the representation of the geophysical parameters, especially graupel, in the databases or forward models is physically accurate.

Using a new physically based statistical methodology, Biggerstaff and Seo (2010) investigated the differences between observed and simulated TBs and found large discrepancies due to deficiencies in cloud models. These deficiencies result in relationships between TBs and associated precipitation profiles that can be systematically different from those suggested by the TMI and PR observations. In brief, the databases exclusively based on the cloud model have statistical properties that do not exist in nature and, conversely, they do not have those properties that exist in the observations.

On the other hand, radiation–precipitation relationships derived exclusively from observations (e.g., TMI and PR) are incomplete from the microphysical perspective and potentially biased as well. Indeed, any given radar beam (with ~4-km horizontal resolution) observes a single vertical profile of reflectivity factors, which is not enough to restore the spatially varying vertical profiles of hydrometeor properties (cloud, rain, snow, and graupel), thermodynamic variables, and surface conditions that are needed for a complete description of the radiometer observations. Because of this lack of a complete observation set, cloud models still have a role in the construction of “a priori” precipitation–radiation databases.

Starting with cloud model simulations, the optimal strategy to improve the accuracy of Bayesian retrievals is to restrict the cloud and precipitation profiles included in the a priori database only to those that are fully consistent with actual observations over the region of interest and, indeed, to enforce this consistency by adjusting the underlying microphysics. Several attempts have been made to improve cloud/rain profiles using TRMM observations (Haddad et al. 1997; Grecu et al. 2004; Masunaga and Kummerow 2005; Grecu et al. 2011; Munchak and Kummerow 2011; Kummerow et al. 2011, 2015).

Indeed, one could make indirect use of radar and radiometer measurements of clouds by restricting the simulated cloud and precipitation profiles only to those actually observed, that is, producing radar and radiometer observations similar to the actual observations as performed in previous studies. Although radar reflectivity and radiometer TBs are integral measurements, they are sensitive to detailed information on the vertical structure of clouds and precipitation. To take advantage of these observations, in this study we use physical models to simulate radar reflectivity and microwave TBs as a function of a small number of independent variables and parameterizations whose a priori values are the values they have in the original cloud-resolving model simulations, and whose final value will be derived to maximize consistency with observations. A distinct feature of this approach is that it optimally utilizes the multilayer and multivariable relations among various hydrometeor contents that are produced by cloud-resolving model simulations, since those relations cannot be obtained in observational datasets. The independent variables are determined to minimize the disagreement between the simulated radar/radiometer observations and the actual TRMM observations by reducing efficiently the dimension of unknown variables to be retrieved and
circumventing many unrealistic assumptions. Our approach is similar in principle to that of Boukabara et al. (2011)—that is, it derives a solution by minimizing the differences between the physical-based simulations and the observations—but it differs from Boukabara et al. (2011) in the fact that here spaceborne radar reflectivity observations are also included in the minimization. In addition, the background information used in our approach is derived using high-resolution cloud-resolving simulations and no numerical weather prediction (NWP) data are required in the retrievals. The high-resolution cloud-resolving model simulations are used to condition the retrievals and eliminate the need for ad hoc assumptions and parameterizations to make the retrieval problem well posed [see Grecu and Anagnostou (2002) for examples of ad hoc parameterizations]. Precipitation–radiation databases derived using this approach are expected to be superior to those derived exclusively from either simulations or retrievals because these new databases are constrained to accurately represent observations while incorporating the physical consistency of cloud-resolving models. As illustrated by Sanò et al. (2013) and Casella et al. (2013), the accuracy of radiometer retrievals can improve through the selective use [i.e., as a function of dynamical–thermodynamical–hydrological (DTH) geographical–seasonal (GS) factors] of the underlying cloud-radiation database. In our approach, we mitigate the impact of GS variability on the quality of the derived cloud-radiation database by focusing exclusively on a specific region and season, that is, the seas and oceans east of Asia during the summer season. Given the use of vertical reflectivity profiles in derivation of the cloud-radiation database, we do not expect that the DTH dimension is as important as the GS dimension in the current study.

To quantify the variability of precipitation over the oceans east of Asia, a contoured frequency by altitude diagram (CFAD; Yuter and Houze 1995) analysis was conducted for the PR radar reflectivity observed in the summer season during five consecutive years. The analysis indicates that, on average, the oceanic region near the Korean Peninsula is characterized by stronger and deeper convective clouds in the vertical and a more intense rain rate than in the tropical and subtropical oceans off East Asia, although the latter regions are located in warmer environments (e.g., Seo 2011; Seo et al. 2015). Also, shallow convective clouds producing very strong rain intensity are frequently observed near the Korean Peninsula (e.g., Seo 2009, 2011; Ryu et al. 2012; Sohn et al. 2013; Song and Sohn 2015), while they are rather unusual over the continental United States. From this perspective, rainfall systems around the Korean Peninsula appear to be quite distinct in terms of cloud dynamics and microphysics. As such, a globally optimized database, like the one used in the standard GPROF, is unlikely to perform optimally in retrieving the hydrometeor profiles associated with precipitation systems around the Korean Peninsula. Therefore, developing a database optimized for the oceans of East Asia is of paramount importance in regional applications, although the method described in this study may be similarly applied to other regions as well.

The approach consists of several steps (Fig. 1). In step 1, we filter out the unrealistic model-generated profiles and restrict the database of simulated cloud and precipitation profiles to only those that agree reasonably well with the observations. Step 2 is to reduce the number of unknowns in order to perform the optimization that otherwise would be mathematically ill posed. After being represented in a reduced dimensionality, the preselected (deemed realistic) simulated clouds are optimized against PR and TMI observations in several steps, using sequential termwise optimization procedures (steps 3–6). In step 3 the optimization is performed against the radar-only observations, at the higher spatial resolution of these observations. In this step of the optimization, we perform scaling of the hydrometeor contents and modification of the drop size distribution by allowing adjustments in the $N_{dw}$ (the intercept of the exponential raindrop size distribution), often increasing the Marshall–Palmer $N_{d}$ (equivalent to decreasing the mean drop size diameter) in agreement with other studies. Next, in step 4, we use microwave radiometer simulators (radiative transfer models) to compute the passive microwave signatures (brightness temperatures) of the radar-optimized hydrometeor profiles, still at the higher spatial resolution of the radar observations. Step 5 convolves the high-resolution simulated brightness temperatures to the scale of the TMI observations. The final step, step 6, optimizes the simulated profiles, now with respect to the passive microwave observations. The end result is the generation of an optimal precipitation radiation database.

The TRMM observations, cloud-resolving model simulations, and radiometer and radar simulators are described in section 2. The retrievals of geophysical variables in clear sky are explained in section 3. Section 4 describes how we preselect simulated clouds that agree with PR observations, before the minimization is performed. Sections 5 and 6 describe how we perform the minimization of the difference between TRMM observations and simulations in the reduced dimension of the hydrometeor content profiles by adopting the multivariate relations among hydrometeor species in simulated clouds. Last, the conclusions are presented in section 7.
2. Data and simulators used in this study

The primary data for this study are TRMM observational data (TMI and PR) and the outputs of cloud-resolving model simulations of geophysical fields, as well as the corresponding synthetic observations that are produced by instrument simulators. The datasets and simulators are briefly explained in this section. They are the key ingredients for minimizing the disagreement between the simulations and the actual TRMM observations.

a. TRMM observational data

Datasets used in this study are TMI TBs (1B11), PR reflectivity profiles (2A25), and rain types (2A23) of version 7 of the TRMM products. TMI is a conically scanning radiometer, operating at a number of frequencies and polarizations. The PR is a cross-track scanning radar operating at 13.8 GHz (Iguchi and Meneghini 1994). Its sensitivity is about 17 dBZ, which corresponds to about 0.7 mm h\(^{-1}\) in rain rate. PR pixels have a horizontal resolution of 4.3 km at nadir and a vertical resolution of 250 m from the earth’s surface to 20-km altitude. Radar reflectivity observations were corrected for attenuation using the Hitschfeld–Bordan formulation and an alpha-adjustment technique to incorporate information from the surface reference method into the correction (Iguchi and Meneghini 1994; Iguchi et al. 2000). Radar-observed rain profiles are classified into three types: stratiform, convective, and others (Awaka et al. 1998).

For every TMI pixel, the collocated PR pixels are found over ocean only. In the collocation process, we consider only the subset of TMI observations consisting of the 12 TMI pixels around the center of the TMI swath (plus/minus six pixels on each side of the swath’s center), which facilitates a better geometrical match between the two types of measurements. For each of these TMI pixels, we select only those PR pixels that fall within the 19-GHz TMI field of view (FOV) and convolve their corresponding PR rain rates using the 19-GHz antenna gain function to represent the averaged PR rain rate over the TMI footprint scale. Only TMI pixels that have rain intensity greater than 0.05 mm h\(^{-1}\) (either by PR or TMI) are considered in the following optimization procedure. The collocated TMI–PR observations are used as input in the construction of the precipitation radiation database using the methodology outlined in
the introduction. This procedure was applied to all the data obtained during the summer months of June–August between 2010 and 2012 over the oceans off the East Asia region (28°–38°N, 110°–140°E).

b. Simulators to construct the a priori database

1) CLOUD-RESOLVING MODEL

The WRF Model, version 3.5, is used to simulate clouds and precipitation profiles. This study employs the new Kain–Fritsch cumulus parameterization scheme (Kain 2004), the Yonsei University (YSU) planetary boundary layer (PBL) (Hong and Pan 1996; Hong 2010), a simple cloud-interactive radiation scheme (Dudhia 1989), and the Rapid Radiative Transfer Model (RRTM) longwave radiation scheme (Mlawer et al. 1997). The WRF single-moment 6-class microphysics scheme (WSM6) is utilized. Graupel and associated processes are explicitly simulated in this scheme using parameterizations based on the work of Lin et al. (1983) and Rutledge and Hobbs (1984). The prognostic water substance variables include the mixing ratios of water vapor, cloud liquid water, rain, cloud ice, snow, and graupel.

Using the WRF Model, we simulated four summer rainfall systems, including two stationary frontal cases (one typhoon case and one light rainfall case) that are representative of summer rainfall around the Korean Peninsula and in East Asia. A 24-h forecast was produced for each case. Initial and boundary conditions were derived from the National Centers for Environmental Prediction Final Analysis data (Kanamitsu et al. 2002). All experiments involved one-way interactive triple-nested domains with a Lambert conformal map projection. The finest grid domain had a resolution of 2 km and was nested in a 6-km-resolution domain, which in turn was nested in an 18-km-resolution domain. All precipitation points in the 2-km gridded dataset were classified into either convective or stratiform following the procedure of Tao and Simpson (1993).

2) MICROWAVE RADIOMETER AND RADAR SIMULATORS

Radiometer and radar simulators were used to convert geophysical fields generated by cloud models into radiometric variables, that is, TBs and radar reflectivity factors. The two simulators use the same vertical discretization and are driven by the same input. Specifically, the atmosphere is divided into 40 layers, and each layer is 500 m thick. Within each layer, hydrometeors and relative humidity are assumed to be constant, while the temperature varies linearly with height. To ensure consistency with the WRF microphysical scheme, five categories of hydrometeors are assumed for the radiative transfer calculations, that is, cloud liquid water, cloud ice, rain, snow, and graupel particles. Cloud ice and snow particles can be either spheres with specified density or nonspherical particles given by Liu (2008). For the nonspherical ice particles, the single-scattering properties are calculated based on the discrete dipole approximation (DDA; Draine and Flatau 1994) assuming random orientation. The absorption by atmospheric gases is computed using the line-by-line method based on the work of Rosenkranz (1998) and Schwartz (1998). The refractive indices are based on Liebe et al. (1991, 1993) for liquid water and Mätzler (2006) for ice water.

The radar simulator produces the equivalent radar reflectivity factor (Ze) along with two-way path-integrated attenuation (PIA) at each layer of the atmosphere. Factor Ze is calculated from the backscattering cross sections of the five categories of hydrometeors using either Mie theory for spheres or DDA for nonspherical ice particles. PIA is calculated from the extinction efficiencies of atmospheric gases and hydrometeors.

To calculate the brightness temperatures, the radiative transfer equation is solved using a four-stream discrete ordinate method. Surface emissivity over ocean is calculated based on the emissivity model of Guillou et al. (1998) with modification by surface wind as suggested by Schlussel and Luthardt (1991). This radiative transfer simulation model was initially developed by Liu (1998) and later extended to allow a user to implement various particle shapes and size distributions. The accuracy of this model was analyzed by comparing its output to that from a polarized 32-stream discrete ordinate model with Mie scattering phase function and to the output from an unpolarized 32-stream discrete ordinate model with Henyey–Greenstein scattering phase function. The analysis showed that the four-stream model is fast and accurate for practical use in microwave remote sensing and data assimilation. Specifically, the differences between this model and the more accurate polarized 32-stream model were found to be less than 3 K for a broad range of microwave frequencies used in currently available satellite sensors. The model was also analyzed relative to a two-stream Eddington model. Results showed that it is more accurate than the Eddington model, especially for ice cloud conditions (Liu 1998).

3. Retrievals of geophysical variables in clear sky

To correctly quantify the relationship between observed TBs and precipitation over oceans, information regarding other geophysical variables, such as the sea
surface temperature (SST), the surface wind, and the total precipitable water (TPW), is necessary. The estimation of these geophysical variables inside precipitating regions is difficult and subject to large uncertainties. However, since these variables are more slowly varying spatially than precipitation, it is beneficial to estimate them in the surrounding clear-sky regions and to interpolate them into rainy regions (e.g., Elsässer and Kummerow 2008; Kummerow et al. 2011; Hristova-Veleva et al. 2013). Considering that thermodynamical and dynamical environments are different between clear-sky and rainy regions, this approach certainly has limitations. Although not error free, it seems to work satisfactorily in the derivation of global precipitation radiation databases directly from observations (Kummerow et al. 2011). In the following paragraphs, we explain how this methodology has been implemented in this study, which is focused on the construction of optimal precipitation radiation databases over the oceans of East Asia.

a. Creation of a priori database

To derive geophysical variables in clear-sky (non-rainy) conditions, we use the simulation data for the four cases described in the previous sections. We select only clear skies (with zero condensed water mass) from all the simulations, and we prepare a clear-sky database consisting of geophysical variables (SST, surface wind, and TPW) and the associated TBs at the seven TMI frequency/polarization combinations (10H, 10V, 19H, 19V, 21V, 37H, and 37V GHz). Since, in general, TBs at frequencies equal or higher than 85 GHz are sensitive to optically thin clouds, they are not used in clear-sky retrievals.

As in other previous studies (e.g., Kummerow et al. 2011, 2015), the database is not stratified by season or storm morphology. This study includes only observations from a single season (namely, summer) and a single region. Therefore, the combination of TMI TBs at the seven channels is expected to appropriately characterize the variability of geophysical variables in the region and season of interest.

b. Clustering of clear-sky TBs

To efficiently retrieve clear-sky geophysical parameters from observed TBs, we use a clustering-based database. Specifically, the a priori database of clear-sky brightness temperatures and associated geophysical parameters is partitioned into a relatively small number of clusters. The k-means clustering algorithm, which is one of the simplest unsupervised learning algorithms (MacQueen 1967), is used. The algorithm partitions n observations into k clusters based on their similarity.

In our study, the large set of simulated TBs is grouped into a significantly smaller number of classes based on their similarity in the seven-dimensional TB space. Details regarding the particular k-means TB space implementation we used can be found in de Hoon et al. (2004). We used the k-means clustering procedure to group about two million clear-sky TBs into 2000 clusters. For each cluster, the mean of TBs and their corresponding geophysical variables are computed. The 2000 clusters derived by the clustering algorithm are likely to capture well the distribution of clear-sky TBs over East Asia seas.

Although practically the clusters are grouped in the seven-dimensional TB space, Fig. 2 shows the distribution of all the clusters in the space of SST versus TPW, SST versus surface wind space, and TPW versus surface wind for the sake of convenience. Each circle in the figure represents a cluster and its circle size represents the number of data within the cluster. The mean data number for a cluster is ~180. The mean distance among all cluster centroids (i.e., from a cluster centroid to a cluster centroid) is about 3.5 K. The mean distance is the mean of the distances in the seven TMI-frequency dimensions among the centroids of all the clusters. The small values in mean distance imply that since the original TB datasets are grouped relatively closely to each other, some subtle changes in their seven-dimensional TBs are likely to be well captured by the TB clusters. The clusters appear to significantly overlap in the two-dimensional space. However, the overlap might be considerably alleviated in the original seven-dimensional TB space. On the other hand, a smaller number of clusters are likely to significantly increase the random errors in clear-sky geophysical retrievals.

c. Retrievals of geophysical variables from TMI TBs

The clustering algorithm compresses the initial database of simulated clear-sky TBs into a smaller database containing only the mean properties of each cluster. To retrieve clear-sky geophysical parameters from TMI observations, we simply compute the Euclidean distances between observations and each of the 2000 clusters in the compressed database. We then determine the cluster with the smallest distance and extract its corresponding geophysical variables linked to the cluster. The retrieved clear-sky geophysical variables are further interpolated into rain areas and used in the optimization process for the simulation of TB in rain areas.

4. Preselection of simulated clouds

In general, simulated cloud structures are known to be quite different from observations, especially from the
perspective of vertical PR structures and 85-GHz TMI TBs (Biggerstaff and Seo 2010; Seo and Biggerstaff 2012). To minimize the negative impact that cloud-resolving model deficiencies might have on the construction of the precipitation radiation database, only simulated precipitation profiles whose corresponding radar reflectivity profile is similar to actual PR observations need to be considered. The similarity between the $i$th simulated cloud radar reflectivity profile ($Z_s^i$) and the $j$th observed radar reflectivity profile ($Z_o^j$) is evaluated using the Euclidean distance as

$$d_{ij} = \frac{|Z_s^i - Z_o^j|}{|Z_s^i|} \times 100 \quad \text{for} \quad j = 1, \ldots, n,$$

where $n$ denotes the number of observed radar reflectivity profiles. It should be noted that the radar reflectivity vectors in Eq. (1) are taken into account for only above-2-km altitudes, since PR reflectivities below this level are affected by the surface returns. As evident in Eq. (1), if $d_{ij}$ is small, the profiles $Z_s^i$ and $Z_o^j$ are similar.

To evaluate the realism of a simulated cloud and precipitation profile, we use Eq. (1) to compare the associated simulated reflectivity profile with a large number of observed PR reflectivity profiles collected in the domain of interest. The selection of the simulated clouds is shown as step 1 in the flowchart (Fig. 1) that outlines our approach from starting with the cloud simulations to the generation of the optimized cloud-radiation database at the end. A simulated reflectivity profile $Z_s^i$ is considered similar to one of the observed profiles if the minimum distance $d_{ij}$ is smaller than 40% for convective rain and 10% for stratiform rain. Since close similarity between simulated and observed reflectivity profiles is rarely found for convective rain, the two different thresholds are employed. The cloud-resolving model seems to have more difficulty in reproducing the observed variability of vertical cloud structure in convective rain than in stratiform rain. It should be noted that due to the PR’s minimum detectable reflectivity ($\sim 17 \text{dBZ}$), only a subset of simulated cloud profiles are considered in the matchup process—only those profiles for which the maximum reflectivity in the column above-2-km altitude is greater than 17 dBZ. Meanwhile, it is intended that simulated reflectivity profiles with echoes below 17 dBZ are not subject to the similarity evaluation process.

Fig. 2. Distribution of the 2000 clusters in the space of (a) SST and TPW, (b) SST and surface wind, and (c) TPW and surface wind. Each circle in shades of gray represents a cluster. The data size of the reference circle in the upper-left corner of each panel is 180. The circle size is proportional to the data size of each cluster.
About 55% (80%) of the simulated convective (stratiform) cloud profiles exhibit structural differences smaller than 40% (10%). The simulated cloud profiles exhibiting structural differences smaller than 40% are considered more realistic and further used in the database construction for the convective rain type. The CFADs of their radar reflectivity profiles and corresponding observational PR profiles are shown in Figs. 3a and 3c, respectively. The two CFADs appear to be in a good agreement except for the area of low radar reflectivity ($Z < 17$ dBZ), which is not detected by the PR instrument. Regarding the simulated cloud profiles whose differences are greater than 40%, high reflectivities at upper levels are apparent (Fig. 3b). The closest observational clouds to these profiles look quite different from the simulated clouds (cf. Figs. 3b and 3d). These simulated vertical cloud profiles (Fig. 3b) are unlikely to occur in nature. The joint distribution of the simulated TBs at 85 and 37 GHz at PR resolution for both cases (i.e., min $d_{ij} < 40\%$ and min $d_{ij} \geq 40\%$) are shown in Figs. 3e and 3f. The profiles that do not closely resemble the PR observations appear to be extended to fairly low TBs at both frequencies, compared to the other profiles. These low TBs are in poor agreement with observed TBs (e.g., Seo and Biggerstaff 2006). Therefore, the evaluation of simulated cloud profiles based on the radar observations is consistent with their evaluation from a TB perspective, and it provides an effective and simple way to filter unrealistic simulated profiles characterized by low TBs at high frequencies.
For stratiform rain, the poorly matched simulated profiles exhibit unusual joint distributions of frozen hydrometeor and rain (Fig. 4); that is, these clouds exhibit a relatively moderate rain layer with weak ice layer, compared to well-matched clouds. This is also apparent in the two-dimensional TB relationships shown in Fig. 4f. Relatively medium to high 37- and high 85-GHz TBs in Fig. 4f are removed based on this selection process, which corresponds to clouds having week to moderate rain layers with possibly absent or weak ice layers. All this suggests that the PR consistency analysis is capable of depicting well-defined structures that are present only in the simulations but are absent in observations. Hence, the simulated hydrometeor profiles and their associated TBs need to be refined to better reproduce the

![Fig. 4. As in Fig. 3, but for stratiform rain clouds.](image)

5. Optimization of simulated clouds using high-resolution synthetic observations

In section 4, an analysis was performed to identify simulated precipitation profiles similar to actual PR observations. However, the CFAD analysis of the simulated profiles deemed to be similar to PR observations still reveals discrepancies relative to observations. Furthermore, this analysis does not guarantee that all precipitation profiles occurring in nature have a close match in the simulated profiles. Hence, the simulated hydrometeor profiles and their associated TBs need to be refined to better reproduce the
joint precipitation–TB distribution of observed clouds over the oceans near East Asia.

Given the lack of in situ observations, we resort to indirect and integral measurements (such as radar reflectivity and passive microwave radiances), and cloud-resolving model simulations to create optimal precipitation radiation database. The simulated clouds deemed realistic by the analysis described in the previous section are used as input to an optimal estimation procedure that is designed to minimize the disagreement between the measurements and simulations for both radar reflectivity and TB by adjusting the water content in each hydrometeor species. In this way, we can adjust the synthetic cloud database, which is done in several steps (steps 3–6 in Fig. 1). The adjusted cloud database is expected to be superior to those that are derived exclusively from simulations or from observations alone.

a. Reduction of dimensionality and defining the optimization approach

The squared difference between the measurements and simulations for radar reflectivity and TBs can be defined as a cost (or objective) function ($\Psi$). The long vector ($Q_j$) for the $j$th cloud profile contains vertical content profiles of all hydrometeor species, such as cloud liquid water ($q_{j}^{l}$), rain ($q_{j}^{r}$), cloud ice water ($q_{j}^{i}$), snow ($q_{j}^{s}$), and graupel ($q_{j}^{g}$) vectors as its components. Thus, the long vector is defined as

$$Q_j = \begin{bmatrix} q_{j}^{l} \\ q_{j}^{r} \\ q_{j}^{i} \\ q_{j}^{s} \\ q_{j}^{g} \end{bmatrix}. \quad (2)$$

Function $\Psi$ (which is a convex function) is minimized as a function of $Q_j$ through a gradient descent technique; that is, we change $Q_j$ (the tentatively optimal cloud content profile) infinitesimally in the direction determined by the gradient of the cost function until a minimum point is reached.

The dimension of $Q$ is the number of hydrometeor species times the number of vertical levels in the content profile of a specific hydrometeor species. For example, for five different hydrometeor species at $N$ levels, the number of unknowns in $Q$ is $5N$. On the other hand, if we have observations of radar reflectivity at $N$ levels and TBs at nine frequency channels, then the number of knowns is $N + 9$. Thus, the number of unknowns is larger than the number of knowns (equations) for any $N > 2$, and so the associated mathematical problem becomes ill posed. To reduce the number of unknowns instead of the full vector $Q$, a reduced vector, including a single ice species, and a low-order ice parameterization have been considered in several past studies (e.g., Grecu and Anagnostou 2002; Grecu et al. 2004; Masunaga and Kummerow 2005; Munchak and Kummerow 2011).

In this study, instead of using ad hoc ice parameterization to reduce the dimension of $Q$ in an effective way, we will make use of an empirical orthogonal function (EOF) decomposition; that is, the long vector for the $j$th data point is the sum of the mean vector $\bar{Q}$ and the deviation vector $Q'_j$. If the deviation vector is decomposed with EOFs, then the long vector can be written as

$$Q_j = \bar{Q} + Q'_j = \bar{Q} + \sum_{k=1}^{N} \alpha_{jk} \hat{e}_k. \quad (3)$$

where $\hat{e}_k$ and $\alpha_{jk}$ denote the $k$th EOF and the EOF coefficient, respectively, at the $j$th data point, and $N$ denotes the number of EOFs for the appropriate expression of $Q'_j$. The EOF structures incorporate the intrinsic multilayered and multivariate relations among hydrometer species. Since there are insufficient in situ observations to accurately characterize the observed $Q$ using a reduced number of variables, we will rely on the EOF structures of simulated hydrometeor contents derived from cloud-resolving model simulations (e.g., Seo et al. 2007). The EOF decomposition is marked as step 2 in the flow diagram in Fig. 1.

Figure 5 shows the mean and standard deviation of the long vector and two major EOF structures for convective and stratiform, respectively, for the dataset of the selected rain clouds. For convective rain clouds, four major EOFs can explain more than 90% of the total variance in the hydrometeor profiles. The first EOF with the eigenvalue of 62.3% shows that rain and graupel have a tight linkage and also a positive connection with other hydrometeor contents; that is, all the hydrometeor contents increase or decrease simultaneously in the same direction. The second EOF (eigenvalue of 15.2%) suggests additional rain and graupel processes. It represents a rain content increase at lower levels in conjunction with a graupel content decrease at upper levels or vice versa; that is, 1) rain content can increase (or decrease) by the consumption (deficit) of graupel and 2) the shallow convection without an ice layer can be interpreted as an anomaly. For stratiform rain clouds, on the other hand, more EOFs are needed to explain about 90% of the total variance, showing larger variability in hydrometeor profiles. This can be attributed to more complex vertical structures in the stratiform rain clouds, where mesoscale downdraft (updraft) exists underneath (above) the
FIG. 5. Mean (first column) and standard deviation (second column) of hydrometeor content profiles and the two dominant EOF structures (first and second EOFs in the third and fourth columns, respectively) for hydrometeor content profiles for convective (first row) and stratiform (second row) rain clouds. The solid and dotted lines denote mean and standard deviations, respectively. The terms, $q_c$, $q_r$, $q_i$, $q_s$, and $q_g$, denote cloud water, rain, cloud ice, snow, and graupel content (g m$^{-3}$), respectively.
freezing level, resulting in quite different microphysical processes than in the convective clouds. The mean profile indicates that snow becomes the most relevant frozen hydrometeor species. In addition, snow becomes important in explaining the vertical hydrometeor variability, unlike in convective rain clouds, where the rain and graupel hydrometeor species are the most relevant variables.

Once the EOFs are determined, $a_{jk}$ in Eq. (3) are the only variables needed to fully reconstruct $Q$. In general, because only several major EOFs can explain more than 95% of the total variance of $Q$, significant dimensionality reduction can be achieved using EOF coefficients instead of $Q$. If we use only $p$ EOF coefficients to represent $Q$, then a vector $X$ containing the coefficients is expressed as follows:

$$X = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}.$$  \hspace{1cm} (4)

After having reduced the dimensionality of the problem, we are now ready to proceed with the optimization. The general cost function can be evaluated as a function of TB and radar reflectivity observations as follows:

$$\Psi(X) = [Z' - Z''(X)]^T O_Z^{-1} [Z' - Z''(X)] + [TB' - TB''(X)]^T O_{TB}^{-1} [TB' - TB''(X)].$$  \hspace{1cm} (5)

The variables in the above-mentioned equation are described in Table 1. Although the number of unknowns in the cost function is highly reduced from $5N$ to $p$ (generally from $5 \times 40$ to less than 20), $X$ accurately describes the original hydrometeor structure of $Q$. Thus, Eq. (5) constitutes a one-dimensional variational (1DVAR) optimization estimation combined with the dimensional reduction using EOFs—referred to as a hybrid EOF–1DVAR approach. In this approach, we assume that surface, thermodynamic atmospheric conditions, and water vapor are known, and that the errors in radar reflectivity and TB are associated only with hydrometeor content profiles derived from the first $p$ EOFs. The error covariances are calculated based on the WRF simulations with these assumptions. Specifically, the known variables are set to their average values and the first $p$ EOFs are used to simulate reflectivity profiles and TBs. The simulated reflectivity and TBs are compared to those derived by using the actual surface and atmospheric conditions and water vapor and hydrometeor contents, and then the error covariance vectors $O_Z$ and $O_{TB}$ are computed. Overall, it is found that the major EOF modes can reproduce very accurately their original hydrometeor content profiles. Since a larger $p$ makes the gradient of the cost function more expensive, we need to choose a reasonable number of modes in dealing with both the restoration of the long vectors and the computationally intensive minimization of the cost function. Considering both factors, 16 EOFs ($p = 16$) for the convective and stratiform rain database were found to be adequate.

**b. Evaluation of the hybrid EOF–1DVAR approach using the synthetic data**

The performance of the optimization approach based on the minimization of Eq. (5) can be evaluated using synthetic data. Specifically, we used the full hydrometeor information in the $Q$ vectors and produced simulated reflectivity and TBs from the WRF simulations described in section 2. These synthetic data were considered the “truth” and were used as input in Eq. (5). Only profiles that resemble actual PR observations are considered. These “true” values of the precipitation profiles are compared with the EOF–1DVAR retrievals. Although the objective function is not separable (i.e., it cannot be written as the sum of multiple functions that can be optimized independently), we found that a sequential termwise optimization works better than a global optimization. Specifically, in the first step (step 3 in Fig. 1), $X$ is estimated exclusively based on the radar observations, that is, by minimizing

### Table 1. Description of the variables in the cost function.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Psi, \Phi$</td>
<td>Cost function</td>
</tr>
<tr>
<td>$Q$</td>
<td>A vector containing hydrometeor contents</td>
</tr>
<tr>
<td>$Z'$</td>
<td>A vector containing observed radar reflectivity in the vertical</td>
</tr>
<tr>
<td>$Z''$</td>
<td>A radar model operator that generates a vector containing synthetic radar reflectivity in the vertical, for a given $Q$</td>
</tr>
<tr>
<td>$TB'$</td>
<td>A vector containing observed TBs at TMI frequencies</td>
</tr>
<tr>
<td>$TB''$</td>
<td>A TB model operator that generates a vector containing simulated TBs at TMI frequencies, for a given $Q$</td>
</tr>
<tr>
<td>$O_Z$</td>
<td>An error covariance vector for radar reflectivity</td>
</tr>
<tr>
<td>$O_{TB}$</td>
<td>An error covariance vector for TBs at TMI frequencies</td>
</tr>
</tbody>
</table>

Unauthenticated | Downloaded 04/06/24 11:38 AM UTC
\[\Psi(X) = \left[ Z_{\text{Ka}}^o - Z_{\text{Ka}}^m(X) \right]^T O_{Z(\text{Ka})}^{-1} \left[ Z_{\text{Ka}}^o - Z_{\text{Ka}}^m(X) \right] + \left[ \text{PIA}_{\text{Ka}}^o - \text{PIA}_{\text{Ka}}^m(X) \right]^T O_{\text{PIA(Ka)}}^{-1} \left[ \text{PIA}_{\text{Ka}}^o - \text{PIA}_{\text{Ka}}^m(X) \right] + \left[ Z_{\text{Ka}}^o - Z_{\text{Ka}}^m(X) \right]^T O_{Z(\text{Ka})}^{-1} \left[ Z_{\text{Ka}}^o - Z_{\text{Ka}}^m(X) \right] + \left[ \text{PIA}_{\text{Ka}}^o - \text{PIA}_{\text{Ka}}^m(X) \right]^T O_{\text{PIA(Ka)}}^{-1} \left[ \text{PIA}_{\text{Ka}}^o - \text{PIA}_{\text{Ka}}^m(X) \right] + \sum_{j=1}^{m} \left( \chi_j^2 / \sigma_j \right), \]

where \(x_j\) and \(\sigma_j\) represent the \(j\)th EOF coefficient in \(X\) and the variance of \(x_j\), respectively, and \(\text{PIA}\) is the path-integrated attenuation. Although real Ka-band observations are not used in this study, whose focus is on TRMM observations, the retrieval methodology can process Ka-band observations through the inclusion of Ka-band observations in the objective function [Eq. (6)] for GPM. In addition, the \(\text{PIA}\) terms in Eq. (6) are not used, since corrected radar reflectivity of PR was employed. A fifth term has been introduced on the right-hand side of Eq. (6) to prevent the estimate of an unrealistic (noisy) EOF coefficient. This term is usually referred to as a regularization term. Regularization is also beneficial in the fact that it makes the cost function more convex and faster converging to its optimal value.

The steepest gradient method (e.g., Fletcher and Powell 1963; Shewchuk 1994) and a finite difference method to evaluate the derivative of the cost function are used to minimize the objective function. Once \(\Psi(X)\) is minimized, its solution is used as an initial guess of the following cost function \(\Phi(X)\) in association with TB observations at radiometer frequencies:

\[\Phi(X) = \left[ \text{tb}^o - \text{tb}^m(X) \right]^T O_{\text{TH}}^{-1} \left[ \text{tb}^o - \text{tb}^m(X) \right]. \quad (7)\]

After minimizing \(\Phi(X)\) using the gradient method, the final \(X\) is used to compute the estimated \(\text{Q}\). Its corresponding radar reflectivity (\(Z^m\)) and TBs (\(\text{tb}^m\)) are compared to their respective truth—the synthetic \(Z^o\) and \(\text{tb}^o\) computed from the full \(\text{Q}\) vector. Note that here the optimization is performed assuming that the radar and the radiometer have the same resolution.

The approach to perform the optimization at the resolutions of the satellite observations is outlined in the following section (section 6). In this synthetic data evaluation, we used the radar and radiometer observations at the model grid scale. To set up the stage for a GPM-based optimization, we used here synthetic data at the DPR and GMI frequencies. In particular, we performed a three-stage optimization using at the first stage synthetic radar data at Ku band, followed by a second stage optimization that used Ka-band radar data. The third stage performed the optimization using the synthetic brightness temperatures. The Ka-band optimization is deemed beneficial, as the Ka-band radar data carry additional information (with respect to the Ku-band observations) on the vertical structure of the cloud and precipitation fields.

A CFAD analysis indicates that the retrieved radar reflectivity is in good agreement with the observations used in their derivation (Fig. 6). In particular, the modes (the most frequent values) of the height-dependent retrieved reflectivity distributions are the same as those of the synthetic reflectivity observations. The structural differences between the mean true and retrieved radar reflectivity profiles are only 6.4% and 4.8% for the Ku-band and Ka-band observations, respectively. A comparison of observed and retrieved TBs shows good agreement for the lower-frequency channels, while there is more scatter in the comparison of the high-frequency channels. Still, even for these channels the majority of points fall along the one-to-one line. Thus, it can be concluded that the hybrid EOF–1DVAR is effective in retrieving multiple species of hydrometeor profiles from radar and radiometer observations. Nevertheless, larger uncertainty is expected in satellite applications involving real data.

6. Optimization of the radiation-cloud databases using TRMM observations

In the previous section, the performance of the hybrid EOF–1DVAR is investigated at model grid scale using synthetic data. Here we apply the retrieval methodology to actual PR and TMI observations. Unlike the synthetic observations investigated in the previous section, actual TRMM observations are characterized by multiple resolutions (Fig. 7).

Specifically, the TMI footprint sizes differ from channel to channel and are larger than PR footprint sizes. The retrieval methodology developed in the previous section needs to be extended to account for these difficulties. The flowchart that describes the hybrid EOF–1DVAR optimization procedures for the cases of satellite observations is shown in Fig. 1.

a. EOF–1DVAR at satellite resolution

For the retrieved precipitation profiles from multiple resolutions of PR and TMI observations, we resort to the same termwise methodology (radar-based optimization followed by a radiometer-based one) used in the model resolution retrieval investigated in the previous section. Specifically, radar-based optimization using the formulation in Eq. (6) are performed
FIG. 6. CFADs for the synthetic radar reflectivity observations (first row) and the radar reflectivity of the hydrometeors obtained from the hybrid EOF–1DVAR approach (second row) in convective rain clouds in model grid scale. The solid and dotted lines represent mean profiles of the synthetic observations and the retrievals, respectively. The numbers in the panels denote structural differences between mean observations and retrievals. The comparison of the observed TBs (y axis) with the retrieved TBs (x axis) at low (third row) and high frequencies (fourth row). The numbers in panels denote bias and RMSE of the retrieved TBs. The lines in the first and second rows represent mean radar reflectivity profiles.
for each PR profile classified as rain (Awaka et al. 1997) within a given 10-GHz TMI footprint (see Fig. 7). It should be noted that the radar reflectivity vectors in Eq. (6) are also taken into account for only above-2-km altitudes due to the surface return reflectivity contamination. Up to this point the optimization approach follows the same steps (1–3, in Fig. 1) as described in section 5. At this point the satellite-based optimization diverges in order to account for the difference in the resolution between the PR and the TMI observations. In step 4 (Fig. 1) the radar-adjusted hydrometeor contents are then used as inputs to the radiative transfer model to compute high-resolution TBs and the TMI frequencies and polarizations. Next, in step 5, the TBs are convolved to the lower resolution of the TMI observations. The computed TB sets within a given TMI footprint are

$$\Phi(X) = [TB^o - TB^m(X)]^T O^{-1}_{TB} [TB^o - TB^m(X)], \quad (8)$$

where $TB^m$ represents the modeled TB vector at the TMI scale. The radiometer resolution $TB^m$ at the $j$th channel is derived from $tb^m_k$ ($k$ is from 1 to $n$, where $n$ denotes the number of the PR pixels within the TMI footprint.) at the PR resolution through a convolution process. The convolution weights $w_{jk}$ are derived based on the antenna gain pattern that is an instrument- and frequency-specific function of the distances between the radiometer footprint centers and the PR footprint centers. The simulated radiometer TB at the $j$th channel is then calculated as

$$TB^m_j(X) = \sum_{k=1}^n w_{jk} tb^m_k,$$ \quad (9)

where the subscript $k$ denotes a PR pixel and $w_{jk}$ represents the weighting factor from the antenna gain function at the $j$th TMI frequency.

Similar to the minimization problem in section 5, the PR-only retrievals based on Eq. (6) can significantly differ from their initial values derived from Eq. (6). This is because TBs alone do not provide enough information to keep the retrievals consistent within the range of possible solutions suggested by the radar observations. To circumvent this limitation, we introduce a set of scaling variables $r$ that relate the PR-only retrievals to the final retrieval in the following way:

$$Y_k(r) = r^o Q^o_k = r_c q^o_c + r_r q^o_r + r_q q^o_q + r_g q^o_g,$$ \quad (10)

where the five elements in $r$ denote the ratios for cloud liquid water, rain, cloud ice, snow, and graupel, respectively; and $Q^o_k$ is obtained from $X_k$, which is the solution of Eq. (6) with $p$ EOF modes for the $k$th PR pixel within the TMI footprint. Therefore, $Y_k(r)$ is a radiometer adjustment of $Q^o_k$ derived at PR resolution by adjusting a scaling variable $r$ while maintaining the vertical shape of $Q^o_k$. Therefore, the cost function that quantifies the agreement between the simulated and observed radiometer TBs given by Eq. (8) can be expressed as a function of $r$:

$$\Phi(r) = \{TB^o - TB^m[Y_k(r)]\}^T O^{-1}_{TB} \{TB^o - TB^m[Y_k(r)]\}$$

$$+ \sum_{j=1}^5 \left( \frac{r_j^2 - 1}{\sigma_j^2} \right),$$ \quad (11)

where $TB^m$ is given by Eq. (9) and the last term in Eq. (11) is a regularization term that prevents $r$ from taking unrealistic values. This is achieved by penalizing deviations of $r_j$ from $r$ that are significantly larger than their expected variances ($\sigma_j$, where $j$ denotes the $j$th hydrometeor species). The optimal hydrometeor profiles consistent with both PR and TMI observations are derived by minimizing $\Phi(r)$ with respect to $r$. Step 6 in Fig. 1 represents this final step.

A preliminary analysis revealed that the derivations of hydrometeor profiles based on the minimization of
the functions of Eqs. (6) and (11) do not always provide good agreement between observed and simulated reflectivity and TBs. Specifically, it was found that sometimes while simulated radiometer observations were in good agreement with the observations, the simulated reflectivities were not in good agreement with the observations. Converse situations—that is, good agreement in terms of reflectivity but poor agreement in terms of TBs—were also encountered. To prevent large adjustments in \( r \) that would result in poor agreement between simulated and observed reflectivity, we included the intercept \( N_0r \) of the retrieved rain drop size distribution (DSD) as a variable in the optimization framework of Eq. (6), as illustrated in step 3 in Fig. 1. We based our decision to adjust the intercept parameter on a recent study by Varble et al. (2011), in which they found that varying the intercept and shape parameters of a gamma distribution had much larger effects on radar reflectivity than varying the ice water content. In the optimization, we employed six different intercepts around the intercept \( N_0r \) used by the cloud-resolving model simulations. For each intercept candidate, the minimization of Eq. (6) was performed and then the corresponding cost was recorded. The radar-adjusted hydrometeor contents and intercept number with the minimum cost among the six costs were chosen for the input for Eq. (11). To improve the agreement between simulated and observed TBs without negatively affecting the agreement between radar simulations and observations, we devised a procedure to update the values of the DSD intercepts. Specifically, the DSD intercept adjustments were determined to keep the simulated reflectivities derived from the minimization of Eq. (6) constant while adjusting the rain contents as suggested by the gradients of Eq. (11). This procedure is similar to those used by Masunaga and Kummerow (2005) and Viltard et al. (2000). Its application resulted in significant improvement in the agreement between simulated and observed TBs. The \( N_0r \) adjustments consisted preponderantly in increases relative to the initial values, suggesting DSDs with a larger number of smaller particles than the initially assumed Marshall–Palmer distribution. Indeed, this is consistent with the microphysics of convective clouds near the Korean Peninsula during summer. This region exhibits strong rain rate more frequently than any other region for comparable radar observations (Seo 2011; Ryu et al. 2012; Sohn et al. 2013), corroborating our finding that drop size distributions that are dominated by smaller particles are more realistic, as they produce satellite observables that are in closer agreement with observations.

b. Comparison with the TRMM observations

The five hydrometeor species that are derived as a solution to the optimization problems described above are archived together with their corresponding computed and observed radar reflectivity and TBs. The rain rate, hydrometeor contents, and radar reflectivities at PR resolution are convolved using the approximate antenna gain function at 19 GHz. The computation of the satellite-scale brightness temperatures was described in section 6a [Eq. (9)] above.

As previous research has pointed out, convective precipitation has microphysical and thermodynamic characteristics that are very different from that in the stratiform regions. To analyze whether the newly constructed precipitation–radiation database represents equally well both of these major precipitation types, we separate the database into convective and stratiform types and evaluate them separately.

To identify the precipitation types of the lower-resolution radiometer's FOV, we use the convective (stratiform) areal fraction convF (stratF), which is defined as the ratio of the number of convective (stratiform) points (at PR resolution) to the total number of points within the nominal footprint scale. The procedure can be found in Seo et al. (2015). Based on these areal fractions, precipitating FOVs at the TMI footprint scale are classified into one of the following three categories:

\[
\text{category} = \begin{cases} 
1(\text{convective}), & \text{convF} \geq 0.5 \\
2(\text{stratiform}), & \text{stratF} \geq 0.5 \\
3(\text{mixed}), & \text{convF} < 0.5 \text{ and } \text{stratF} < 0.5.
\end{cases}
\]

The CFADs of the computed radar reflectivities and their corresponding observed PR reflectivities (which are convolved over the nominal TMI footprint) are shown in Fig. 8. Overall, the retrieved radar reflectivities are in a better agreement with the observed when both PR and TMI observations are used during the retrieval. This is an indication that including more observational information in the analysis and relaxing the conditions imposed by the principal component analysis are conducive to better results; that is, the inclusion of both PR and TB observations in the optimization leads to retrievals that match better the PR observations compared to the case when only PR observations were used. Hence, in real situations more varied constraints might be required to make the problem stable, since the reality is more complex than this synthetic environment.

In both convective and stratiform rain, the retrieved radar reflectivity exhibits small discontinuities near
freezing level. Because the forward models used in this study do not explicitly model melting particles and only a small number of EOF modes are utilized, larger uncertainties are expected in the retrievals in the mixed-phase layer. Above the freezing level, the EOF–1DVAR approach produces slightly larger radar reflectivity than the observations in both rain types. However, the frequency distribution of the retrieved radar reflectivity in Fig. 8 is much more similar to that of the observations than the CFADs of the preselected profiles in Figs. 3 and 4.

As previously described, the hydrometeors retrieved for every PR observation profile are used to compute TBs at the TMI frequency channels and convolved into TBs at the TMI footprint scale. The simulated TBs are directly compared with the observed TMI TBs. Results are shown in Fig. 9 in terms of scatterplots. To show the comparison at all TMI frequencies, the horizontal and vertical

![Image](https://example.com/image.png)
FIG. 9. The comparison of the retrieved TBs (y axis) with the observed TBs (x axis) at (a),(b) 10H, (c),(d) 19H, (e),(f) 37H, and (g),(h) 85H GHz in TMI footprint scale for (left) convective and (right) stratiform rain clouds.
polarizations are selected for convective and stratiform rain types, respectively. Overall, the simulated TBs are in good agreement with the observed TBs at the TMI frequencies in stratiform rain. In convective rain, a warm bias is apparent at 10 GHz, while a cold bias is apparent at 19 GHz. In all rain types, the TB root-mean-square (RMS) differences at 10V and 19V GHz and channels are 6.4 and 17.2 K, respectively. The correlations at the two frequencies are 0.91 and 0.85, respectively. The RMS differences are somewhat large, but the correlations are high. Since low frequencies are highly sensitive to surface conditions, some of the discrepancies at these frequencies are likely not caused only by uncertainties in precipitation retrievals. Nevertheless, most retrieved TBs at the low frequencies resemble closely the observed TBs. In all rain types, the TB RMS differences at the 37V and 85V channels are 15.0 and 15.8 K, respectively. The correlations are 0.57 and 0.80, respectively. The large differences can be attributed mainly to a small number of the simulated TBs that are distant from the one-to-one line. At 37 GHz, the convective region shows a cold bias, but the stratiform region exhibits a very good match between the retrieved and observed TBs. At 85 GHz, a very good agreement is found, particularly in stratiform rain. Such a good consensus at high frequencies cannot be accomplished without reasonable retrievals in frozen hydrometeors; that is, the EOF–1DVAR approach appears to retrieve a well-balanced distribution of snow and graupel and rain.

Figure 10 shows the mean profiles and CFADs of five hydrometeor species retrieved by the hybrid EOF–1DVAR approach using both PR and TMI observations. Especially in convective rain clouds, the mean graupel contents are considerably reduced compared to the mean rain contents (cf. Figs. 5 and 10). These results show the superiority of our methodology over cloud model–based approaches and mitigate deficiencies associated with the production of too much ice content by the models at upper levels. In particular, the retrieved graupel contents in stratiform rain clouds are about a half of the mean graupel contents in Fig. 5, while the retrieved snow contents remain largely unchanged, with an amount comparable to the mean snow content before the optimization. This result is consistent with the expectation that stratiform rain clouds produce abundant low-density ice-phased hydrometeors but small amounts of high-density ice-phased hydrometeors and it illustrates a major benefit of our approach.

7. Conclusions

Cloud profiles simulated by cloud-resolving models are frequently employed in building cloud-radiation databases that are then used in passive microwave rain retrieval algorithms, such as Bayesian retrievals. However, these model-produced profiles show dissimilarity from observed clouds when compared to vertical PR structures and microwave brightness temperatures. Hence, a critical step toward improving rain estimates from algorithms that use cloud-radiation databases is to bring agreement between simulated and observed clouds, at least as judged by the joint distribution of the associated microwave signatures.

Lacking in situ observations, we make indirect use of actual radar and radiometer measurements of clouds, which provide detailed information on their vertical structure. To take advantage of these observations, we use physical models to simulate radar reflectivity and microwave TBs as a function of a small number of independent variables and parameterizations derived from cloud-resolving model simulations. The independent variables are determined to minimize disagreement between the simulated radar and radiometer observations and the actual TRMM observations.

To do so, we developed an efficient and unique method (the so-called hybrid EOF–1DVAR approach). The approach consists of several steps (Fig. 1). First, we filtered out the unrealistic model-generated profiles and restricted the database of simulated cloud and precipitation profiles to only those that agree with the observations. We achieved that by investigating the similarity of simulated and observed radar reflectivity profiles for both convective and stratiform rain types.

The next step is to reduce the number of unknowns in order to perform the optimization that otherwise would be mathematically ill posed. Past studies have addressed this problem by sometimes making strong assumptions such as removing some hydrometeor species and using a single ice category or using a low-order ice parameterization with simple relations among the hydrometeor species. Instead, we reduce the dimension of the hydrometeor content profiles by adopting the multivariate relations (EOF analysis) among hydrometeor species in simulated clouds, thus efficiently alleviating the mathematically ill-posed problem.

In the next couple of steps, the preselected (deemed realistic) simulated clouds are optimized against PR and TMI observations, after first being represented in the reduced dimensionality (EOF space). This EOF–1DVAR approach uses sequential termwise optimization procedures. The optimization approach results in scaling of the hydrometeor contents and modification of the drop size distribution by allowing adjustments in the $N_{w0}$ (the intercept of the exponential distribution of rain), often increasing the $N_{w0}$ in agreement with other studies. The end result is the generation of an optimal precipitation radiation database.

We performed a statistical comparison between simulated and observed radar reflectivity and brightness
Fig. 10. CFADs for (a),(b) cloud liquid water, (c),(d) rain, (e),(f) cloud ice, (g),(h) snow, and (i),(j) graupel content obtained from the hybrid EOF–1DVAR approach in TMI footprint scale. The line in each panel represents the mean hydrometeor content profile.
temperatures, over the oceans near East Asia during the summer. The frequency distribution of the retrieved radar reflectivity with the optimization process is much more similar to that of the observations than those without the optimization; that is, the simulated TBs with the optimal retrieved hydrometeor species for every PR observation are in good agreement with the observed TBs at the TRMM frequencies in all rain types, especially at 85 GHz. This implies our EOF–1DVAR approach can retrieve a well-balanced distribution of ice species such as snow and graupel. By applying the new method for improving the vertical hydrometeors profiles, the simulated cloud-radiation database is considerably adapted toward the observed clouds using the TRMM PR and TMI observations.

Finally, although we demonstrated the optimization method using TRMM observations over the oceans near East Asia during summer, the approach itself can be extended to other specific oceanic regions, seasons, or observing platforms, as long as simultaneous precipitation radar and microwave radiometer observations are available. Therefore, it is equally applicable to GPM Core Observatory satellite data and other geographical regions.

Acknowledgments. We acknowledge the anonymous reviewers for their constructive comments. TRMM data were provided by NASA’s Precipitation Processing System. The WRF simulation data were provided by Prof. Song-You Hong at Yonsei University. This research was supported by “the Development of Meteorological Data Utilization and Operation Supportive Technology” of the National Meteorological Satellite Center (NMSC) of the Korea Meteorological Administration (KMA).

REFERENCES


