Adaptation of a model-generated cloud database to satellite observations

Eun-Kyoung Seo,[1,2] Guosheng Liu,[1] Wei-Kuo Tao,[3] and Sang-Ok Han[4]

Received 16 August 2006; revised 16 October 2006; accepted 13 December 2006; published 3 February 2007.

[1] Cloud resolving model outputs are often used to build databases for satellite microwave remote sensing of precipitating clouds. A known problem of this approach is that cloud resolving models tend to systematically produce excessive amount of high density frozen hydrometeors, causing the cloud/radiation model database to have stronger scattering signatures at high microwave frequencies than those observed by satellite or airborne sensors. Consequently, it lowers the performance of the cloud and precipitation retrieval algorithms that utilize the model database. Since multi-frequency satellite observations contain information on hydrometeors' properties, measured brightness temperatures can give guidance as to how the modeled cloud database may be modified to better mimic natural clouds. Following this philosophy, in this study, we propose a method to adapt the modeled database toward observations. The newly constructed database results in a better match to the characteristics of the satellite observed brightness temperatures. **Citation:** Seo, E.-K., G. Liu, W.-K. Tao, and S.-O. Han (2007), Adaptation of a model-generated cloud database to satellite observations, *Geophys. Res. Lett.*, *34*, L03805, doi:10.1029/2006GL027857.

1. Introduction

[2] Retrieving precipitating cloud parameters from satellite observations often relies on a cloud/radiation model-generated database [*Kummerow et al.*, 1996; *Olson et al.*, 1996, 1999]. Mismatch between the model-generated and satellite measured radiances significantly affect the performance of retrieval algorithms [*Panegrossi et al.*, 1998; *Bauer*, 2001; *Seo and Biggerstaff*, 2006]. The database, commonly constructed by combining cloud-generating models and radiative transfer models, relates cloud properties to their corresponding radiances, which are often expressed by brightness temperatures (T_Bs) in microwave frequencies. In order to obtain accurate cloud retrievals from measured T_Bs, the database should consist of both cloud structure and its corresponding T_Bs that are realistically occurring in nature.

[3] In the context of replicating clouds in nature, cloud models still have limitations especially due to imperfect microphysical parameterizations, which commonly use bulk schemes with several categories in representing liquid and ice hydrometeors. These limitations, in turn, influence the replication of T_Bs calculated by radiative transfer models. In particular, cloud models are known to have a tendency to produce excessive amount of high-density ice particles, resulting in stronger microwave scattering signatures than observations [*Bauer*, 2001; *Biggerstaff et al.*, 2006; *McFarquhar et al.*, 2006]. Hence, multi-frequency T_B relations in the model-generated database often differ from those of the observations. By changing microphysical, macrophysical, and environmental factors individually in the database, *Panegrossi et al.* [1998] explored the possibility of mitigating the mismatch between measured and modeled multi-frequency T_B relations (i.e., manifolds). With the same intention, we propose a new approach in this study to mitigate the mismatch problem. However, our approach seeks to ensure the statistical consistency between satellite observations and hydrometeor profiles in the cloud database, which does not ensure the consistency of physical processes within the cloud resolving model. Since satellite microwave T_Bs indirectly represent the various cloud systems occurring in nature, it might be possible to extract the cloud structure information hidden in the measured T_Bs. The idea of our approach is to use measured T_Bs to modify cloud properties in the existing model-generated database. The modified cloud properties then serve as new inputs to the radiative transfer model to generate a new database, which has been adapted to the characteristics of the satellite measurements.

2. Data and Models

[4] Datasets used in this study include satellite data from NOAA-15, -16, and -17 Advanced Microwave Sounding Unit – B (AMSU-B) and hydrometeor structures from a cloud-resolving model. The AMSU-B provides measurements of scene radiance at five frequencies of 89-, 150-, 183.3 ± 1, ±3 and ±7 GHz, which are very useful in estimating, especially, ice amounts in clouds from space. The first two frequencies enable deeper penetration through the atmosphere to the Earth's surface. The radiances measured by the 89-GHz channel are mostly affected by the surface and the lowest layer of atmosphere depending on optical thickness of clouds. The highest three frequencies span the strongly opaque water vapor absorption line at 183 GHz. The sensitivity of each channel to various clouds has been discussed by *Seo and Liu* [2006]. AMSU-B T_B data used in this study are collected over the Atlantic ocean in the longitudes 40° ∼ 120°W and latitudes 0 ∼ 40°N during July in 2004.

[5] The Goddard Cumulus Ensemble model [*Tao and Simpson*, 1993; *Tao et al.*, 2003] outputs simulated for a

[1]Department of Meteorology, Florida State University, Tallahassee, Florida, USA.
[2]School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea.
[3]Laboratory for Atmospheres, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[4]Texas A&M University, College Station, Texas, USA.

Copyright 2007 by the American Geophysical Union.
0094-8276/07/2006GL027857

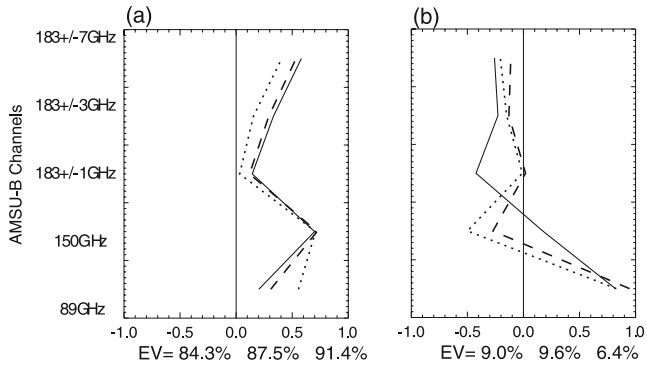

Figure 1. (a) The first EOFs and (b) the second EOFs of TBs at AMSU-B channels for the measurements (solid), the model-generated database (dotted), and the adapted database (dashed). The three numbers at bottom of the figures represent eigenvalues for the measurements, the model-generated database, and the adapted database, respectively.

5-day episode during 7–12 August 1999 are used as model-generated hydrometeor structures. Cyclic lateral boundaries are used. At the top of the model, a free-slip condition is used for horizontal winds, temperature, and specific humidity, and zero vertical velocity is applied. The horizontal domain is 512 × 512 km^2 with a horizontal grid resolution of 2 km. For the present study, a stretched vertical coordinate with 41 levels is used. The model has finer resolution (about 80 meters) in the boundary layer and coarser resolution (about 1000 meters) in the upper levels. The time step is 10 s. Surface fluxes from the ocean and cloud-radiation interactive processes are explicitly included. The observed sea surface temperature (SST) was used for latent and sensible heat flux calculations. The model is forced by zonally-uniform vertical velocity, zonal wind, and thermal and moisture advection based on 6-hourly KWAJEX observations. Kwajalein island is located at 8.44°N, 167.43°E and has the area of 15 km^2, which is the largest island in Kwajalein Atoll (2200 km^2). *Shie et al.* [2003] conducted 3D simulations and compared them with radar observations during three active periods in KWAJEX. The simulation and observation show reasonable agreement in rainfall, apparent heat and moisture sources. The microphysical scheme has a parameterized two-class liquid water (cloud water and rain) and three-class ice-phase (cloud ice, snow and graupel) hydrometeors. For rain, snow and graupel, an exponential drop size distribution was used [*Marshall and Palmer*, 1948]. The densities of snow and graupel were 100 kg m^{-3} and 400 kg m^{-3}, respectively. The horizontal domain size was 512 × 512 km^{-2}. All model-generated variables were archived every one hour. Further details about this case are given by S. Gao et al. (Convective moist vorticity vectors associated with tropical oceanic convection: A three-dimensional cloud-resolving model simulation, submitted to *Geophysical Research Letters*, 2006).

[6] Model-generated data are used to compute T_Bs by a radiative transfer model [*Liu*, 1998] to form a database connecting hydrometeor profiles with T_Bs at AMSU-B channels. One unique feature in the radiative transfer model developed for this study is the inclusion of nonspherical effect of snow particles in computing single scattering properties [*Liu*, 2004]. The area ratio and equivalent density of the ice particles used for designing the nonspherical ice particles follows recent observational results of *Heymsfield and Miloshevich* [2003] and *Heymsfield et al.* [2002]. The usage of nonspherical ice scattering in radiative transfer models results in more accurate brightness temperatures than using spheres for ice particles [*Liu*, 2004]. The database connecting cloud parameters and T_Bs consists of about a half million data points.

3. Mismatch Between Measurement and Model T_Bs in EOF Space

[7] Since all hydrometeors in the radiometer's field of view affect the upwelling radiation, the T_Bs at the five AMSU-B frequencies should reflect the properties of the hydrometeors. To show multivariate relations among T_Bs at the five channels, EOF analysis was conducted. The EOF analysis represents the correlative structure of the T_B variability at the AMSU-B channels. Let $\mathbf{T}_i$ represent a column vector, which contains the depression (clear sky T_B minus cloudy sky T_B) of T_Bs at the five AMSU-B channels at the ith data point in the observations or the model database. Clear sky background T_Bs are obtained by searching the maximum occurrences of T_Bs with bin size of 1 K over an area of 10° (latitude) by 10° (longitude) for every 10-day period. The difference between $\mathbf{T}_i$ and $\bar{\mathbf{T}}$ (mean T_B depression vector) defines the anomaly $\mathbf{T}_i'$. The anomalies are then expanded with EOFs, $\boldsymbol{\varphi}_j$, so that $\mathbf{T}_i' = \sum_{j=1}^{N} \alpha_{i,j}\, \boldsymbol{\varphi}_j$, where $\alpha_{i,j}$ and N represent coefficients for the jth EOF of the ith data point and the number of EOFs, respectively.

[8] The EOF analysis for the measured and modeled T_Bs provides their EOFs ($\boldsymbol{\varphi}_j$) and corresponding coefficients ($\alpha_{i,j}$). Figure 1 shows two dominant EOF patterns for the T_Bs. The two EOFs explain about 93–97% of their variances. The first EOF variability pattern of the modeled T_Bs indicate strong scattering signature at both 89– and 150-GHz. On the other hand, the measurements yield weak scattering signature at 89-GHz but strong scattering signature at 150-GHz. This implies that the model-generated database has excessive amount of large frozen ice particles, which result in strong scattering at both frequencies [*Biggerstaff et al.*, 2006].

[9] Figure 2a shows the manifolds of the observation $\alpha_{i,j}(o)$ and the model-generated $\alpha_{i,j}(m)$ in their first and second EOF space. Their manifolds show relatively large overlap. However, since their corresponding EOFs [$\boldsymbol{\varphi}_j(o)$ and $\boldsymbol{\varphi}_j(m)$] are not exactly the same (Figure 1), the model-generated $\alpha_{i,j}(m)$ should be compared with the observation $\alpha_{i,j}(o)$ in the observational EOF space [$\boldsymbol{\varphi}_j(o)$]. The frequency distribution of the projected model generated $\alpha_{i,j}(p)$ onto $\boldsymbol{\varphi}_j(o)$ is plotted with that of $\alpha_{i,j}(o)$ in the same EOF space (Figure 2b). About 94% of the model-generated database is overlapped with the observed T_Bs occurring over the target ocean. However, their detailed probability density functions show differences. The different manifolds between the observations and model-generated database can result in erroneous cloud retrievals, especially in retrieval algorithms based on Bayesian theorem, in which the retrieval algorithms depend heavily on a priori probability [*Seo and Liu*, 2006]. Therefore, the similarity in the manifolds of observations and model-generated database

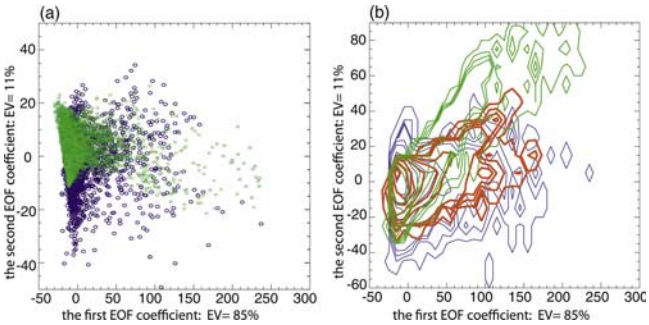

Figure 2. (a) TB distribution of the measurements (blue dots) and the model-generated database (green dots) in their first and second EOF space. (b) TB manifolds of the measurements (blue), the model-generated database (green), and the adapted database (red) in observational EOF space. Contours represent frequencies.

becomes a prerequisite for reliable performance of the retrieval algorithms.

4. Adaptation of Model-Generated Clouds to Measured T_Bs

[10] Because of the inherent mismatch between the measured and model-generated T_B manifolds, the model-generated clouds need to be modified toward more realistic clouds in nature. To achieve a better match between model-generated T_Bs with observations, *Bauer* [2001] performed subselection of the cloud resolving model hydrometeor profiles, while keeping individual profiles unchanged. Our approach is to modify hydrometeor profiles in the cloud database so that T_Bs have better match with observations in EOF space. By doing so, more realistic cloud structures can be extracted from satellite microwave observations, and the new database is adjusted more close toward the observed cloud characteristics.

[11] To accomplish this goal, let us first conduct EOF analysis for a long vector, that is, hydrometeor structure ($\mathbf{Q}$). The long vector contains vertical profiles of all hydrometeor species, such as cloud liquid water ($\mathbf{q}_c$), rain ($\mathbf{q}_r$), cloud ice water ($\mathbf{q}_i$), snow ($\mathbf{q}_s$), and graupel ($\mathbf{q}_g$) as its components. Thus, the vector is defined as:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{q}_c \\ \mathbf{q}_r \\ \mathbf{q}_i \\ \mathbf{q}_s \\ \mathbf{q}_g \end{bmatrix}. \tag{1}$$

The long vector for the ith data point is the sum of mean vector, $\overline{\mathbf{Q}}$, and deviation vector, $\mathbf{Q}_i'$. If the deviation vector is decomposed with EOFs, the long vector can be written as:

$$\mathbf{Q}_i = \overline{\mathbf{Q}} + \mathbf{Q}_i' = \overline{\mathbf{Q}} + \sum_{k=1}^{N} \beta_{i,k} \varsigma_k, \tag{2}$$

where $\beta_{i,k}$ and ς_k denote the kth EOF coefficient and EOF, respectively, at the ith data point, and N denotes the number of EOFs for the expression. The EOFs possess concurrent multilayered and multivariate relations among hydrometer species (Figure 3). The hydrometeor structure ($\mathbf{Q}$) can be related to T_B vector ($\mathbf{T}$) by various approaches, for instances, multivariate linear regression or Bayesian theorem. In particular, the dimension reduction of $\mathbf{Q}$ and $\mathbf{T}$ is very useful to multivariate linear regression, which relate the hydrometeor EOF coefficients $\beta_{i,k}$ to the T_B EOF coefficients $\alpha_{i,j}$. Multivariate linear regression has the benefit to expand the relations outside the database boundary, but is less suitable for highly nonlinear relations. On the other hand, Bayesian method can hold for nonlinear relations locally, but is limited by the boundary of its a priori database. Therefore, both multivariate linear regression and Bayesian theorem will be used in this study for deriving the relations connecting hydrometeor profiles and T_Bs in the model-generated database.

[12] Based on multivariate linear regression, the kth EOF coefficients of a microphysical variable can be derived from satellite AMSU-B T_Bs in a linear aspect as follows:

$$\widetilde{\beta}_k = c_0 + \sum_{j=1}^{4} c_j \alpha_j', \tag{3}$$

where $\widetilde{\beta}_k$ is the kth estimated EOF coefficient of the long hydrometeor vector, α_j' represents the jth EOF coefficient of observed T_Bs projected onto $\boldsymbol{\varphi}_j(m)$, and c_j is a regression coefficient obtained from the multivariate linear regression relating α_j and β_k in the model-generated database. By doing so, new long hydrometeor vectors were built using $\widetilde{\mathbf{Q}} = \overline{\mathbf{Q}} + \sum_{k=1}^{4} \widetilde{\beta}_k \varsigma_k$.

[13] Based on Bayesian approach, the expected vector $\hat{E}(\mathbf{Q})$ given a set of $\mathbf{y}_0$ (the observed state) is expressed in a discrete form [*Lorenc*, 1986; *Olson et al.*, 1996; *Seo and Liu*, 2005] by

$$\begin{aligned} \widetilde{\mathbf{Q}} &= \hat{E}(\mathbf{Q}) \\ &= \sum_{\mathrm{j}} \mathbf{Q}_{\mathrm{j}} \exp\Big\{-0.5\big[\mathbf{y}_0 - \mathbf{y}_s(\mathbf{Q}_{\mathrm{j}})\big]^{\mathrm{T}} \\ &\quad \cdot (\mathbf{O}+\mathbf{S})^{-1}\big[\mathbf{y}_0 - \mathbf{y}_s(\mathbf{Q}_{\mathrm{j}})\big]\Big\}/\hat{A}, \end{aligned} \tag{4}$$

where $\mathbf{O}$ and $\mathbf{S}$ are the observation and simulation error covariance matrices, respectively; $\hat{A}$ is a normalization factor and $\mathbf{y}_s$ is the radiative transfer model simulated brightness temperature vector. Details on the Bayesian formulation using AMSU-B data are given by *Seo and Liu* [2006].

[14] Using both multivariate linear regression and Bayesian methods, new cloud structures ($\widetilde{\mathbf{Q}}$) adjusted to satellite measurements are derived, and then utilized by radiative transfer model to calculate T_Bs. So, a new database having relations between adjusted cloud structure and its corresponding T_Bs was built.

[15] To examine the newly constructed database, the mean and standard deviation of hydrometeor profiles were compared with those for the original model-generated database. The new database has less graupel and rain, more cloud ice and snow, and smaller variability in rain and graupel than the original model-generated database

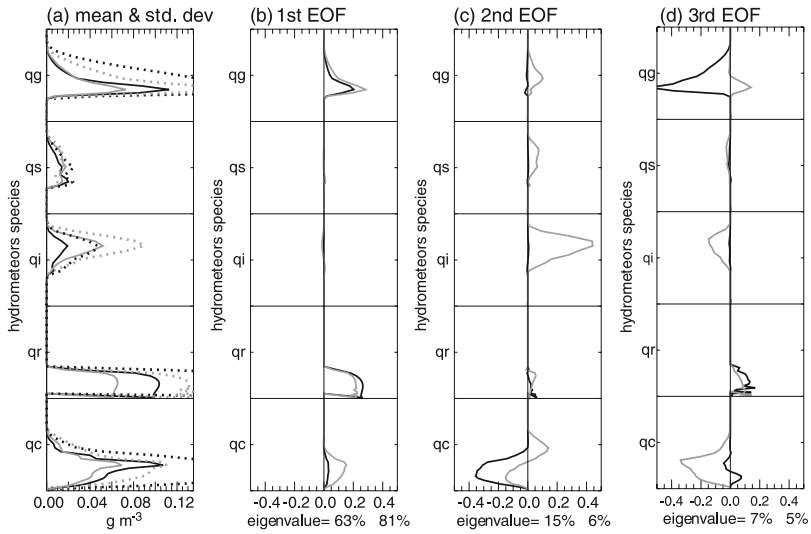

Figure 3. (a) Mean and standard deviation of hydrometeor content profiles and (b, c, d) three dominant concurrent EOF patterns for hydrometeor content profiles. Black and gray lines represent the model-generated and the adapted databases, respectively. In Figure 3a, solid and dotted lines denote mean and standard deviations, respectively. The terms, qc, qr, qi, qs, qg, denote cloud water, rain, cloud ice, snow and graupel content (g m^{-3}), respectively.

(Figure 3). Producing more snow and less graupel is very encouraging in light of the known problem that cloud models tend to produce excessive graupels. Meanwhile, the first EOF pattern of T_Bs in the new database lies between the observation and model-generated database, especially at 89-GHz (Figure 1), indicating that the adjusted database has less scattering by large high-density ice particles than the model-generated database.

[16] As another indication of improvement of the database, the occurrence frequency distribution of the EOF coefficients (α_j) of the new database was compared to the observed ones in EOF space after projecting them onto the observed EOFs (Figure 2b). For the original model-generated T_Bs, about 94% of its total occurrences are overlapped with the observations, while for the newly adapted database about 99% are overlapped. For a more quantitative comparison, we define a structural difference between the two manifolds as:

$$diff = \sum_i \sum_j \left\{ f_{i,j}(m) - f_{i,j}(o) \right\}^2 \Big/ \sum_i \sum_j \left\{ f_{i,j}(o) \right\}^2,$$

where $f_{i,j}(o)$ and $f_{i,j}(m)$ denote occurrences for the measurements and another database, respectively, and i and j denote the ith and jth pixel in the first and second EOF coefficient axes, respectively. The structural difference of the occurrences of the newly adapted database from the observations is about 5%, while that of the original model-generated database is about 20%. Therefore, the new database truly has been adjusted toward observations in terms of cloud structures and T_B signatures.

5. Conclusions

[17] It is known to be an inherent problem that cloud/radiation model-generated database has colder T_Bs at 85 GHz microwave channel than observations because the model database often possesses excessive amount of high-density frozen particles. Hence, the objective of the study is to modify the model-generated cloud properties by using the guidance of observed T_Bs in an effort to improve the supporting database utilized by cloud retrieval algorithms. A basic assumption to fulfill the modification is that the model-generated database supports some part of intrinsic relations between cloud properties and T_B signatures in observations. If so, we can infer cloud properties from observed T_Bs using the relations obtained by the model-generated database. That is, the relations can be interpolated and extrapolated within and beyond the boundary of the model-generated T_Bs in various ways.

[18] By doing so, we could construct new cloud properties and then calculate T_Bs at satellite observation channels. To investigate properness of the newly constructed database, the multivariate relations between T_Bs and hydrometeors (that is, EOFs) and the T_B manifolds in two major EOF space were compared to those in observations and the model-generated database. The results indicate that cloud characteristics in the new database lie more closely to those of observations compared to the original model-generated database. In the meantime, stronger efforts should be made to produce hydrometeor distributions from cloud-resolving model simulations that are more realistic.

[19] **Acknowledgments.** This research has been supported by DOE ARM grant DE-FG02-03ER63526 and NASA grant NNG05GJ17G. The first author was partly supported by the BK21 Project of the Korean Government through the School of Earth and Environmental Sciences, Seoul National University in 2006.

References

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

Biggerstaff, M. I., E.-K. Seo, S. Hristova-Veleva, and K.-Y. Kim (2006), Impact of cloud model microphysics on passive microwave retrievals of cloud properties. Part I: Model comparison using EOF analyses, *J. Appl. Meteorol. Climatol.*, *47*, 930–954.

Heymsfield, A. J., and L. M. Miloshevich (2003), Parameterizations for the cross-sectional area and extinction of cirrus and stratiform ice cloud particles, *J. Atmos. Sci.*, *60*, 936–956.

Heymsfield, A. J., S. Lewis, A. Bansemer, J. Iaquinta, L. M. Miloshevich, M. Kajikawa, C. Twohy, and M. R. Poellot (2002), A general approach for deriving the properties of cirrus and stratiform ice cloud particles, *J. Atmos. Sci.*, *59*, 3–29.

Kummerow, C. D., W. Olson, and L. Giglio (1996), A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors, *IEEE Trans. Geosci. Remote Sens.*, *34*, 1213–1232.

Liu, G. (1998), A fast and accurate model for microwave radiance calculations, *J. Meteorol. Soc. Jpn.*, *76*, 335–343.

Liu, G. (2004), Approximation of single scattering properties of ice and snow particles for high microwave frequencies, *J. Atmos. Sci.*, *61*, 2441–2456.

Lorenc, A. C. (1986), Analysis methods for numerical weather prediction, *Q.J.R. Meteorol. Soc.*, *112*, 1177–1194.

Marshall, J. S., and W. M. Palmer (1948), The distribution of raindrops with size, *J. Meteorol.*, *5*, 165–166.

McFarquhar, G. M., H. Zhang, G. Heymsfield, R. Hood, J. Dudhia, J. B. Halverson, and F. Marks (2006), Factors affecting the evolution of Hurricane Erin (2001) and the distributions of hydrometeors: Role of microphysical processes, *J. Atmos. Sci.*, *63*, 127–150.

Olson, W. S., C. D. Kummerow, G. M. Heymsfield, and L. Giglio (1996), A method for combined passive-active microwave retrievals of cloud and precipitation profiles, *J. Appl. Meteorol.*, *35*, 1763–1789.

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

Panegrossi, G., S. Dietrich, F. S. Marzano, A. Mugnai, E. A. Smith, X. Xiang, G. J. Tripoli, P. K. Wang, and J. P. V. P. Baptista (1998), Use of cloud model microphysics for passive microwave-based precipitation retrieval: Significance of consistency between model and measurement manifolds, *J. Atmos. Sci.*, *55*, 1644–1673.

Seo, E.-K., and M. I. Biggerstaff (2006), Impact of cloud model microphysics on passive microwave retrievals of cloud properties. Part II: Uncertainty in rain, hydrometeor structure and latent heating retrievals, *J. Appl. Meteorol. Climatol.*, *47*, 955–972.

Seo, E.-K., and G. Liu (2005), Retrievals of cloud ice water path by combining ground cloud radar and satellite high-frequency microwave measurements near the ARM SGP site, *J. Geophys. Res.*, *110*, D14203, doi:10.1029/2004JD005727.

Seo, E.-K., and G. Liu (2006), Determination of 3-D cloud ice water contents by combining multiple data sources from satellite, ground radar and numerical model, *J. Appl. Meteorol. Climatol.*, *45*, 1494–1504.

Shie, C.-L., W.-K. Tao, and J. Simpson (2003), Simulated KWAJEX convective systems using a 2D and 3D cloud resolving model and their comparisons with radar observations, paper presented at 31st Conference on Radar Meteorology, Am. Meteorol. Soc., Seattle, Wash.

Tao, W.-K., and J. Simpson (1993), The Goddard cumulus ensemble model. Part I: Model description, *Terr. Atmos. Ocean Sci.*, *4*, 35–72.

Tao, W.-K., et al. (2003), Microphysics, radiation and surface processes in the Goddard Cumulus Ensemble (GCE) model, *Meteorol. Atmos. Phys.*, *82*, 97–137.

S. O. Han, Texas A&M University, College Station, TX 77843-3150, USA.

E.-K. Seo and G. Liu, Department of Meteorology, Florida State University, 404 Love Building, Tallahassee, FL 32306-4520, USA. (eunkyoungseo@gmail.com)

W.-K. Tao, Laboratory for Atmospheres, NASA Goddard Space Flight Center, Code 6131, Greenbelt, MD 20771, USA.