Cluster analysis of tropical clouds using CloudSat data

Yuying Zhang,[1] Steve Klein,[1] Gerald G. Mace,[2] and Jim Boyle[1]

[1] The mesoscale patterns of cloud/precipitation radar reflectivity from early CloudSat data are used to identify distinct tropical cloud regimes via a cluster analysis. Five basic cloud regimes are identified, and the geographical distribution of their occurrence frequency is quantified. Although the contemporary MODIS observations show some limitations to CloudSat observations, comparison with traditional passive satellite observations shows that CloudSat describes the major features of the vertical structure of the tropical cloud regimes. Using the monthly mean vertical velocity at 500 hPa as an indicator, the elements of each cloud regime are sorted into different dynamical regimes, and the results demonstrate the links between clouds and the atmospheric circulation. **Citation:** Zhang, Y., S. Klein, G. G. Mace, and J. Boyle (2007), Cluster analysis of tropical clouds using CloudSat data, *Geophys. Res. Lett.*, *34*, L12813, doi:10.1029/2007GL029336.

1. Introduction

[2] Current Global Climate Models (GCMs) cannot accurately and consistently simulate cloudiness [*Bony et al.*, 2004]. In order to improve models, it is first required to identify where the simulated clouds differ from observations of the real world. Cloud simulations have been evaluated against observations from satellite and ground sites in the past. The CloudSat radar is the first spaceborne millimeter wavelength radar, and the new global survey of cloud vertical profiles provides many opportunities to describe cloud structures and apply that knowledge to improve cloud parameterizations in models.

[3] Due to irregular space-time sampling from surface or space platforms, the traditional method to evaluate GCM simulated clouds relies on comparing large spatial and temporal means of model output with observations [*Weare et al.*, 1996]. But because of compensating errors, this method cannot effectively constrain cloud simulations [*Norris and Weaver*, 2001]. A complementary method is to group cloud properties using the meteorological parameters, such as 500-hPa vertical velocity or Sea Surface Temperature, with which they co-vary [e.g., *Wyant et al.*, 2006; *Bony et al.*, 2004]. The most important disadvantage associated with this method is a lack of reliable data for some atmospheric variables, such as the vertical velocity from weather analyses. Another technique, termed 'cluster analysis', aims to objectively identify cloud regimes based on observed cloud information alone without any knowledge of other meteo-

rological parameters. This simple approach can classifies objectively distinct cloud regimes which can then be linked to atmospheric characteristics. In this approach, the character of the individual clusters is objectively described by the data themselves.

[4] The cluster analysis method has been implemented on passive sensing satellite data in tropics [*Jakob and Tselioudis*, 2003; *Rossow et al.*, 2005], in a single midlatitude spatial domain [*Gordon et al.*, 2005], and on Tropical Rainfall Measuring Mission (TRMM) data [*Boccippio et al.*, 2005]. We show early results from CloudSat radar data using a cluster-analysis method, and relate the regimes and radar reflectivity to 500 hPa vertical pressure velocity. Section 2 briefly describes the observations and the cluster method. Results are presented in section 3, while the potential application of these results and future work with CloudSat data are discussed in section 4.

2. Observations and Analysis Method

[5] CloudSat, launched on 28 April 2006, is orbiting in formation with other NASA spacecraft in the A-Train [*Stephens et al.*, 2002]. The Cloud Profiling Radar (CPR) [*Im et al.*, 2006] on CloudSat is a 94-GHz nadir-looking radar which measures the energy backscattered by clouds and precipitation within a 1.5 km across-track by 2.5 km along-track radar footprint. Cloud vertical boundaries can be estimated at a 250 m sampling interval. The radar, providing simultaneously observations on cloud condensate and precipitation, will complement the remote sensing data collected by other A-Train instruments. The CloudSat ground track repeats every 16 days. The sample rate is 0.16 second per vertical profile. The data used here is from one of the CloudSat Standard Data Products, Level 2B Cloud Geometrical Profile (2B-GEOPROF) [*Mace*, 2004]. At this stage, CALIPSO data are not used in this study.

[6] The vertical profiles of radar reflectivity collected during 84 days (Jun 15 ∼ Sep 6 2006) form the basis of the analysis for the tropical regions (±30° latitude). We look for distinctive patterns in the joint frequency distributions of the height and radar reflectivity (hereafter H-dBZ) from 200 profile sequences occurring within ∼2° regions (hereafter referred to as an element). The length of each element is similar to the horizontal size of the regions used in the analysis of passive ISCCP D1 data. The seven pressure boundaries used to bin CloudSat data coincide with those in the ISCCP D1 dataset, and the height-pressure conversion is based on the standard tropical profile. Since the domain analyzed is different for this study than for the other studies, the clustering may find somewhat different clusters. Due to surface contamination [*Mace et al.*, 2007], the signal returns from the first 1 km are not considered. Reflectivity above the CPR minimum detectable signal of −28 dBZ is binned into seven categories with a bin interval of 10 dBZ. The relative frequency of occurrence

[1]Energy and Environment Directorate, Lawrence Livermore National Laboratory, Livermore, California, USA.
[2]Department of Meteorology, University of Utah, Salt Lake City, Utah, USA.
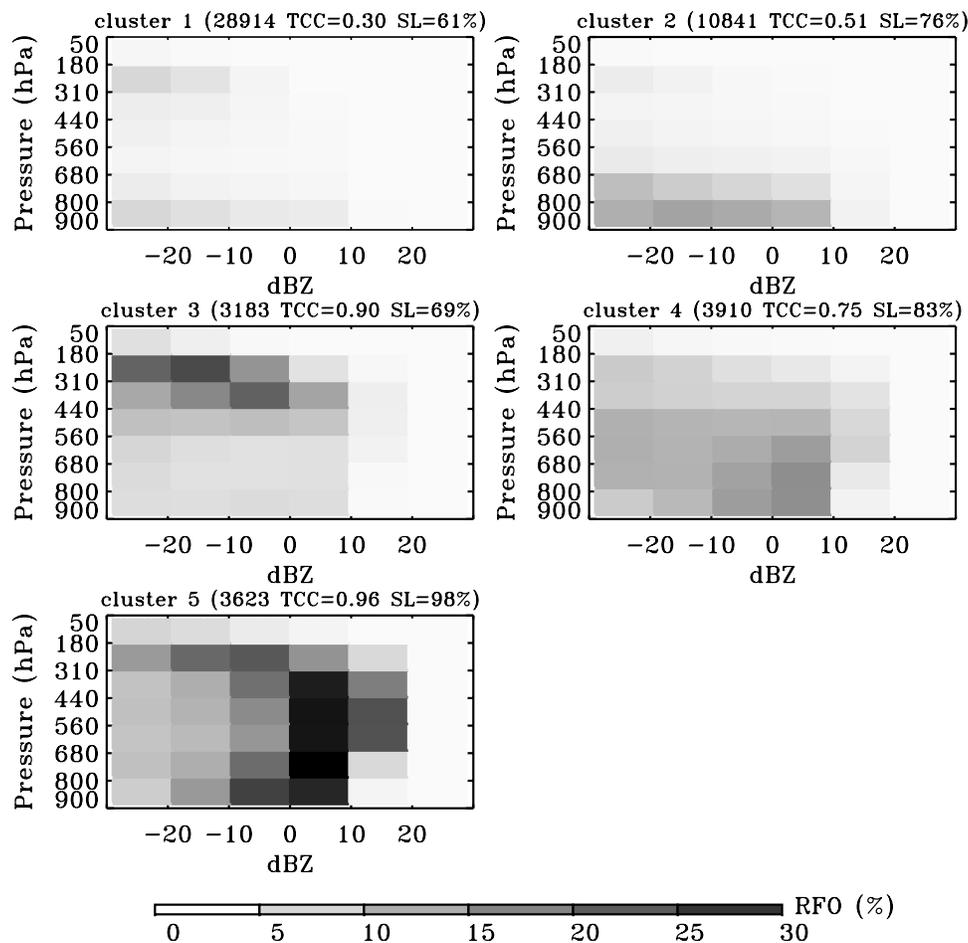
Figure 1. H-dBZ histograms of the centroids of the five clusters. Cluster 1, low cloud and thin cirrus; Cluster 2, subtropical marine stratus; Cluster 3, anvil cirrus cloud; Cluster 4, cumulus congestus; and Cluster 5, deep convection and heavy precipitation. The first number in brackets is the number of elements in each cluster. TCC is the total cloud cover which is the fraction of CloudSat profiles that contain cloud or/and precipitation. SL is, by considering a 200 profile average, the percentage of elements with a single cloud layer, where separate layers are defined by a separation of cloudy range gates by 2 or more non-cloudy range gates.

(RFO) represents the frequency of data within a dBZ range at a given level. The diagrams, similar to contoured frequency by altitude diagrams (CFADs) [*Yuter and Houze*, 1995], summarize the frequency distribution of radar reflectivity factor in the vertical dimension, and are a better way to describe the mesoscale cloud patterns for the purpose of cluster analysis in a large-scale sense. When fewer than 10 of the 200 profiles in an element have cloud, the element will be defined as clear sky and not enter the cluster analysis. With this definition, 22,565 and 50,471 elements are classified as clear sky and cloudy sky, respectively.

[7] We also make use of passive-sensing observations from the Moderate Resolution Imaging Spectroradiometer (MODIS). Since CloudSat does not scan, only a single vertical curtain of measurements are collected along the satellite subtrack. On the other hand, MODIS scans a 2300-km across-track swath with a high horizontal resolution. Cloud properties included in MOD06 [*Platnick et al.*, 2003], including cloud top pressure and visible optical thickness, are derived from MODIS multi-spectral observations. Only daytime MODIS data are used.

[8] The basic issue addressed here is how well the measurements from CloudSat can describe the spatial distribution and the vertical characteristics of distinct and recurring tropical cloud regimes. The H-dBZ histograms are classified using a K-means cluster algorithm [*Anderberg*, 1973]. The algorithm iteratively searches for a set of a predefined number (K) of clusters. The cluster centroids represent specific histogram patterns which have a minimum within-cluster distance measured in a Euclidian sense. K is the only arbitrary parameter within the cluster analysis method, and the initial K cluster centroids are specified by random selection. Every histogram is assigned to one of the clusters according to the minimum Euclidian distance with the initial centroids, and, then, new centroids are calculated. The iteration continues until the cluster set is obtained with the minimum sum of all the distances between individual histograms and centroids.

3. Results and Discussion

[9] The cluster algorithm is applied to all histograms in the tropical region except clear elements. The total cloud
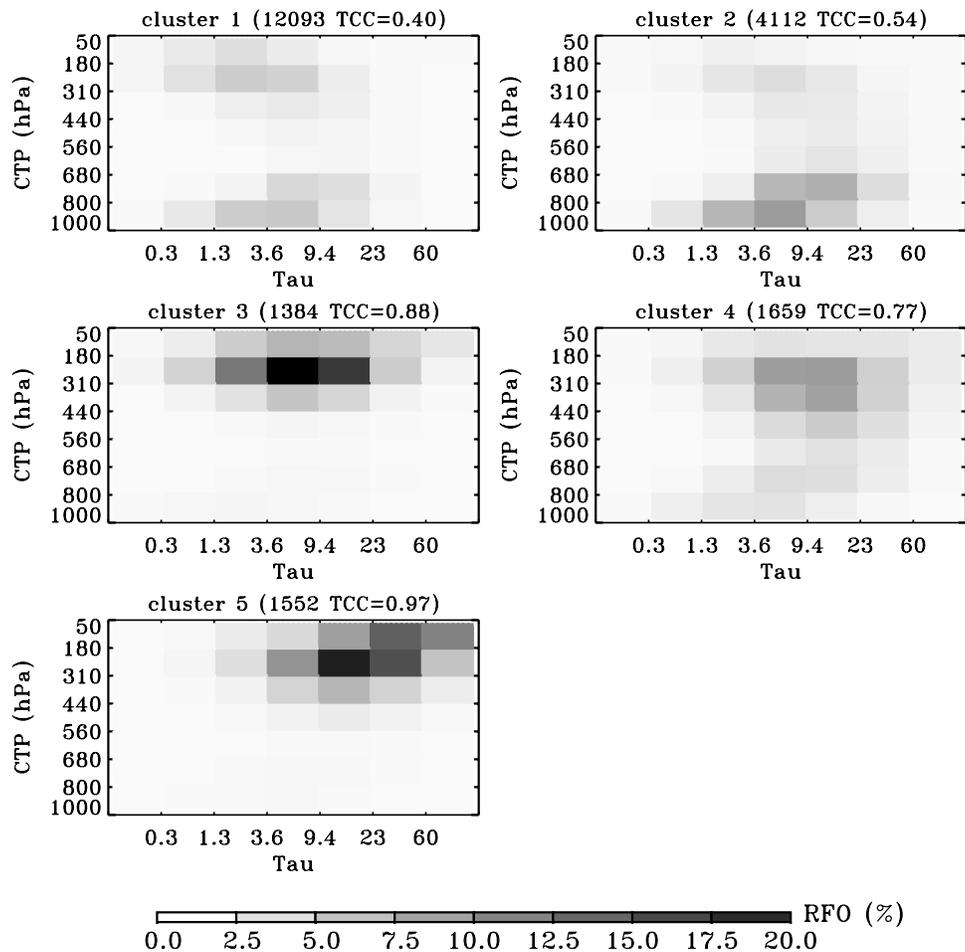
Figure 2. The cloud top pressure (CTP) – optical thickness (Tau) frequency histogram patterns from contemporary MODIS observations composited into the clusters from CloudSat data. The first number in brackets is the case number, and TCC is the MODIS total cloud cover and the integral over the entire histogram.

cover (TCC) is the fraction of profiles in a cluster that contains cloud or precipitation. Following the four criteria mentioned by *Rossow et al.* [2005] to judge the outcome, we show the histograms of the five clusters' centroids in Figure 1. The first two clusters with low correlation between their geographic distributions will merge together if K is four, and the last cluster will separate into two clusters with their cloud patterns highly correlated if K is six. The pattern correlation coefficients between any two clusters have a maximum of 0.65, and a minimum of −0.12. Tests show that the five clusters do not depend on the selection of initial clusters. The joint histograms of cloud top pressure and optical thickness from the contemporaneous MODIS cloud data corresponding to each of the CloudSat elements are created by matching the position of the CloudSat elements. These histograms are composited into the CloudSat clusters without reclustering (Figure 2), so can be used to quantify the cloud properties of the CloudSat cloud regimes. After organizing clouds into regimes, the frequency of occurrence maps for each cluster (Figure 3) can also represent spatial variations of the cloud clusters.

[10] The first cloud regime is the most frequent of the five, and is characterized by a mixture of shallow low clouds and thin cirrus with low TCC. These cloud regimes are found frequently in the large subsidence regions off the west coasts of South America and Africa and higher latitudes in the northern hemisphere. The second regime is the subtropical marine stratus regime as named by *Rossow et al.* [2005], and frequently happens in the marine regions between 10S and 30S in the southern hemisphere and higher latitudes in the corresponding northern hemisphere. The third regime is dominated by high clouds of moderate dBZ which we will call anvil cirrus clouds with an averaged TCC of about 0.9, and predominantly found in the western and eastern Pacific and the Asian Monsoon regions of India and south Asia. Cirrus clouds may be either outflow from deep cumulus or associated with synoptic and mesoscale disturbances. The fourth regime consists of clouds over a wide range of dBZ and levels beneath 440 hPa. This fourth cloud regime suggests small-scale isolated convection, and exhibits lower cloud top and lower reflectivities than deep convection. For this reason, we call this cluster cumulus congestus and it has high RFO especially over high topography on the west coast of central and South America, the east central Africa, and south Asia. It is also common in the same regions that the anvil cirrus and the fifth clusters are found. The fifth and last regime is composed primarily of large reflectivity clouds from the surface to 180 hPa. We call this regime deep convection and heavy precipitation. Some of this regime is associated with the large mesoscale
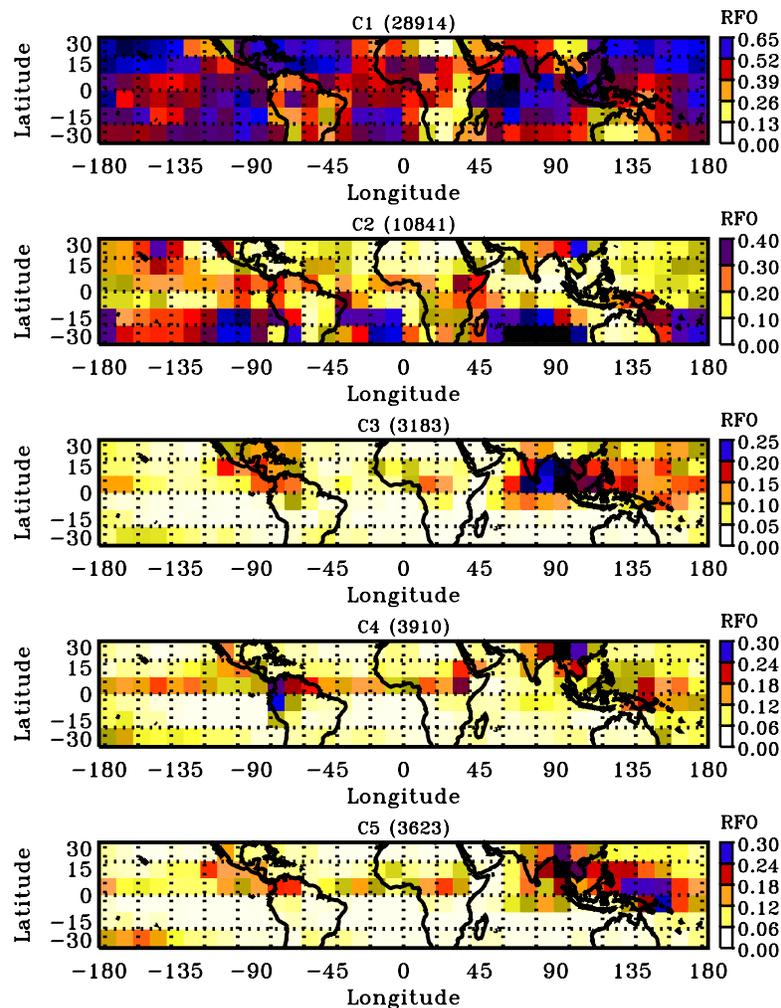
Figure 3. The three-month average occurrence fraction of each cluster. The sum of the frequencies across the clusters in a given 10°-by-10° box represents the frequency of cloudy elements.

systems, and is most common in the west Pacific warm pool and the Asian Monsoon region [*Zipser et al.*, 2006; *Liu and Zipser*, 2005]. The spatial distribution of the anvil-cloud regime is highly correlated with that of deep convection, which may imply that anvil cloud cluster is primarily derived from deep convective outflows.

[11] *Rossow et al.* [2005] found six patterns that represented reasonably well the mesoscale distributions of tropical cloud top pressure and cloud optical thickness using the ISCCP D1 data set [*Rossow and Schiffer*, 1999]. Despite substantial differences in data and period, the clusters defined in this study correspond closely with those studied by *Rossow et al.* [2005], especially CloudSat clusters 2, 3, 4, 5 with ISCCP Weather states (WSs) 6, 2, 3, 1, respectively. The primary advantage of using active remote sensing data is that we are better able to identify the vertical structure of the cloud fields within the clusters. Due to the detection threshold of the radar and the inability of the CPR to detect clouds lower than 1km, it is possible that CloudSat cluster 1 corresponds to the combination of WSs 4 and 5.

[12] The comparison of cloud property patterns between the observations from MODIS and CloudSat shows some substantial differences as well as agreement. First, the

MODIS TCC of the first regime is 30% larger than that from CloudSat, which may be because the clouds detected by MODIS exhibit reflectivity less than the cloud radar threshold and/or the low level clouds exist in the lowest 1 km of the profile. Second, the cloud-top heights from MODIS and CloudSat are mismatched sometimes, especially for cluster 3 and 5. Since there are also uncertainties in estimating cloud top height from MODIS, this issue needs more information to be addressed. On the other hand, CloudSat demonstrates that the vertical structure of clouds can differ dramatically for scenes which appear to traditional satellites to consist of high cloud tops and large optical thickness (e.g., cluster 3, 4, 5). Specifically, the cloud pattern of cluster 3 describes clouds at high altitude with relatively small dBZ, while that of cluster 4 represents clouds at lower altitude with large dBZ. Furthermore, the clouds of cluster 5 have largest dBZ and highest RFO.

[13] The MODIS observations with 25 km in width along the CloudSat footprint are tested when CloudSat defines the observation as clear sky. Considering only the scenes including cloud coverage from MODIS, the TCC of 20% suggests that CloudSat has missed large portion of low cloud and some of thin cirrus cloud. This result is expected and
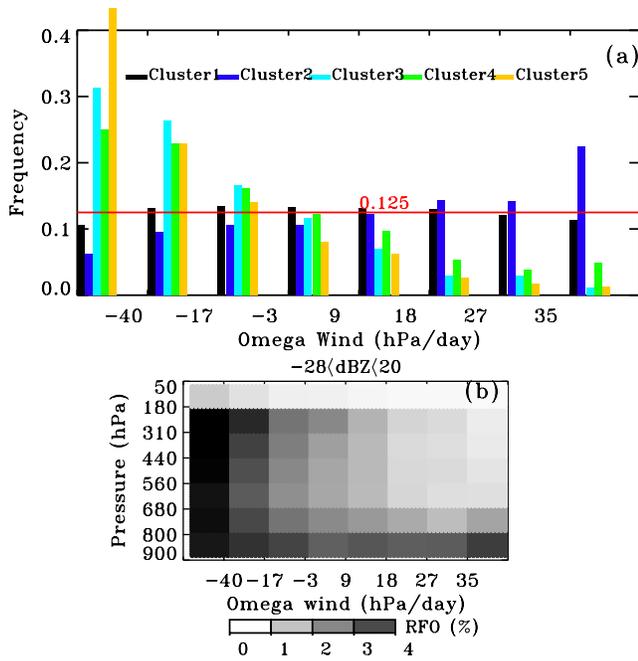
Figure 4. (a) The frequency of the occurrence for each cluster from CloudSat data as a function of the monthly mean vertical pressure velocity (Omega Wind) at 500 hPa calculated from NCEP analysis data. Each bin of the vertical velocity represents the equivalent occurrence frequency for the applied 500 hPa vertical pressure velocity. If there were no association of cluster regime with 500 hPa vertical pressure velocity, the frequency of occurrence for a given cluster would be 0.125 (red line) apart from random fluctuations. (b) The relative frequency of occurrence for valid dBZ range at a given pressure level as a function on the abscissa of the monthly mean 500 hPa vertical pressure velocity.

points to the need to combine CloudSat with other A-Train data streams to fully represent the cloud occurrence and cloud property statistics [*Mace et al.*, 2007]. The cloud patterns have high RFO and high TCC in the west coasts of South America and Africa and south Asia, where the boundary layer clouds are known to be very shallow and often less than 1km deep [*Betts et al.*, 1992].

[14] To show the association of cloud regimes with large-scale dynamics, the oceanic elements in each cluster are grouped using the monthly mean vertical pressure velocity at 500 hPa calculated from National Centers for Environmental Prediction (NCEP) GDAS operational analysis data following the method of *Bony et al.* [2004]. The occurrence frequencies of each cluster in identical dynamic ranges are depicted in Figure 4a. The vertical pressure velocities that separate bins are chosen such that each bin of the vertical velocity represents the equivalent occurrence frequency. In spite of large uncertainties in NCEP analysis data and the short duration and irregular spatial coverage of currently available CloudSat data, we can still see that the anvil cloud regime, the cumulus congestus regime, and the deep convection regime have a tendency to be more common in regions of upward motion. The cirrus and low cloud regime has a less clear association with vertical pressure velocity.

[15] Finally, the radar reflectivity values independent of the clusters they are associated with are grouped by the 500 hPa vertical velocity (Figure 4b). The cloud radar reflectivities are widely used to empirically estimate cloud water content based on in situ measurements [e.g., *Viltard et al.*, 2000; *Hogan et al.*, 2006]. Before the cloud retrievals become available, the radar reflectivities can be used to qualitatively represent the vertical distributions of cloud water. The panel shows the distribution of all significant radar returns and may be used to represent the cloud fraction. As expected, deep-convective and anvil cirrus cloud systems are most common in ascending regions, while lower level cumulus and stratus are more common in subsiding regions.

4. Summary and Future Application

[16] This study reports on early results from cluster analysis using CloudSat data and shows the possibility to derive information on cloud regimes. The five prevalent cloud regimes exist with various frequencies of occurrence. The geographical distributions of occurrence frequency for each cloud cluster suggest a possible connection between distinct cloud regimes and certain atmospheric characteristics. The links of individual cloud clusters to the dynamic components (vertical pressure velocity) are indicated by means of compositing. It is demonstrated that the CloudSat data can describe the major vertical features for prevalent cloud regimes in tropical regions, and the connection between cloud regimes and atmospheric circulation characteristics may provide information about cloud generation mechanisms.

[17] When combined carefully with other information from the A-Train such as CALIPSO lidar data, it may be possible in many circumstances to derive the local radiative and latent heating rate profiles, which are essential to climate research and weather prediction. Also, the CALIPSO lidar will sense most of the optically thin cirrus and shallow PBL clouds that the CPR on CloudSat will miss. The new techniques to organize distinct cloud regimes and sort them by dynamical regime provide a more informative and stringent way to evaluate models with independent and complementary information about the quality of the cloud simulation. The results will help to demonstrate the model sensitivity to changes in circulation, and may help to provide more insight into sources of errors in cloud parameterization since the results can show particular deficiencies not readily apparent in time-space average comparisons.

References

Anderberg, M. R. (1973), *Cluster Analysis for Applications*, 359 pp., Academic, New York.
Betts, A. K., P. Minnis, W. Ridgway, and D. F. Young (1992), Integration of satellite and surface data using a radiative-convective oceanic boundary-layer model, *J. Appl. Meteorol.*, *31*, 340–350.

Boccippio, D. J., et al. (2005), The tropical convective spectrum. Part I: Archetypal vertical structures, *J. Clim.*, *18*, 2744–2769.

Bony, S., et al. (2004), On dynamic and thermodynamic components of cloud changes, *Clim. Dyn.*, *22*, 71–86.

Gordon, N. D., J. R. Norris, C. P. Weaver, and S. A. Klein (2005), Cluster analysis of cloud regimes and characteristic dynamics of midlatitude synoptic systems in observations and a model, *J. Geophys. Res.*, *110*, D15S17, doi:10.1029/2004JD005027.

Hogan, R. J., et al. (2006), The retrieval of ice water content from radar reflectivity factor and temperature and its use in evaluating a mesoscale model, *J. Appl. Meteorol. Climatol.*, *45*, 301–317.

Im, E., et al. (2006), Cloud profiling radar for the CloudSat mission, *IEEE Trans. Aerosp. Electron. Syst.*, *20*, 15–18.

Jakob, C., and G. Tselioudis (2003), Objective identification of cloud regimes in the Tropical Western Pacific, *Geophys. Res. Lett.*, *30*(21), 2082, doi:10.1029/2003GL018367.

Liu, C., and E. J. Zipser (2005), Global distribution of convection penetrating the tropical tropopause, *J. Geophys. Res.*, *110*, D23104, doi:10.1029/2005JD006063.

Mace, G. G. (2004), Level 2 GEOPROF product process description and interface control document, Coop. Inst. for Res. in the Atmos., Fort Collins, Colo.

Mace, G. G., R. Marchand, Q. Zhang, and G. Stephens (2007), Global hydrometeor occurrence as observed by CloudSat: Initial observations from summer 2006, *Geophys. Res. Lett.*, *34*, L09808, doi:10.1029/2006GL029017.

Norris, J. R., and C. P. Weaver (2001), Improved techniques for evaluating GCM cloudiness applied to the NCAR CCM3, *J. Clim.*, *14*, 2540–2550.

Platnick, S., et al. (2003), The MODIS cloud products: Algorithms and examples from Terra, *IEEE Trans. Geosci. Remote Sens.*, *41*, 459–473.

Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, *80*, 2261–2287.

Rossow, W. B., G. Tselioudis, A. Polak, and C. Jakob (2005), Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures, *Geophys. Res. Lett.*, *32*, L21812, doi:10.1029/2005GL024584.

Stephens, G. L., et al. (2002), The CloudSat mission and the A-Train, *Bull. Am. Meteorol. Soc.*, *83*, 1771–1790.

Viltard, C., et al. (2000), Combined use of the radar and radiometer of TRMM to estimate the influence of drop size distribution on rain retrievals, *J. Appl. Meteorol.*, *39*, 2103–2114.

Weare, B. C., et al. (1996), Evaluation of the vertical structure of zonally averaged cloudiness and its variability in the Atmospheric Model Intercomparison Project, *J. Clim.*, *9*, 3419–3431.

Wyant, M. C., et al. (2006), A comparison of low-latitude cloud properties and their response to climate change in three AGCMs sorted into regimes using mid-tropospheric vertical velocity, *Clim. Dyn.*, *27*, 261–279.

Yuter, S. E., and R. A. Houze (1995), Three-dimensional Kinematic and microphysical evolution of Florida Cumulonimbus. Part II: Frequency distributions of vertical velocity, reflectivity, and differential reflectivity, *Mon. Weather Rev.*, *123*, 1941–1963.

Zipser, E. J., et al. (2006), Where are the most intense thunderstorms on Earth?, *Bull. Am. Meteorol. Soc.*, *87*, 1057–1071.

————————

J. Boyle, S. Klein, and Y. Zhang, Energy and Environment Directorate, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94550, USA. (zhang24@llnl.gov)

G. G. Mace, Department of Meteorology, University of Utah, 819 Browning Building, 135 S. 1460 East, Salt Lake City, UT 84112, USA.