

City University of New York (CUNY)

## CUNY Academic Works

---

Dissertations and Theses

City College of New York

---

2015

### Characterization of Tropical Clouds Using Multi-Satellite Observations

Ricardo Anderson  
*CUNY City College*

[How does access to this work benefit you? Let us know!](#)

More information about this work at: [https://academicworks.cuny.edu/cc\\_etds\\_theses/345](https://academicworks.cuny.edu/cc_etds_theses/345)

Discover additional works at: <https://academicworks.cuny.edu>

---

This work is made publicly available by the City University of New York (CUNY).  
Contact: [AcademicWorks@cuny.edu](mailto:AcademicWorks@cuny.edu)

# **Characterization of Tropical Clouds Using Multi-Satellite Observations**

BY

RICARDO C. ANDERSON

THESIS

Submitted in partial fulfillment  
of the requirements for the degree of  
Masters of Science in Earth and Atmospheric Science, City College of New York  
The City University of New York  
May 2015

Mentor: Z. Johnny Luo

5/12/2015

## Table of Contents

<b>Abstract</b> .....	3
<b>Introduction</b> .....	5
1.1 Overview .....	5
1.2 Background .....	5
<b>Methodology</b> .....	7
2.1.1 CloudSat .....	7
2.1.2 TRMM.....	9
2.1.3 ISCCP .....	10
2.1.4 Dataset.....	10
2.2 Contoured Frequency Altitude Diagrams .....	11
2.3 K-Mean Clustering Analysis .....	13
2.4 ISCCP Collocation.....	15
2.5 ISCCP Regimes in Infrared.....	16
<b>Results</b> .....	17
<b>Discussion</b> .....	22
<b>Conclusion</b> .....	25
<b>Acknowledgments</b> .....	26
<b>Figures</b> .....	27
<b>Appendix A- Cluster 1 examples</b> .....	49
<b>Appendix B- Cluster 2 examples</b> .....	53
<b>Appendix C- Cluster 3 examples</b> .....	57
<b>Appendix D- Cluster 4 examples</b> .....	61
<b>Appendix E</b> .....	65
<b>Bibliography</b> .....	69

## Abstract

Clouds are an important component of the earth-climate system and play a critical role in affecting energy and water cycle of the planet. In particular, tropical convective clouds account for the majority of the precipitation that fall on the Earth's surface. Multiple active satellite missions in recent decade such as TRMM (Tropical Rainfall Measurement Mission) and CloudSat have provided fruitful new insight into the internal structures of these tropical convective clouds. In conjunction with cloud data from ISCCP (International Satellite Cloud Climatology Project) that is based passive remote sensing technology in the visible and infrared spectrum, this allows for a more coherent understanding of the dynamic structure of tropical clouds.

In this study, we focus on the synergy between CloudSat and TRMM radar reflectivity data in a CFAD (Contoured Frequency by Altitude Diagram) framework and apply a clustering analysis to identify distinct clusters. The properties of these clusters were also further analyzed with regards to their cloud top height and radar echo top height. In addition, they were compared with both the visible/infrared and infrared-only Weather States (WSs) from ISCCP.

Results show that there are four tropical cloud clusters containing three precipitating cloud regimes and one non-precipitating cloud regime. Signatures of deep convection, cumulus congestus, and shallow precipitating clouds were identified in the three precipitating cloud clusters. Regions of shallow precipitating

clouds are mostly associated with sinking air motion, while deep convective and congestus cloud regimes were present in regions of rising air motion. Comparison with collocated ISCCP WS data shows broad agreement, although that ISCCP tends to show lower frequency of convective cloud regimes and higher occurrence frequency of non-convective cloud regimes due to differences in remote sensing techniques and spectra used.

## Introduction

### 1.1 Overview

Clouds are an important component of the earth-climate system and play a critical role in affecting energy and water cycle of the planet. More so, tropical convective clouds account for the majority of the precipitation that accumulates on the Earth's surface. Active remote sensing satellite missions such as CloudSat and Tropical Rainfall Measurement Mission (TRMM) provide insight into the internal structure of these tropical convective clouds. Together with International Satellite Cloud and Climatology Project (ISCCP) which is a based on passive remote sensing technology in both the visible and infrared (IR) spectrum, they allow for a more coherent understanding of the dynamic characteristics of clouds in the tropics. A few examples of these dynamics are vertical structure, precipitation patterns, and satellite radiance measurements. Although the objectives for each mission were different, together they are able to help paint a more complete picture of the dynamic structure and characteristics of clouds.

### 1.2 Background

Previous cluster analysis (Boccippio, Petersen, and Cecil 2005; Elsaesser et al. 2010; Jakob 2003; Rossow et al. 2005; Tan and Jakob 2013; Zhang et al. 2007) were performed on various missions such as CloudSat, ISCCP, and TRMM. Jakob 2003 first performed clustering analysis on ISCCP for a small region in the South East Pacific. Rossow et al (2005) took this globally and applied it to the entire

tropics. Following the same line, Zhang et al (2007) exploited Contoured Frequency Altitude Diagrams (CFAD) (Yuter and Houze 1995) with radar reflectivity and height data from CloudSat. Similarly, clustering of TRMM data (Elsaesser et al. 2010) utilized this same framework as well. CloudSat and TRMM are complementary of each other although they have very different orbits. The CPR sensor onboard CloudSat is very good at detecting radar reflectivity from -30dBZ upwards to ~15dBZ while the PR sensor onboard TRMM is able to detect radar reflectivity values of 17dBZ and higher. In this study, we focus on the collocated cases of both CloudSat and TRMM extending prior studies by Zhang et al. (2007) and Elsaesser et al. 2010 utilizing K-means clustering of CFAD from both CloudSat and TRMM. We also examine these clusters and compare them to their corresponding ISCCP weather states in both the visible and infrared spectrum to study the correlation between them.

## Methodology

### 2.1.1 CloudSat

Clouds are a natural part of the water cycle in the atmosphere and dominate the energy budget of the planet (Rossow and Schiffer 1999; Stephens et al. 2002). Simultaneously, clouds also tend to cool the Earth by reflecting sunlight and also warm the earth by absorbing thermal radiation (Wielicki et al. 1995). CloudSat is an experimental satellite launched by NASA in 2006 as a part of the Afternoon Constellation (A-Train). The A-Train is a constellation of satellites that all pass the equator at 13:30 local time as well as 01:30 local time. Each satellite within the A-Train plays a synergic role in improving our understanding of different aspects of Earth's ever-so changing climate. Figure 1 shows a summary of the satellites of the A-Train configuration up until June 2011. CloudSat is placed 15 seconds ahead of CALIPSO and 2 minutes behind AQUA. CALIPSO is strategically placed 15 seconds behind CloudSat since it is able to pick up aerosol particles, which are smaller than cloud size particles. As shown in Figure 2, together different members of the A-Train are able to put together dynamics of a storm. This figure shows the synergy that CloudSat, CALIPSO, and MODIS and AIRS sensors onboard AQUA satellite are capable of in viewing different properties of clouds and other atmospheric parameters near simultaneously. All members of the A-Train are Low Earth Orbiting satellites and have a period of ~99 minutes. They are also sun-synchronous orbiting with an inclination angle of 98.2 degrees. CloudSat is ~705km from the

Earth's surface. Each satellite makes ~14 trips around the Earth daily and also follow the same exact orbital path every 16 days.

The primary objectives of the CloudSat satellite mission are to help advance our understanding of the vertical structure, distribution, abundance, and radiative properties of clouds. All of these factors will help to determine the overall changes that clouds play in the hydrological cycle and both positive and negative feedbacks they present within the climate system (Stephens et al. 2002). CloudSat was the first millimeter wavelength cloud radar satellite launched. At 94GHz, the Cloud Profiling Radar (CPR) is able to observe both cloud and precipitating particles from a nadir view using active remote sensing technology. This frequency also allows both ice and liquid forms of precipitation to be detected.

This allows CloudSat to observe both the condensing and precipitating states of clouds from a nadir view. Both ice and liquid forms of precipitation can be observed is another key advantage of using millimeter wavelength radar. The sensor used on CloudSat satellite is called the Cloud Profiling Radar (CPR.) This sensor sends out 94GHz pulses and receives a return signal. At 94GHz frequency, it is able to detect more than 90% of ice clouds and more than 80% of water clouds (Stephens et al. 2002). Some of the key products produced by the CPR are the radar reflectivity, height profile, and also cloud mask. The cloud mask allows us to help remove noise from a typical picture view of CloudSat. Figure 3 shows an example of CPR reflectively profiles of a convective cloud. The cloud mask removed the noisy portion of the radar reflectivity data, which is then plotted against the height profile of the cloud.

### 2.1.2 TRMM

Global rainfall accuracy is still an unsolved issue. It is understood that the majority of rainfall globally rains within the tropics. To help fill this void of understanding rainfall and accurately measuring it, NASA and JAXA launched TRMM in 1998. In addition, this will help to improve tropical rainfall and their influence on global circulation (Kummerow et al. 1998). This satellite contains a variety of sensors that measure different parts of the storm as displayed in shown in Figure 4

Of these different sensors, the precipitation radar (PR) is the most valuable for this study since it provides us with a three-dimensional structural view of a storm. These measurements help provide information on the intensity and distribution of rainfall, and also the point at which snow changes over to rain. Since the PR is designed to measure rainfall, it is not sensitive to non-precipitating particles and is able to detect moderate to extremely heavy precipitating clouds. It is also designed just for the tropics as the name implies; it only captures measurements from 35N - 35S. The orbit time for TRMM is ~93 minutes at an altitude of 403km and makes ~16 orbits per day. Figure 5 shows an example of a TRMM radar reflectivity plotted as a function of both height and latitude.

### 2.1.3 ISCCP

Long-term climate models suffer from the unknown trends of clouds. Starting in 1983, the ISCCP dataset was created to help improve this and the understanding of clouds as a part of the World Climate Research Programme. The properties of clouds, diurnal, seasonal variation, distribution of clouds globally, and radiation balance are a few of the main objective of the ISCCP. Data is recorded every three hours starting from July 1, 1983 0:00 UTC.

In this study, we focus on the weather states derived from the ISCCP for the extended tropics. A total of 8 distinguishable weather states were found (Rossow et al. 2005). These weather states were derived using cloud top pressure and optical thickness of clouds measurements. The optical thickness could only be derived from the visible channel of satellites, which occurs in the daytime. This left out cloud data at night, which is vital towards studying the diurnal cycle of convective clouds. In (Tan and Jakob 2013), they were able to recreate the weather states utilizing the infrared channel only, allowing nighttime clouds to be utilized in their analysis. The infrared channel missed out on very thin clouds, but can handle convective clouds very well.

### 2.1.4 Dataset

This study was carried out utilizing data from CloudSat cloud profiling radar (CPR), Tropical Rainfall Measurement Mission (TRMM) precipitation radar (PR), International Satellite Cloud and Climatology Project (ISCCP) weather states that was derived from the ISCCP D1 dataset (Rossow and Schiffer 1999) and ISCCP IR-

Only Cloud Regimes (Tan and Jakob 2013). Similar orbital periods and different trajectories allow both CloudSat and TRMM satellites to overpass approximately 30 times daily. A collocated data of CloudSat and TRMM is available from CloudSat Data Processing Center at Colorado State University. There is some lag time between the satellites ranging from zero to 50 minutes a part in this dataset. In this study, we focus on the data for cases in which the separation time was from zero minutes to 20 minutes. Each intersection is approximately 220km in length and the data time period ranges from August 2006 to December 2010. The parameters utilized from CloudSat are cloud top height, radar reflectivity, cloud mask, longitude and latitude. Similarly, TRMM products used are radar reflectivity, longitude and latitude. The heights for TRMM PR data are not explicitly given, but it was noted that every bin was an interval of 250 meters. The ISCCP weather state data is available for the extended tropics (35N -35S) containing three hourly readings from July 1, 1983 till July 1, 2008 (Rossow et al. 2005). This dataset contains a weather state for every 2.5 by 2.5 degree globally every 3 hours. It also contains infrared (IR) data, which was utilized in creating IR only regimes of the weather states as well (Tan and Jakob 2013).

## **2.2 Contoured Frequency Altitude Diagrams**

An important tool utilized in analyzing CloudSat CPR and TRMM PR radar reflectivity data is contoured-frequency-altitude-diagrams (CFAD) (Yuter and Houze 1995). CFADs allow us to see a two-dimensional depiction of multiple frequency

distributions for a particular variable along a pre-determined height interval. In this study, 1km altitude bins from 1-20km are used for height, while 5dBZ bins from -30 to 55dBZ are utilized for the radar reflectivity values of CloudSat and TRMM.

To remove situations of clear sky only, filters are placed on CloudSat and also on TRMM data, independent of each other. In the case of viewing the data from a CloudSat perspective, a 1% threshold of valid data is implemented to remove clear sky cases, that is, at least 1% of the CloudSat cross section has to be filled with clouds. Locations where the cloud mask is at least 20 were considered valid since this meant that there was a very high confidence that a cloud existed at that location. In addition to this filter, both the first and last 50 profiles of CloudSat are removed since they do not correlate with any TRMM data as stated in the collocated dataset manual. Figure 6 shows a footprint of both CloudSat and TRMM and the excess footprint that CloudSat contains. Similarly, a 0.1% threshold of TRMM is required, which is determined by the number of points that contained 18dBZ or higher readings over the total area. The 18dBZ threshold is used since that is the minimum valid reading of TRMM (Simpson et al. 1999). Using the cloud mask of 20 on CloudSat, all radar reflectivity values that did not have this minimum cloud mask value are assigned not-a-number value. This removes noise and invalid signal readings of CloudSat radar reflectivity data. Similarly, for TRMM, radar reflectivity values under 18dBZ are assigned not-a-number value to remove the poor signal reading and invalid data as well. For each case, the number of counts of valid dBZ and height are added and normalized per altitude, such that sum of the corresponding altitude will be 100%.

Figure 7 shows an example of a valid CloudSat case being transformed into a CFAD. On the left the figure, valid radar reflectivity and height is plotted, and on the right, the CFAD of the same case is displayed. For each height interval of 1km, the sum of the frequencies adds up to 100%. Contour lines are drawn to connect similar frequencies among different altitudes.

To prepare for the K-means analysis, for every valid case, CloudSat CFAD and the collocated TRMM CFAD are saved in a 1-D array. The 1-D array contains the same data and information that the 2-D version does. The same is done for every valid TRMM case; TRMM CFAD and its corresponding CloudSat CFAD were saved as well. Then, a 2-D array for both CloudSat and TRMM is created where along one dimension contained the CFAD and the other dimension can be seen as time, or total number of instances within the entire time period that does not violate the condition defined for clear skies.

### 2.3 K-Mean Clustering Analysis

K-means clustering algorithm (Anderberg, 1973) allows us to interactively group data in such a way that they have common traits among them. One first must select a certain amount of clusters,  $K$ , which then find the closest sample data points. The position at which each cluster center is located originally is at random. Once each data point has been assigned a cluster number, the average distance from each cluster center, hereby centroid is computed. This average becomes the new centroid locations and each sample point is reassigned to the closest centroid. This process is

repeated until the minimum Euclidean Square Value is achieved between the centroids and their corresponding points. If the process is continued, the centroid locations will not change, therefore the sample points will not be regrouped.

Previous clustering analyses were performed on CloudSat, TRMM, and ISCCP data (Boccippio et al. 2005; Elsaesser et al. 2010; Jakob 2003; Mace et al. 2007; Rossow et al. 2005; Tan and Jakob 2013; Zhang et al. 2007). These previous studies are used to help derive a starting number of K clusters in this study. For the CloudSat perspective of the analysis, four and five clusters analysis are tried. For the TRMM perspective, three and four clusters are tried. Since the starting centroid location varies may yield slightly different results, there were 15 iterations done to ensure the absolute minimum Euclidean Square Value is achieved. Each cluster analysis gave an output of the overall average CFAD for each cluster. These outputs are performed for both CloudSat and TRMM, regardless of the initial perspective. For each trial, the cluster's relative frequency of occurrence is computed as well. The relative frequency of each cluster's geographic location is also calculated for each individual cluster.

In addition to the CFAD framework, each cluster is further analyzed by comparing the cloud top height and the height of the maximum echo for each individual profile per cluster. The distribution of radar reflectivity within TRMM for each cluster is also calculated.

## 2.4 ISCCP Collocation

The collocated CloudSat and TRMM radar reflectivity clustered data is also paired up with ISCCP data. Figure 8 shows the distribution of all the CloudSat-TRMM collocated cases. Due to a high concentration of the data located between 30-35 north and south latitude because of TRMM's highly inclined orbit, these data points are removed from the study to reduce a potential bias towards these regions. Figure 9 shows a revised distribution pattern in regards to the data of interest. Each CloudSat-TRMM collocation contains latitude and longitude data, which can be paired up with a corresponding weather state from ISCCP. Rounding to the CloudSat-TRMM data time to the nearest factor of 3 hour is necessary to finding the correct hour within ISCCP. Once this is done, we can then match where the CloudSat-TRMM collocation is to the corresponding 2.5 x2.5 degree ISCCP weather state. Weather states that are clear skies, nighttime or invalid readings are ignored during this study. In addition, the condition of CloudSat containing at least 1% valid cloud data and TRMM containing at least 0.1% precipitation data is also included. Table 1 shows an overall distribution of data availability after different constraints were implemented in the study. With relation to the original eight ISCCP weather states relative frequency of occurrence over 25 years for the extended tropics (Rossow et al. 2005), a comparison to the breakdown of these collocated CloudSat-TRMM with ISCCP is also done. Figure 10 shows the breakdown of this relative frequency of occurrence comparison.

In addition to comparing the relative frequency of occurrence, analysis of each cluster's individual weather state distribution is performed. This allowed us to see which weather state(s) dominated a corresponding cluster from the previous clustering analyses performed. In addition, the inverse was performed as well, that is the relative frequency of cluster within each individual weather state.

## 2.5 ISCCP Regimes in Infrared

Initial clustering analysis of ISCCP data was based on optical thickness of clouds in the visible and IR spectrum. This biases the diurnal sampling, because all of the nighttime cases were not accounted for since visible spectrum is not available at night time. In understanding convection it is important to see both day and night due to the diurnal cycle of convective systems [Yang and Slingo, 2001; Tian et al., 2004]. To overcome this difficulty, an analysis of the ISCCP data (Tan and Jakob 2013) was done using the infrared spectrum only. This allowed cases at night to be of use. Using the IR regime data, we paired each CloudSat-TRMM collocation with their corresponding IR regime. Once again, rounding of the CloudSat-TRMM interaction data time to the nearest factor 3 hour was necessary to find a corresponding match. A comparison between the regimes and the CloudSat-TRMM clusters is done.

## Results

Over the time period spanning from August 2006 till December 2010, there are a total of 38,395 CloudSat – TRMM collocation recorded within 50 minutes of each other. For this study, to improve the accuracy of these intersections, only cases within 20 minutes of each other were used and only from 30N-30S. This yielded total of 10,111 collocated CloudSat and TRMM cases. After filtering out the cloud free cases with CloudSat, 55.48% (or 5610 cases) is used in the k-means analysis. For filtering using TRMM only, 28.73% (or 2905 cases) of the collocated files are utilized in the k-means analysis.

Viewing the data from each perspective individually, different numbers for the ideal number of clusters are obtained. Using TRMM alone, we found that 3 clusters were ideal as displayed in Figure 11. The first cluster appeared the least frequently at 21.86%. It contained the tallest clouds vertically with also the largest radar reflectivity values as displayed in on the left in Figure 11. On the right in Figure 11, the highest cloud top heights for the first cluster and largest radar reflectivity values out of the three subplots are displayed. The second cluster appeared the second least frequent at 35.31% of the time. The CFAD for this cluster within Figure 11 on the left shows an average height that was more in the medium range near 8-9km. The third cluster appeared the most frequent at 42.82%. The CloudSat view in Figure 11 shows a mixture of different types of clouds. First it shows at high altitudes very low radar reflectivity and a variety of readings at lower altitudes. However, the complementary view with TRMM shows low altitude and

not as intense as the previous two clusters in terms of intensity with respect to radar reflectivity.

On the other hand, using CloudSat for the K-means analysis, we found 4 clusters to be the ideal amount as displayed in Figure 12. As with the previous result with using TRMM, the first cluster appears the least frequent at 13.33% and also the tallest clouds and the largest radar reflectivity values. The second cluster appears the next least frequent, showing very similar characteristics of the second cluster of Figure 11.

Focusing in on the cluster of 4 from CloudSat point of view (Figure 12), the locations each case per cluster is shown in Figure 13. In the first cluster we see that they are mostly prevalent in four regions: Central Africa, Southeast Asia, and over Central America and the northern part of South America. The second cluster locations are similar to those of the first cluster, except more spread out over the tropics. There are some notable regions where the second cluster is largely absent such as the northern part of Africa, and off the west coast of South America, Central America Africa, and Australia. The third cluster is almost the complete opposite of the second and is concentrated off the west coast of large landmasses such as South America, Central America, Africa, and also Australia. The fourth cluster appears in nearly all the areas of the tropics with a few notable absences off the west coasts of South America and Africa.

In Figure 14, each cluster is further examined individually. Within each cluster, the profiles from all of the cases are analyzed and maximum echo top height

(ETH) and cloud top height (CTH) were extracted and plotted. In addition, the concentration of data points is marked up with contour lines. The maximum ETH versus CTH scatter plot presents a different depiction of cloud vertical structure than CFAD. The first cluster showed strong concentration of echo top height readings at high altitudes. The second cluster shows a more uniform concentration of echo top height readings with a reasonable amount in high, mid and low areas. The third cluster is mainly concentrated at a very shallow region of about 2km. The fourth cluster shows peaks in both high and low altitudes. The largest peak is located at the higher altitudes.

A comparison between the clusters produced through k-means analysis and a previously concocted clustering analysis on the patterns of cloud property distribution at mesoscale from ISCCP was done next. The eight previous found weather states distribution for the extended tropics data (Figure 15) overlapped the collocated CloudSat and TRMM data from August 2006 till June 2008. In Figure 10, the first three weather states appear more frequently in CloudSat-TRMM intersecting in comparison to ISCCP when in conjunction with each other. Weather states 4, 5, and 6 were all relatively the same amount, while weather states 7 and 8 appears to be more prevalent in ISCCP.

Figure 16 shows a breakdown of the location of the CloudSat-TRMM cases collocated with ISCCP weather state. The first cluster is mostly over Central America, Southeast Asia and over the tropical African region. The second cluster appears over the Atlantic basin, Southeast Asia, and in the Indian Ocean. The third cluster appears most frequently off the west coast of South America, west coast of

Australia, and also over mainland China. The fourth cluster appears scattered around the world with a notable absence off the west coast of South America and also west coast of Africa. In addition, Table 1 displays the distribution of collocated ISCCP and CloudSat-TRMM data that is used. There is a lot less available since the overlap is only 23 months in comparison to the 53 months of CloudSat-TRMM collocated data available.

In taking a more detailed look at the comparison between ISCCP and CloudSat-TRMM clusters, we examined the breakdown of weather states for each individual cluster. In Figure 17, the first cluster is dominated by WS1 and WS3, both related to convective cloud regimes (Rossow et al. 2005). WS5, WS6, and WS7 are all near non-existent in this cluster. The second and third clusters are both dominated by WS3 and WS8, but with a noticeable difference in other WS. WS1 and WS2 are prevalent in cluster 2, but near non-existent in cluster three. WS5, WS6, and WS7 are also noticeably near non-existent in cluster two, but very relevant in cluster three. The fourth cluster is dominated by WS3 and WS8 as well and near even in WS1, WS5, WS6, and WS7. Each cluster shows a strong presence of WS8, which is not surprising since this is the most dominate WS in ISCCP.

Inversely, we analyze the breakdown of the CloudSat-TRMM clusters within each weather state as shown in Figure 18. Of the eight weather states, WS1 stands out with a relatively low amount of observations for cluster 4. The first and second clusters of CloudSat-TRMM clustering analysis dominated this WS. In WS2, the second cluster and fourth dominate this cluster. WS3 shows cluster 4 as most prevalent with clusters 2/3 both nearly the same amount. In WS4, cluster 4 and

cluster 2 once again dominates similarly to WS2. In WS5, by far cluster 3 dominated this with over 70% belonging to this weather state. Similarly in WS6, WS7, and WS8, cluster 3 stands out with over 50%, 40% and over 30% respectively for each weather state. Cluster 4 is near even in WS6, WS7, and WS8 in terms of its relative frequency of occurrence for each weather state.

The IR-only regimes from Tan and Jakob (2013) are also analyzed in comparison with the clusters derived from CloudSat-TRMM. These regimes filled a big void that was missing from the ISCCP visible weather states, because IR data are available during both day and night. Each cluster was displayed showing the percentage of matches for each IR regime. In Figure 19, the first cluster is mainly of the cirrus clouds (CC) and mid-level congestus clouds (IM) regime. Another note is that this cluster contains the highest percentage of the deep stratified clouds (CD) regime of the four clusters. The second cluster is mainly composed of (trade cumulus clouds) ST and a near even amount of CC and IM. Over 50% of the third cluster is composed of ST and stratocumulus (SS1), while the fourth cluster is ST mostly.

## Discussion

Our results show that using CloudSat for clustering analysis yielded nearly twice as many cases as using TRMM. This is no surprise since CloudSat is designed to be sensitive to cloud size particles while TRMM is designed to detect precipitation sized particles. Clouds that precipitate account for a small fraction of all clouds in the atmosphere.

Clustering analysis that anchors on TRMM results in three clusters. This broadly agrees with previous clustering analysis done on TRMM (Elsaesser et al. 2010), even though this was a joint-analysis invoking CloudSat. In both, there were three modes of precipitation for TRMM found. As shown in Figure 11, these modes are deep convective, cumulus congestus, and shallow precipitating clouds.

Clustering analysis based on CloudSat produces a total of four clusters. This is slightly different from previous clustering analysis using CloudSat such as Zhang et al. (2007), who found five clusters. The CloudSat-based cluster analysis showed three clusters that are associated with different precipitation types and one cluster that corresponds to non-precipitating cloud type. The decision to stay with four instead of three or five clusters was based on criteria that judge the outcomes of these clusters (Rossow et al. 2005).

Each cluster shows characteristics of different cloud types as classified by ISCCP as shown in Figure 20. Each cluster had a lot of WS8 and WS3. The first cluster in Figure 12 relates to deep convective and nimbostratus clouds. These clouds can be characterized by their large radar reflectivity and high cloud top heights, which are continuous from near surface. For this cluster, the corresponding

weather states from ISCCP that dominated are WS1 and WS2. WS1 accounted for over 30%, which is a deep convective weather state. WS2 is mostly cirrostratus, which are cirrus anvils attached towards deep convective cloud types. Of all the TRMM-CloudSat data, this cluster only corresponding to 13% of the data, making it the least frequent of the four clusters. This cluster also pertained to regions of rising air masses as shown in the NCAR reanalysis in Figure 21.

The second cluster in Figure 12 is correlated to both WS1 and WS2 as well, but not as much as the first cluster. This shows this cluster also is a convectively active cluster, but not as much as the first cluster. The cloud types here likewise are very similar in both nimbostratus and deep convective types. These clouds can also be seen as cumulus congestus cloud types as described in Johnson et al. (1999) and Luo et al. (2009). To better understand the difference, it is noted that the overall CFAD for this cluster cloud top heights peak lower than the first cluster. Geographically, they appear in the same regions as cluster 1, but a little more spread out. As with the first cluster, this was located in regions of general rising air masses. This cluster appeared the second least frequent at 21% of the total.

The third cluster in Figure 12 corresponds to all the shallow type of clouds. Cumulus, stratocumulus, and stratus clouds are the most prevalent cloud types in this cluster. There is a lot of light precipitating, and a lot of low non-precipitating clouds in this cluster. This cluster corresponding to regions where there is sinking rather than rising air mass, which explains why the cloud top heights are all shallow. They were mostly found off the west coast of large continent mass such as Africa,

South America, Central America, and Australia. These clouds appeared the most frequent out of all the cloud types at over 33%.

The fourth cluster in Figure 12 is related to cirrus, altocumulus, and cumulus cloud type. The most common trait in these cloud types is the low optical thickness in all height levels. Precipitation isn't associated with these cloud types. These clouds appeared the second most frequently among the four clusters at 31%. However this is comparable to the most, which was cluster 3 at 33%. These clouds can be seen globally at a near equal rate.

## Conclusion

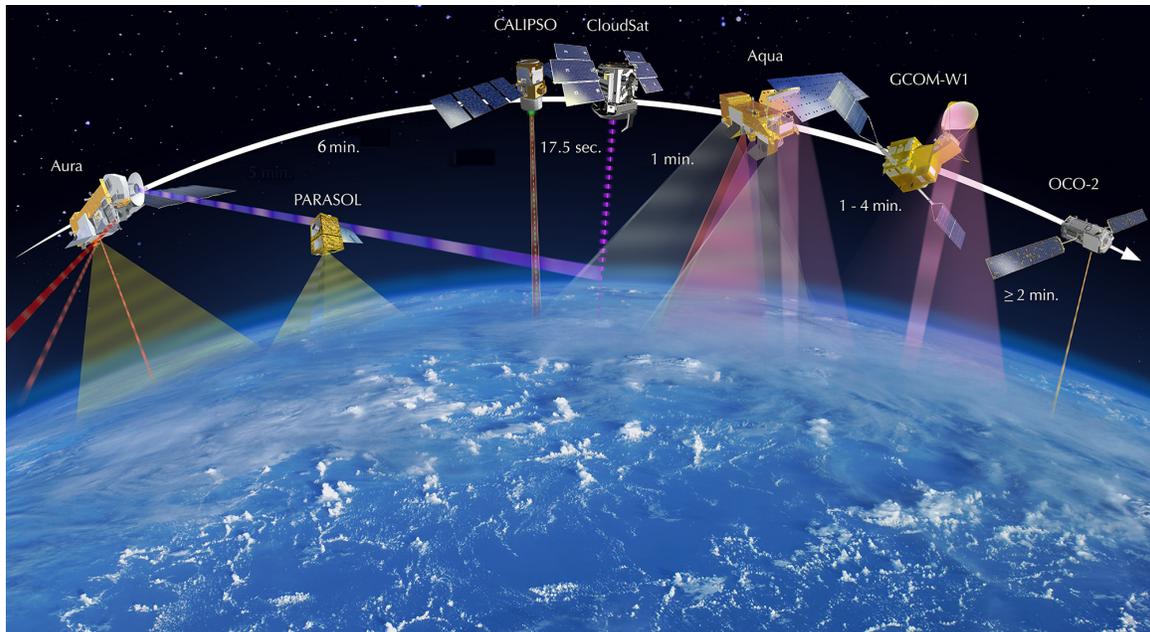
Clouds are an important component of the earth-climate system and play a critical role in affecting energy and water cycle of the planet. In this study, we focus on exploring synergy between CloudSat and TRMM radar reflectivity data in a CFAD (Contoured Frequency by Altitude Diagram) framework and apply a clustering analysis to identify distinct clusters. Jointly, four clusters are found to be optimal. Each cluster showed different characteristics of tropical cloud types. Three of the clusters are precipitating, while the fourth is non-precipitating cloud type. The three precipitating clusters correspond to deep convective, congestus, and shallow cloud types. Comparison with NCAR-NCEP reanalysis data shows that regions of shallow precipitating clouds are mostly associated with sinking air motion, while deep convective and congestus cloud types were present in regions of rising air motion.

Comparisons between CloudSat/TRMM cloud clusters and ISCCP Weather State (WS) data are conducted. Results lend further support to the conclusion that the first two CloudSat/TRMM clusters are convective. ISCCP tends to show lower occurrence frequency of convective cloud types, while CloudSat and TRMM show the non-convective cloud type less frequently. These differences are attributable to the different spectra and remote sensing technique used.

## **Acknowledgments**

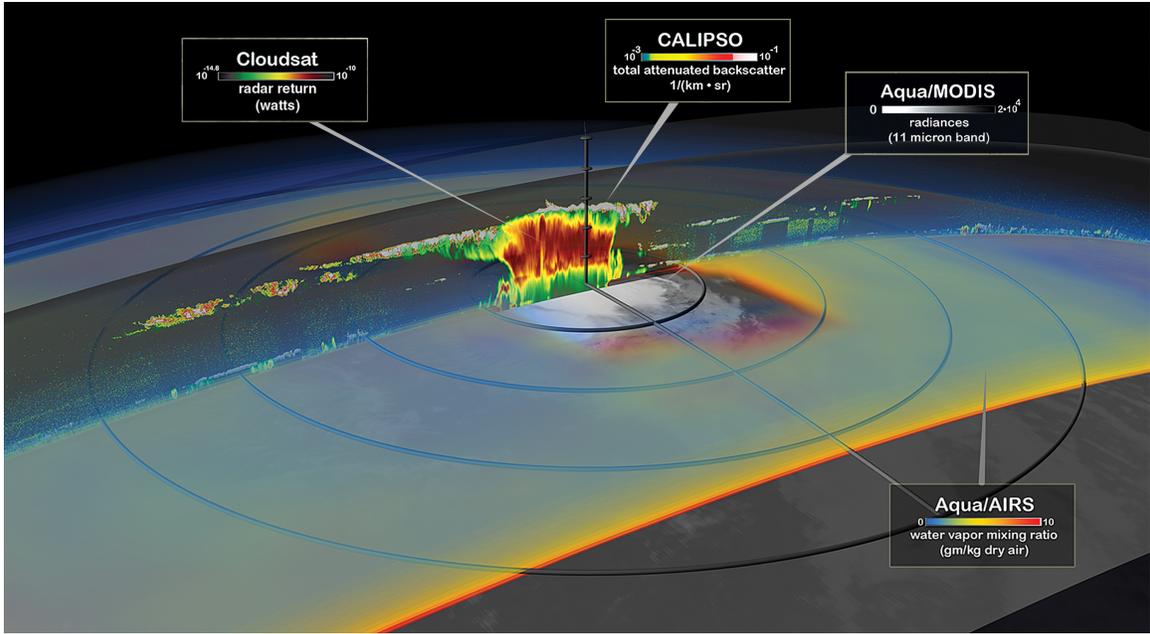
I would like to personally thank Dr. Z. Johnny Luo for his guidance and advice through this project and also graduate studies. His generosity, knowledge and support from beginning to end was remarkable and a great experience. I would also like to thank NASA for funding as well throughout this project, allowing me to obtain valuable new skills and knowledge to use in my future career. Also, I would like to thank all my group members for their support and valuable input through this project.

## Figures



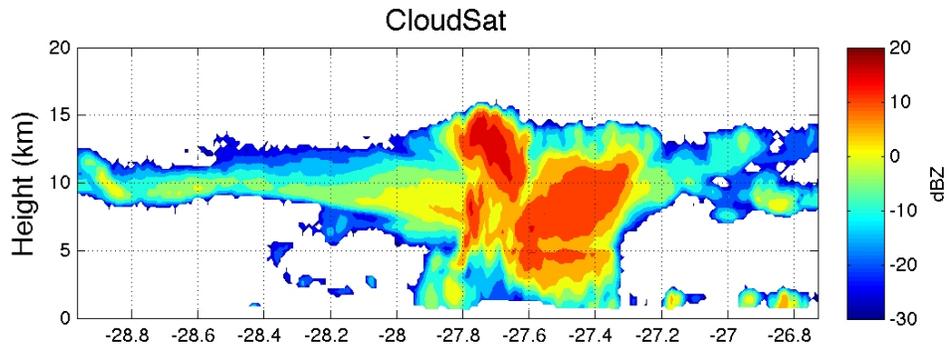
**Figure 1-** Members of the Afternoon Constellation (A-Train) as of June 2011.

Credits: [http://atrain.nasa.gov/historical\\_graphics.php](http://atrain.nasa.gov/historical_graphics.php)

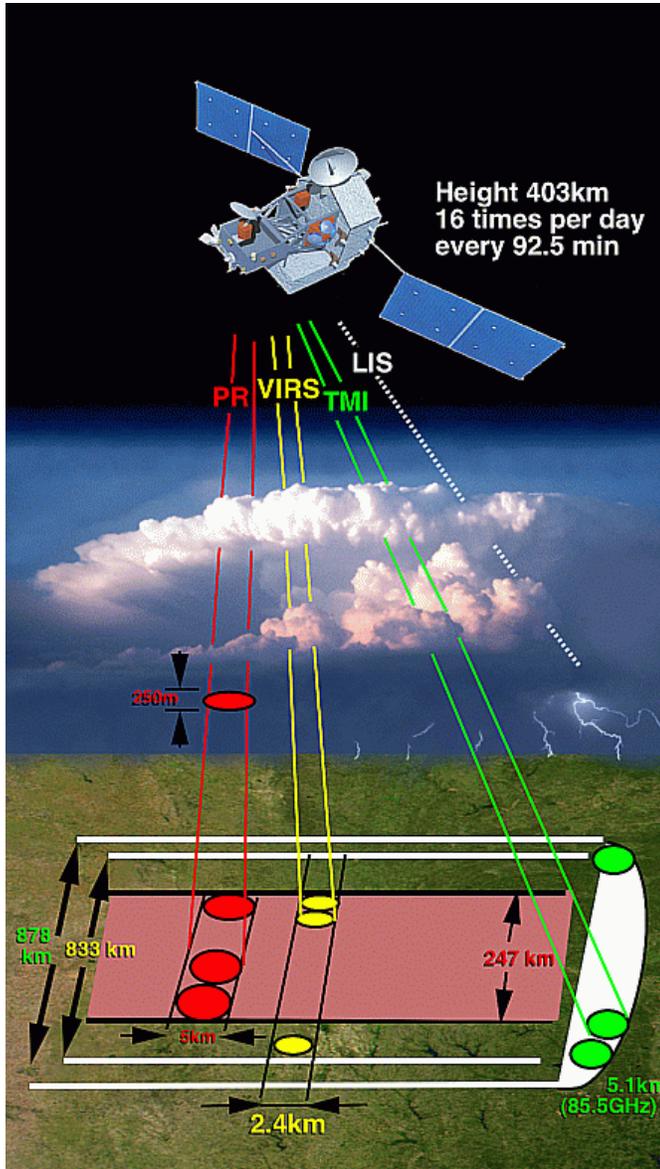


**Figure 2** Tropical Storm Debby (2006) being observed by different members within the A-Train. MODIS sensor provides the optical thickness of the clouds present in the storm while CloudSat and CALIPSO provide a vertical slice of the storm allowing us to view the internal structures that composes the center of this tropical storm.

**Credits:** <http://atrain.nasa.gov/images.php>

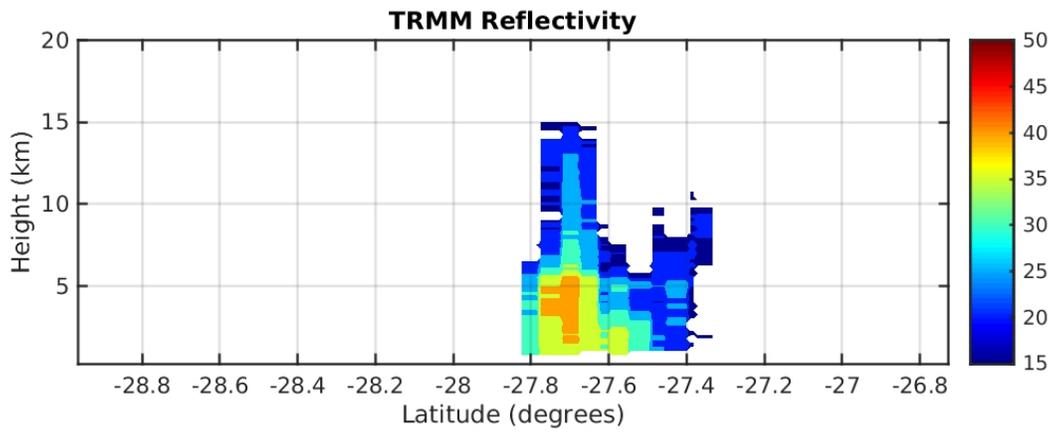


**Figure 3** An example of CloudSat radar reflectivity data plotted along with the height profile of the cloud.

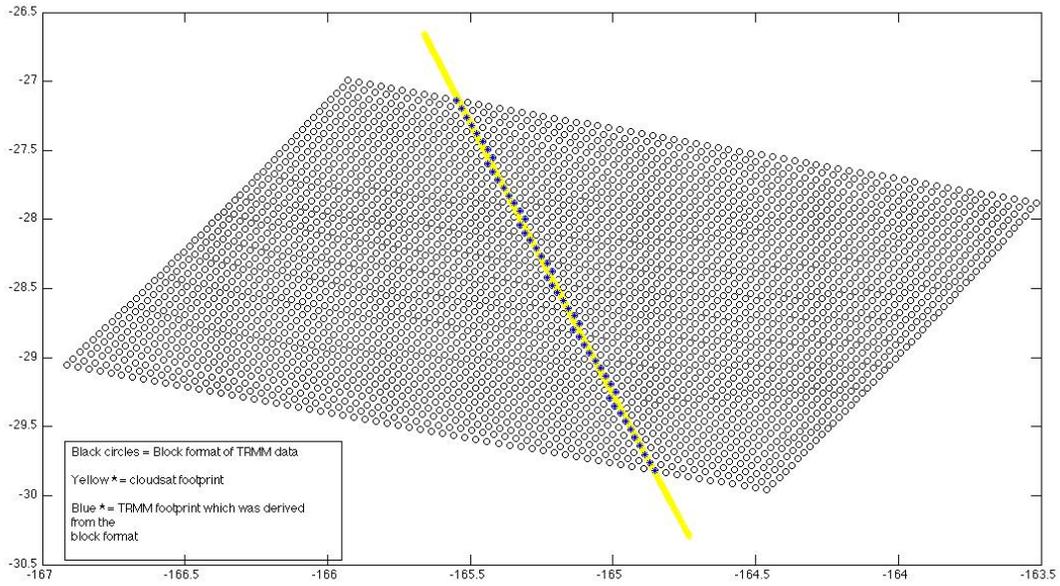


**Figure 4** TRMM satellite and its sensors.

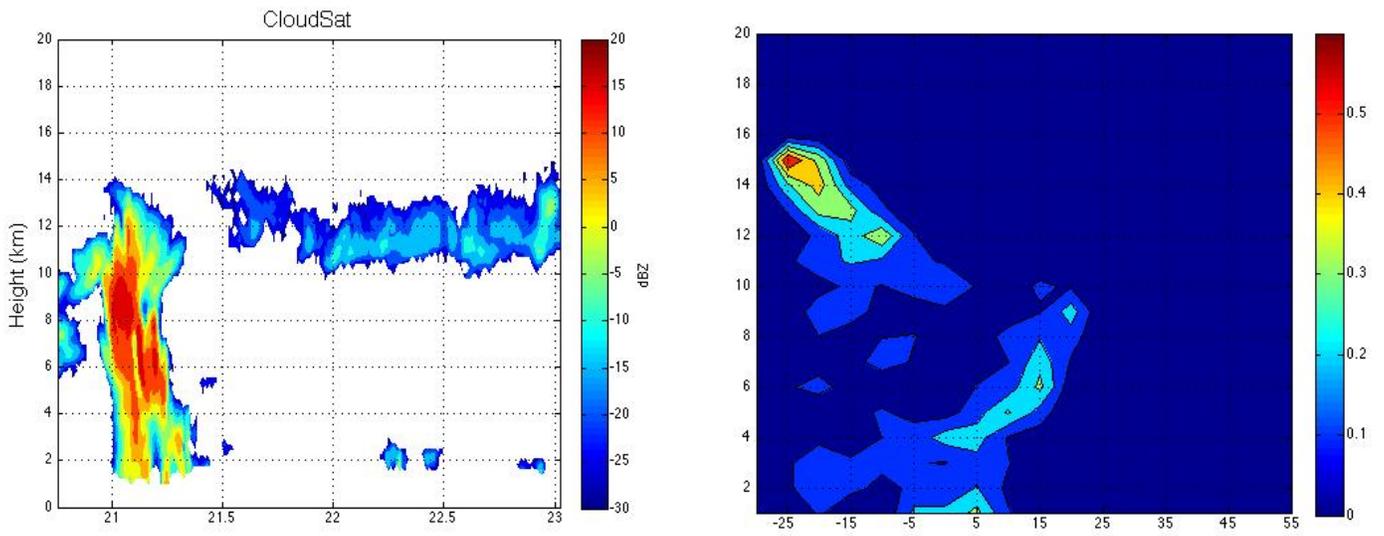
Credits: [http://trmm.gsfc.nasa.gov/overview\\_dir/background.html](http://trmm.gsfc.nasa.gov/overview_dir/background.html)



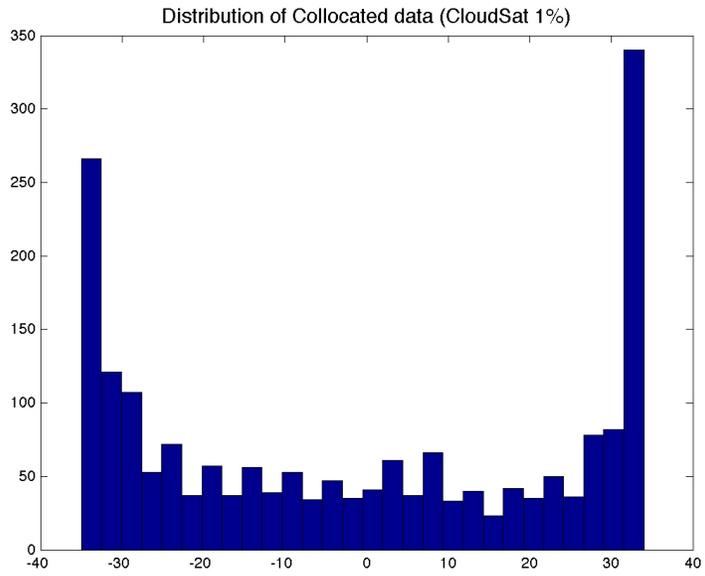
**Figure 5** TRMM radar reflectivity data being plotted against height.



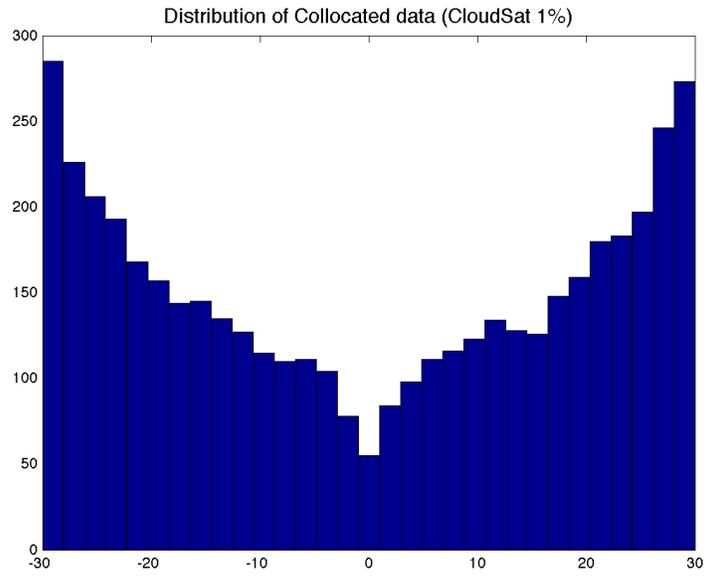
**Figure 6** TRMM and CloudSat collocated. Yellow marks CloudSat footprint, while the black region is TRMM.



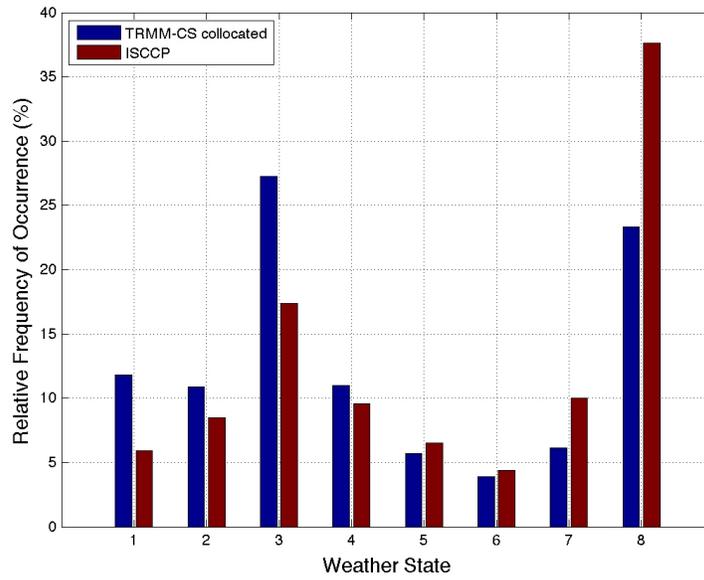
**Figure 7** A valid CloudSat case and its corresponding CFAD. The figure on the left displays the raw radar reflectivity and height readings, while the figure on the right shows how it reorganized using CFAD.



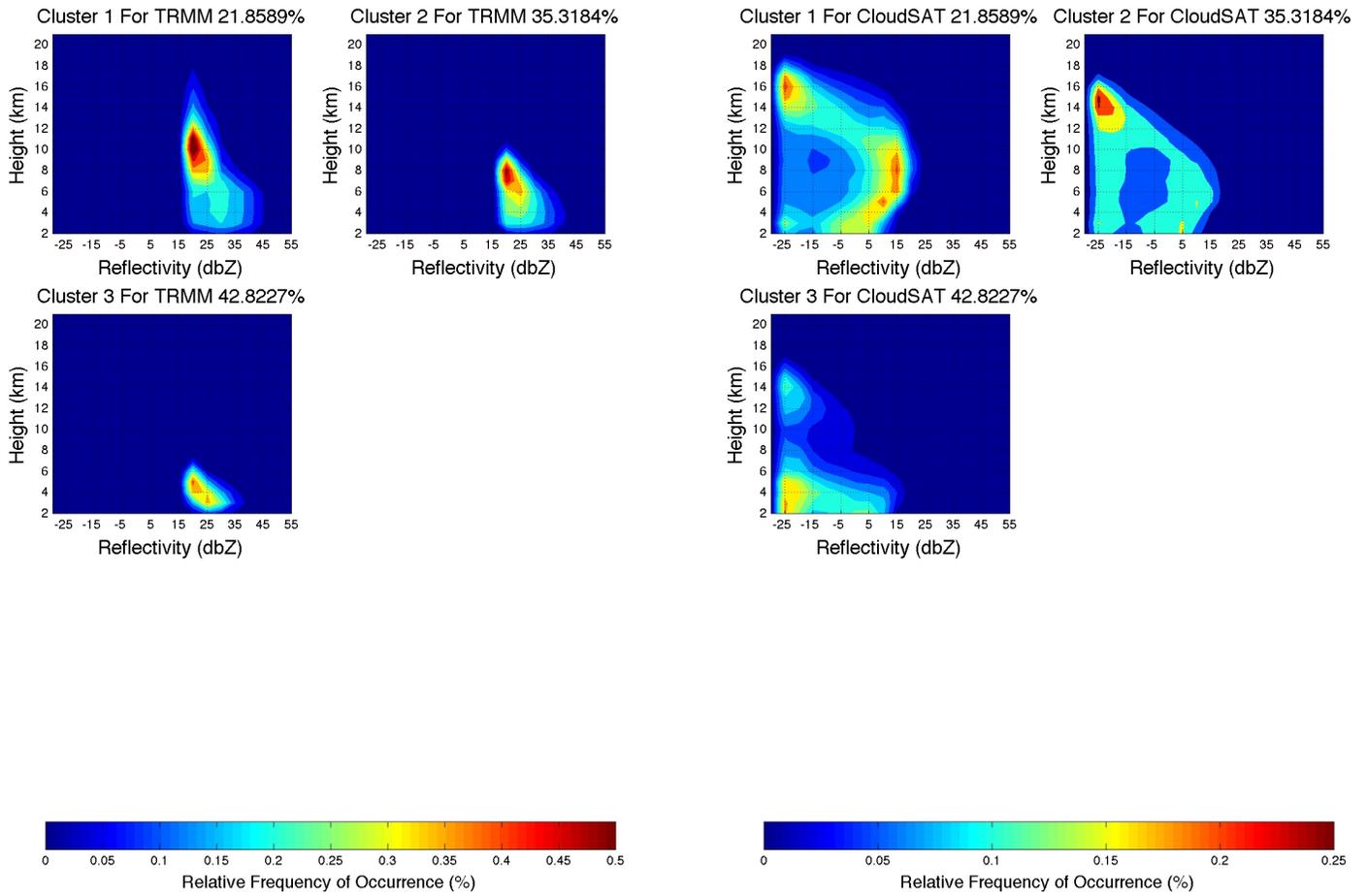
**Figure 8** Distribution of CloudSat-TRMM files with ISCCP from 35N-35S The majority of the cases fell between 30-35N/S



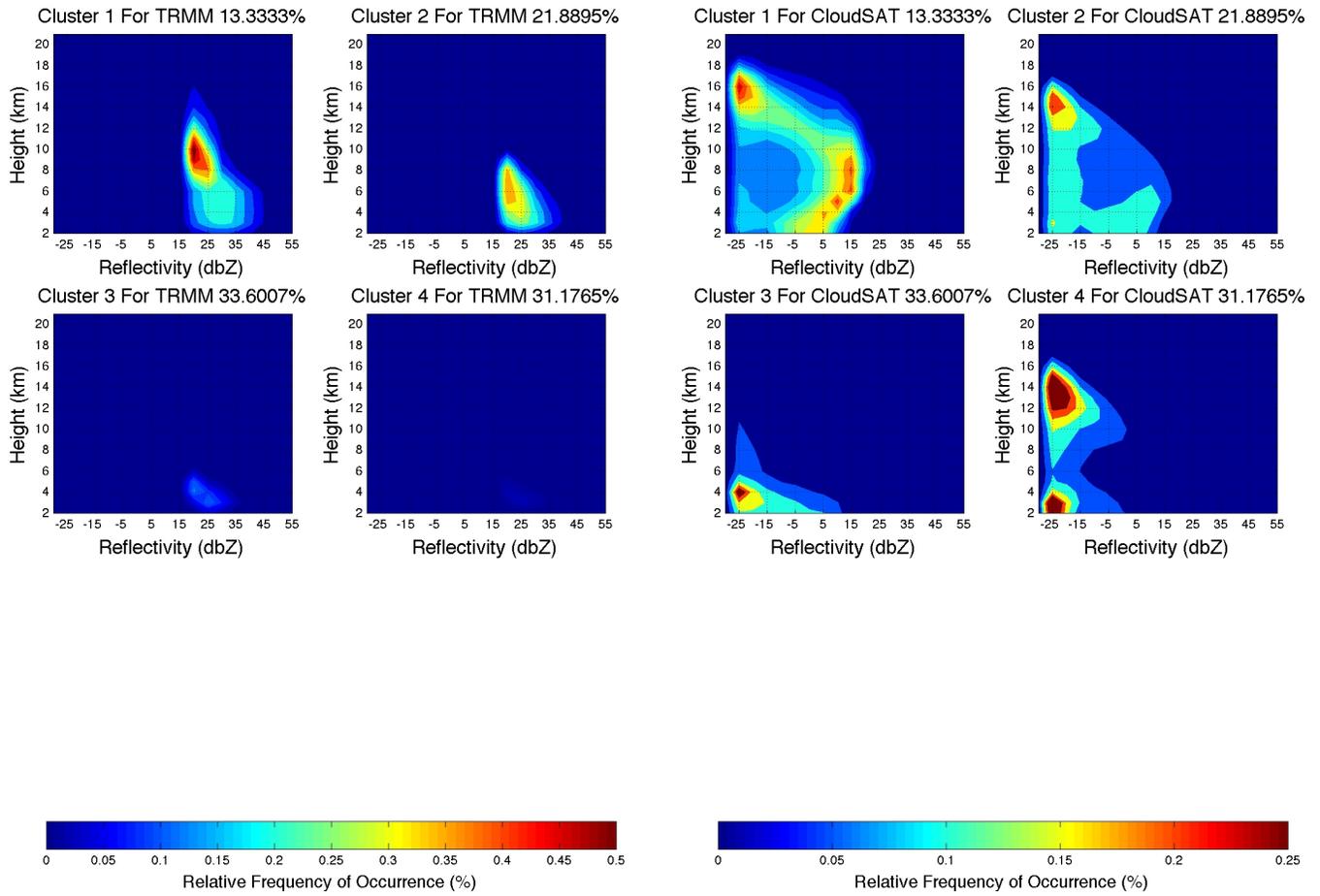
**Figure 9** Revised distributions of collocated CloudSat-TRMM data used with ISCCP.



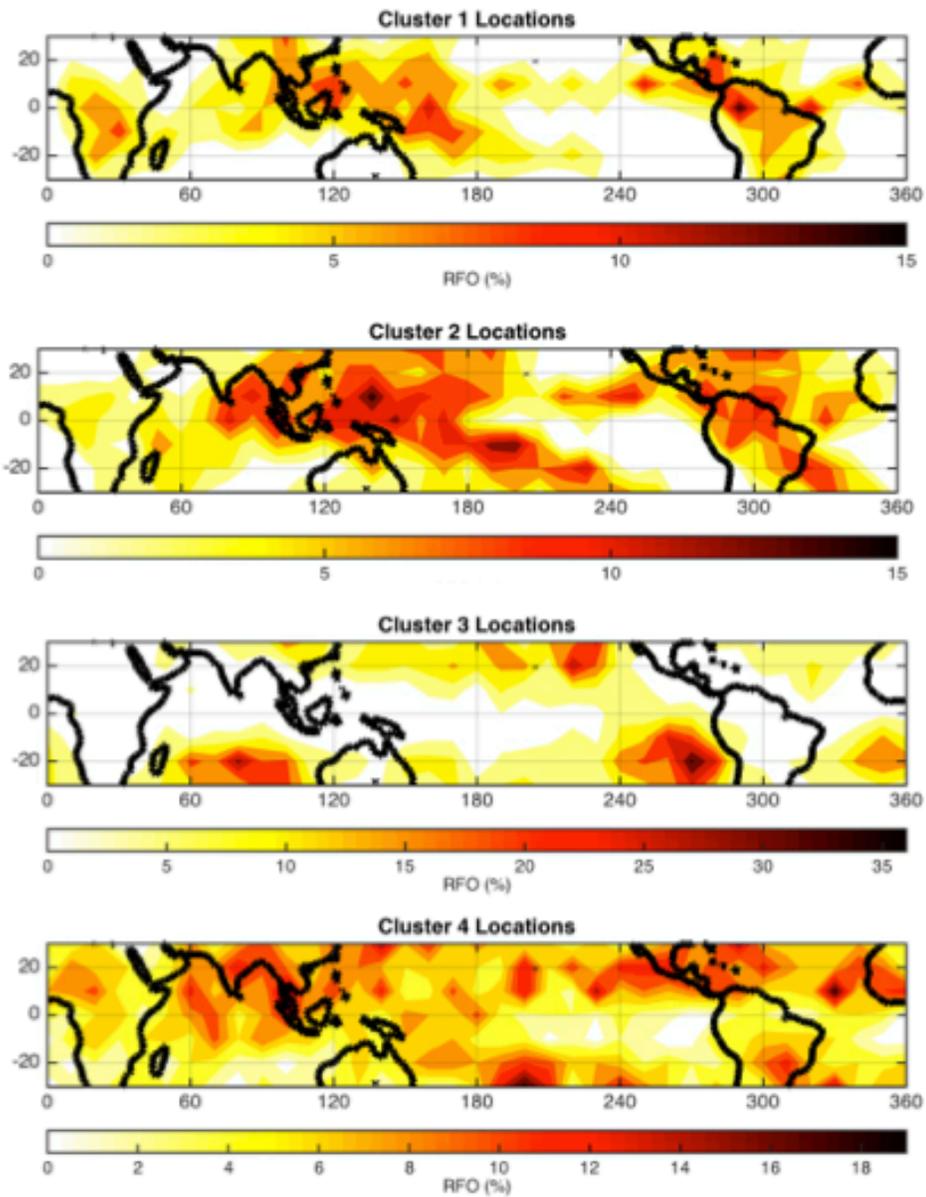
**Figure 10** A comparison of the frequency of occurrence for each weather state between the ISCCP and CloudSat-TRMM intersection cases. In the first three weather states, CloudSat-TRMM sees a lot more frequent cases, while weather states 4,5,6 are near identical. Weather states 7 and 8 are dominated by ISCCP.



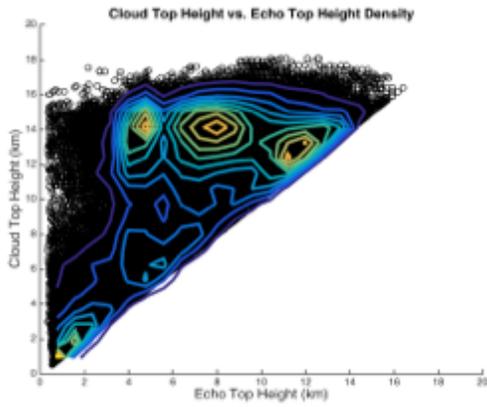
**Figure 11** -Three clusters from TRMM point of view. The figure on the left is the CFAD for CloudSat, while the figure on the right is the CFAD for TRMM data.



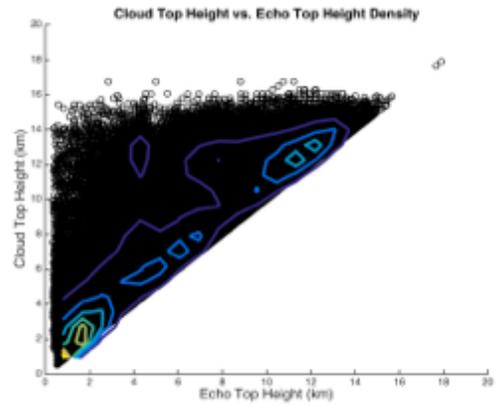
**Figure 12** Four clusters from CloudSat point of view. The figure on the left is the CFAD for CloudSat, while the figure on the right is the CFAD for TRMM data.



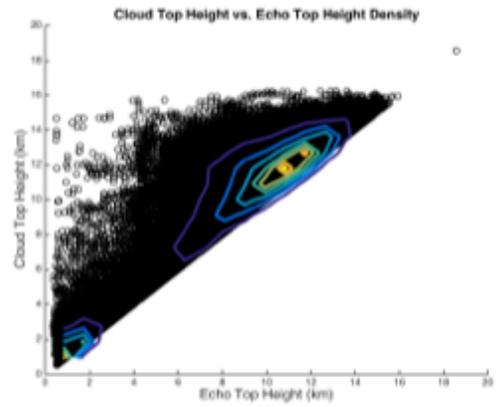
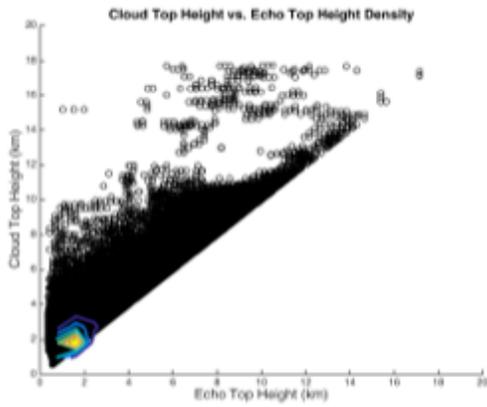
**Figure 13** Global distributions from 30N to 30S in 10 -by- 10 box that represents relative frequency of occurrence for each cluster.



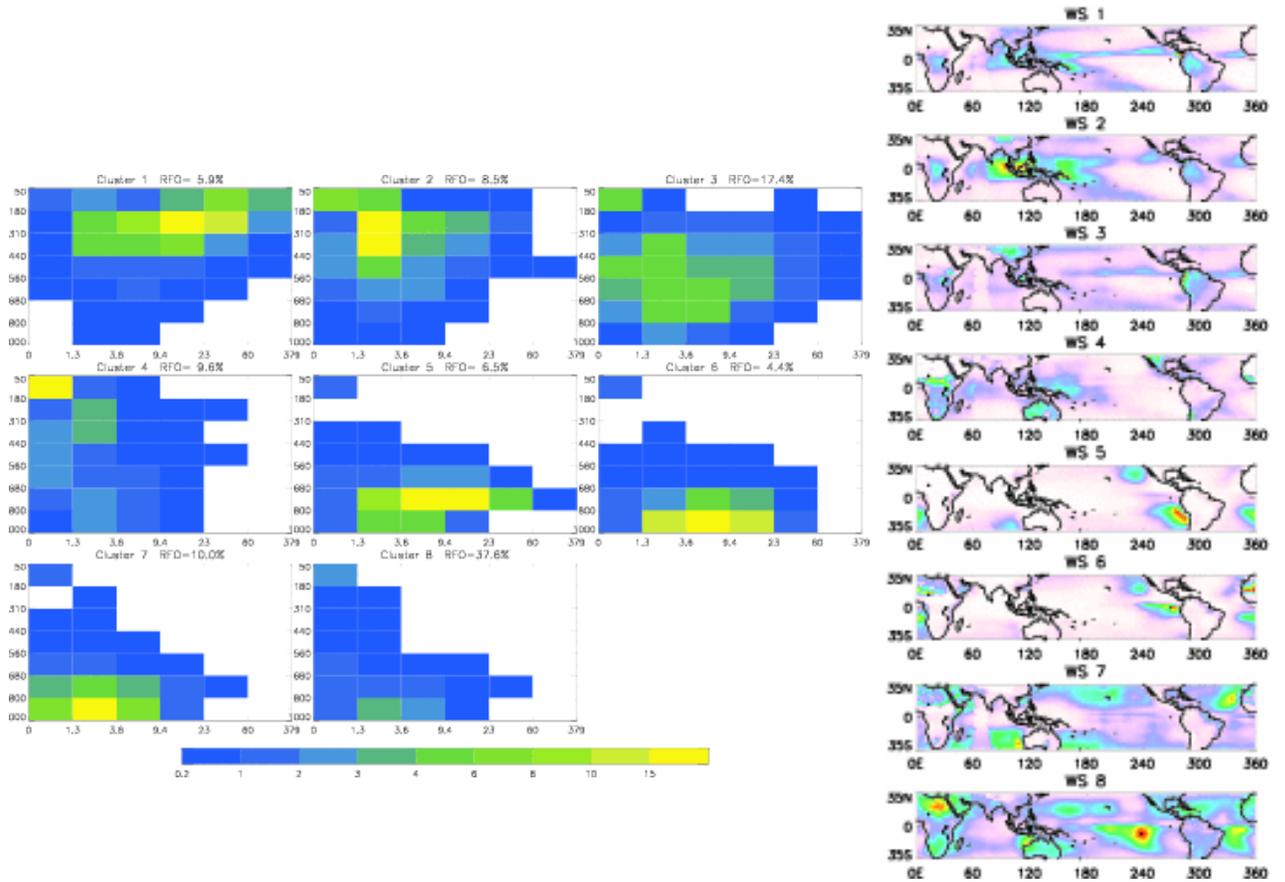
**Cluster 3**



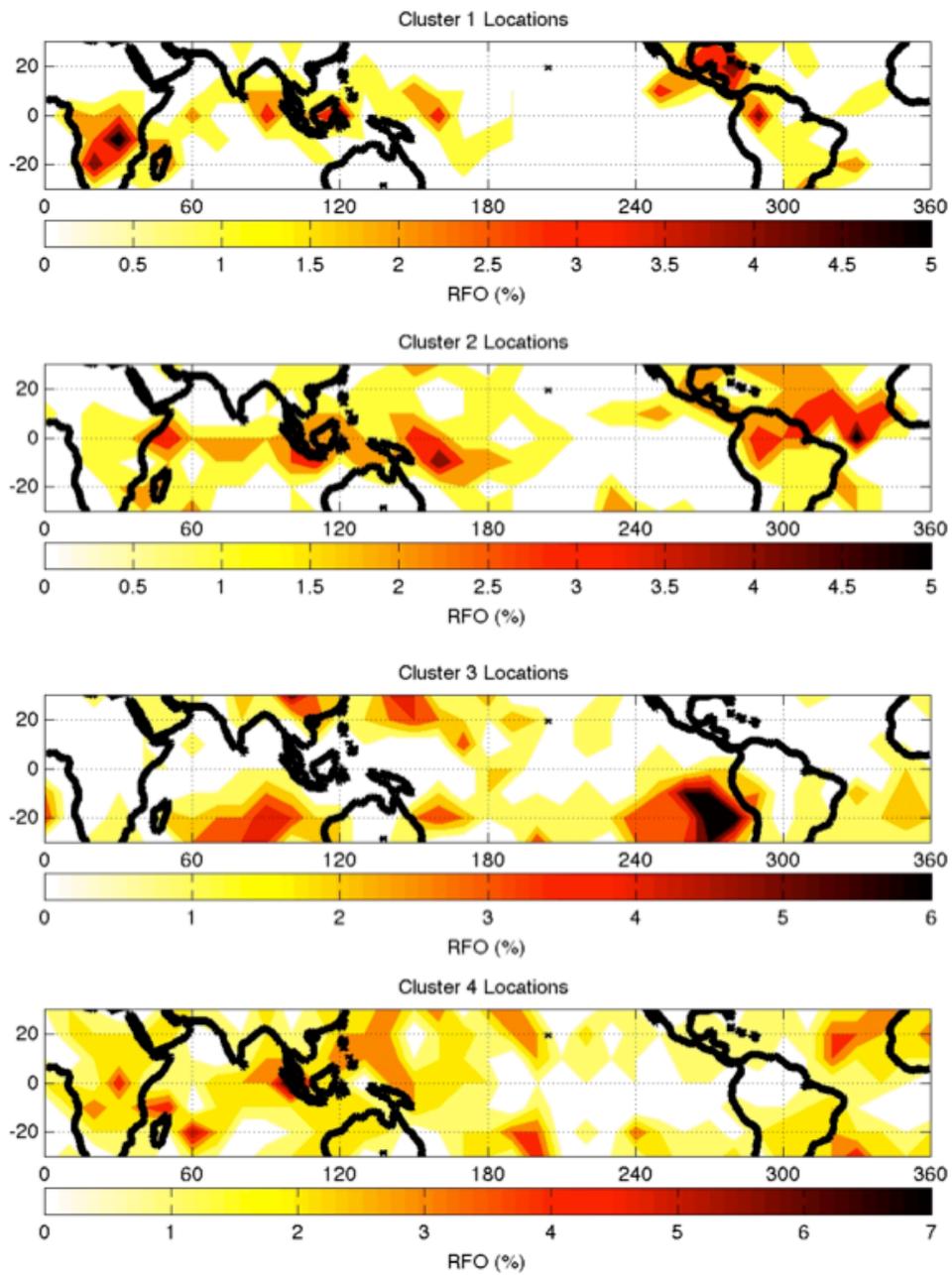
**Cluster 4**



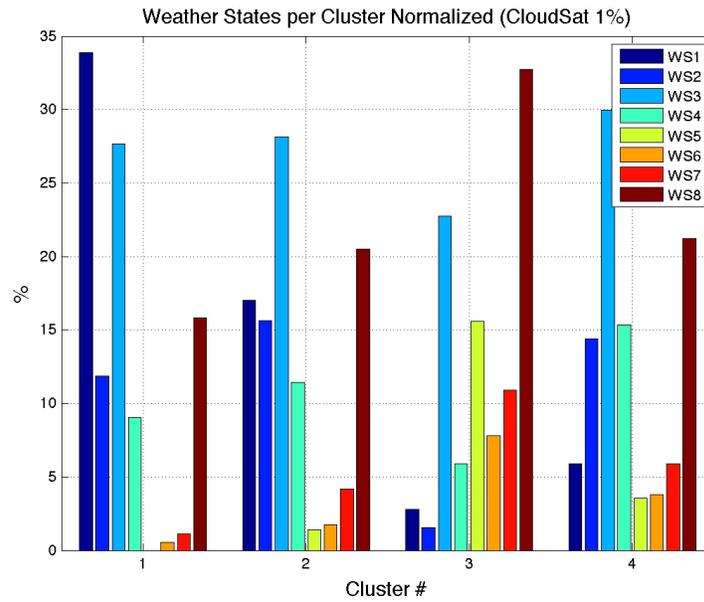
**Figure 14** Echo top height of profiles from CloudSat for each cluster. On top of each are contour lines, which represent the density of all the profiles collectively per cluster.



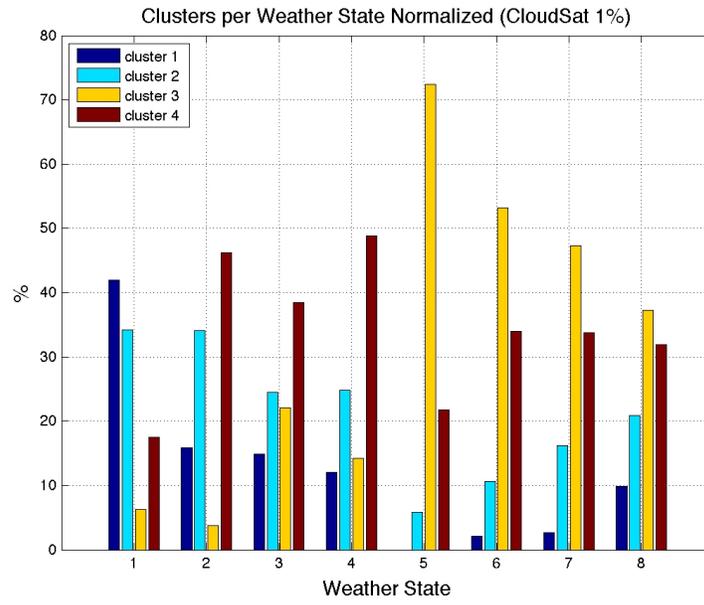
**Figure 15** ISSCP Weather State Distribution from 1983-2008 for the extended tropics. (<http://isccp.giss.nasa.gov/etcluster.html>)



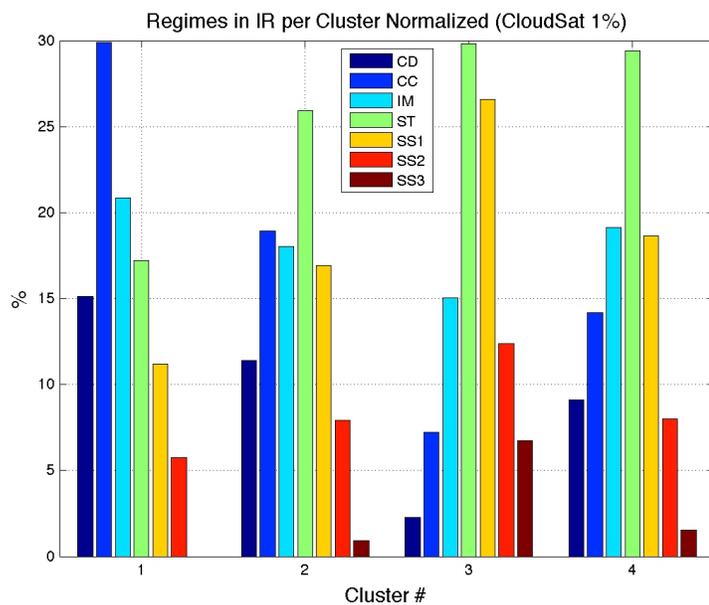
**Figure 16** - Geographic distribution of the collocated TRMM-CS and ISCCP cases



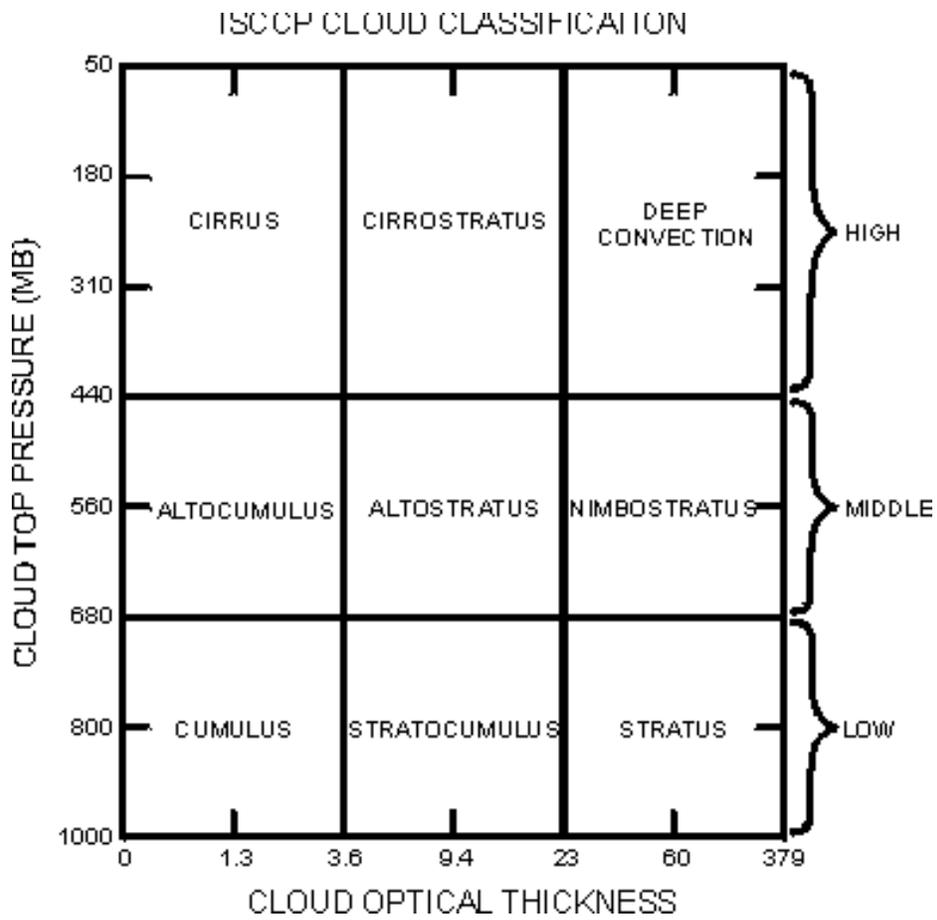
**Figure 17-** Breakdown of occurrence frequency of each Weather States per cluster.



**Figure 18** - Clusters per weather state.

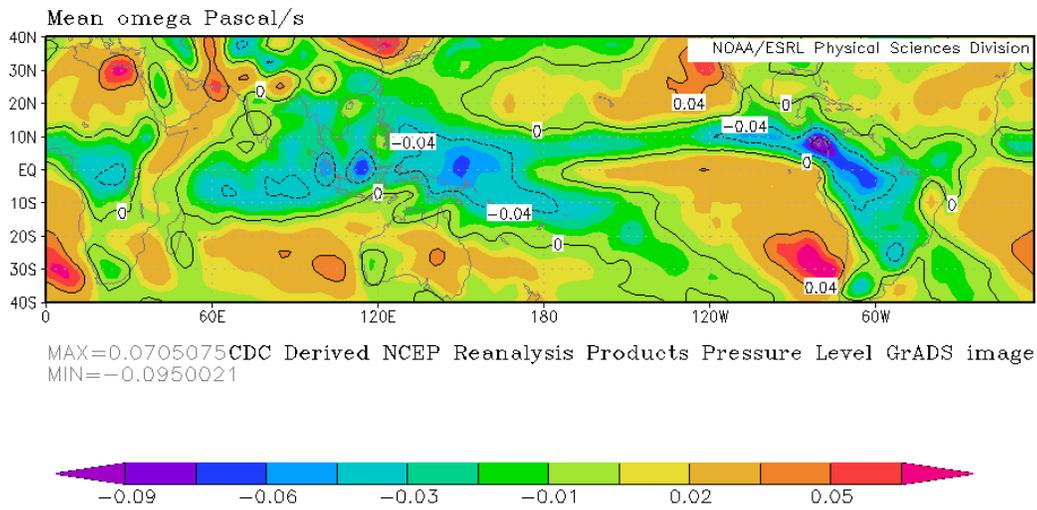


**Figure 19-** Regimes per cluster in the IR.



**Figure 20** ISCCP Cloud Classification chart.

lon: plotted from 0.00 to 357.50  
lat: plotted from -40 to 40  
lev: 500.00  
t: averaged over Aug 2006 to Dec 2010

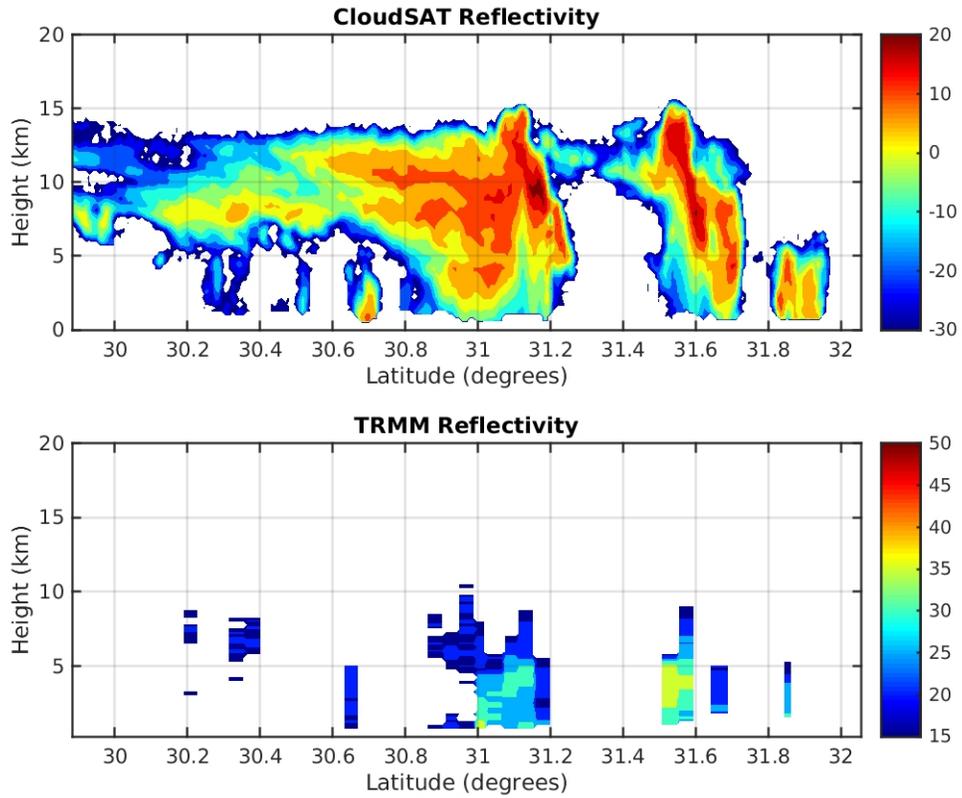


**Figure 21-** NCAR Reanalysis of vertical velocity at 500hPa from August 2006 to December 2010 globally in the tropics.

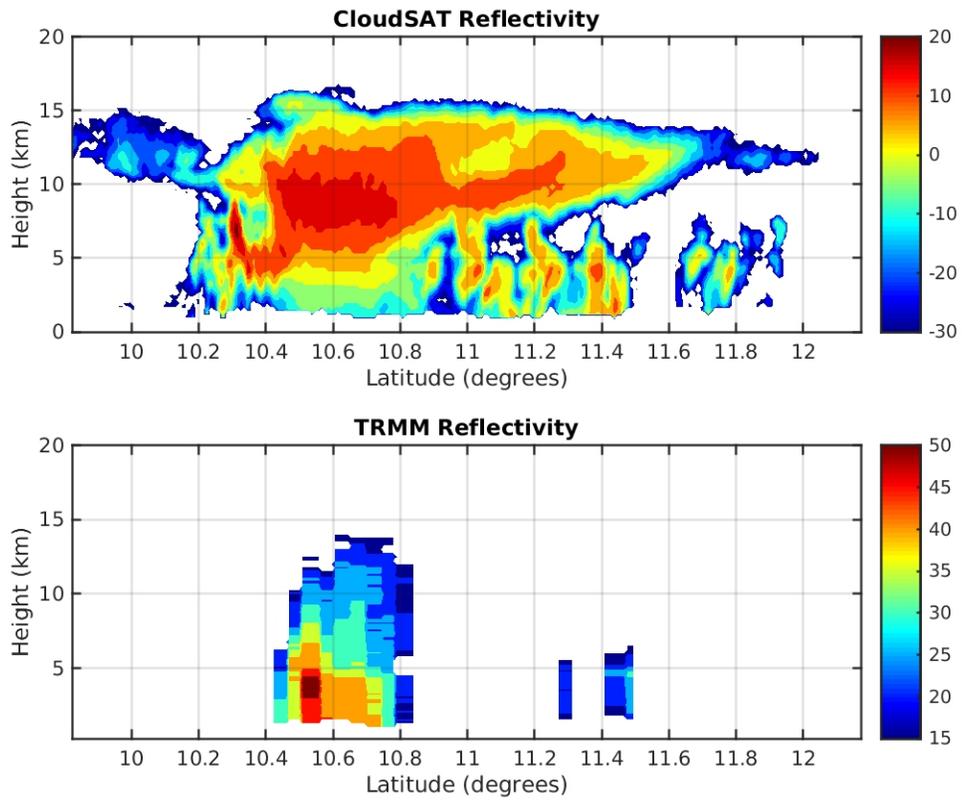
	CS-TRMM	ISCCP w/ CS-TRMM	ISCCP w/ CS-TRMM + w/o night + missing data + clear skies	IR w/ CS - TRMM
20min 30N- 30S	10,111	4,665	2,352	4501
20min + CloudSat 1%	5610	X	1219	2496
20min + TRMM 1%	2905	X	625	1274

**Table 1** - Distribution of HDF files used per condition.

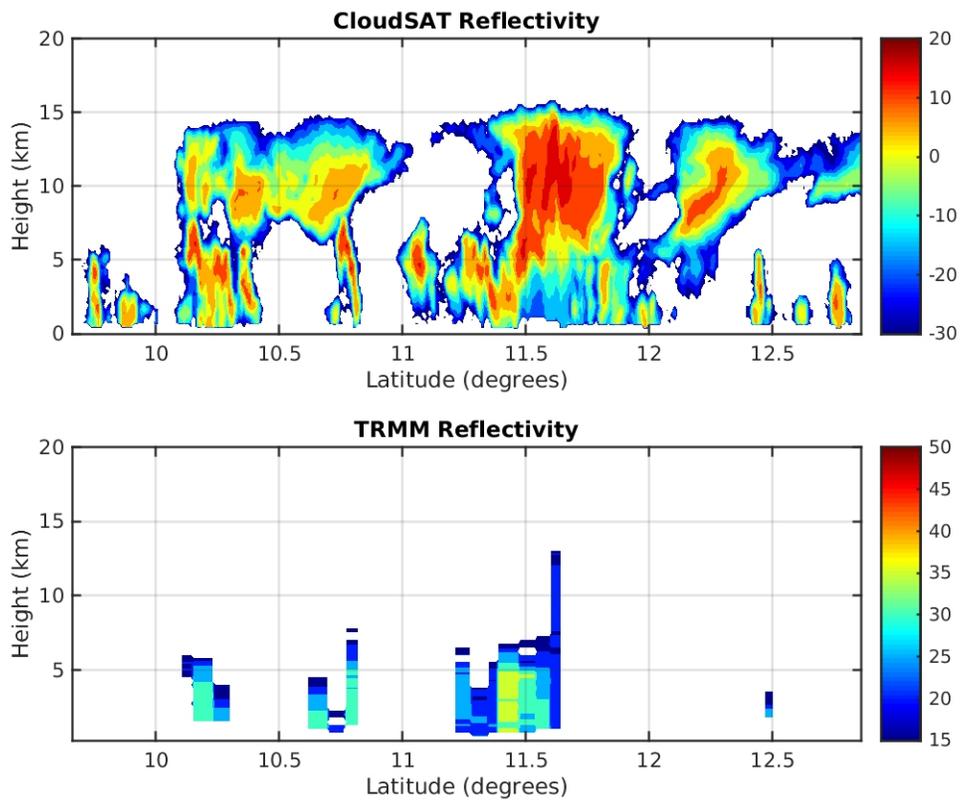
## Appendix A– Cluster 1 examples



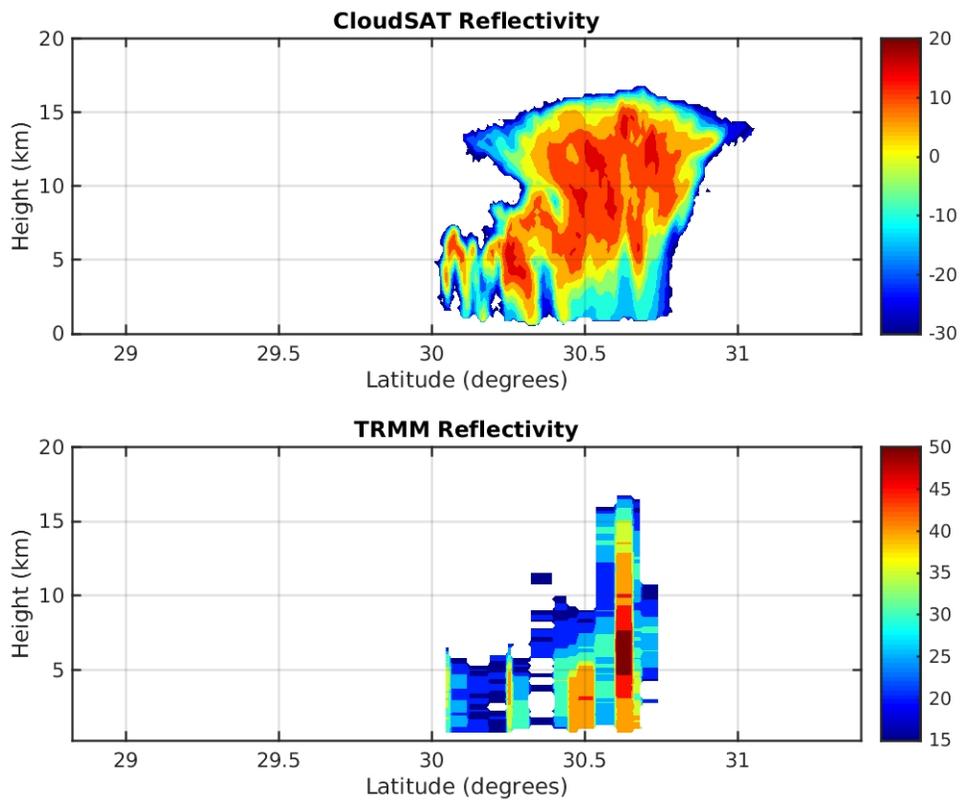
An example of a deep convective cloud type.



This is an example of cluster one; deep convective cloud type.

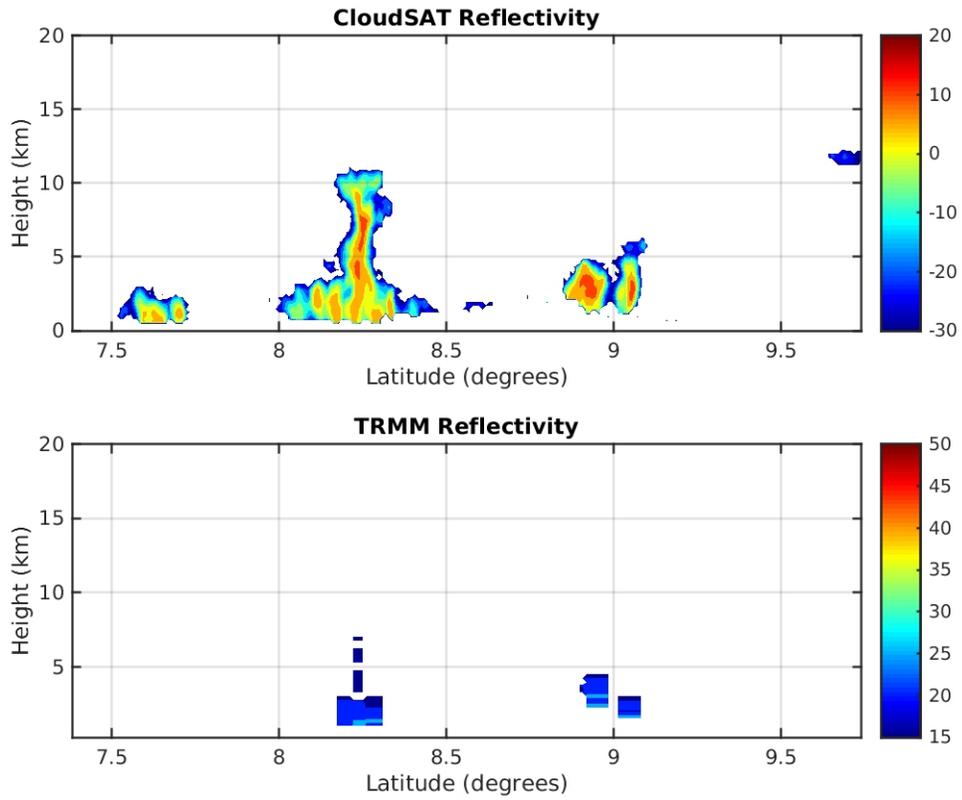


This is an example of cluster one; deep convective cloud type.

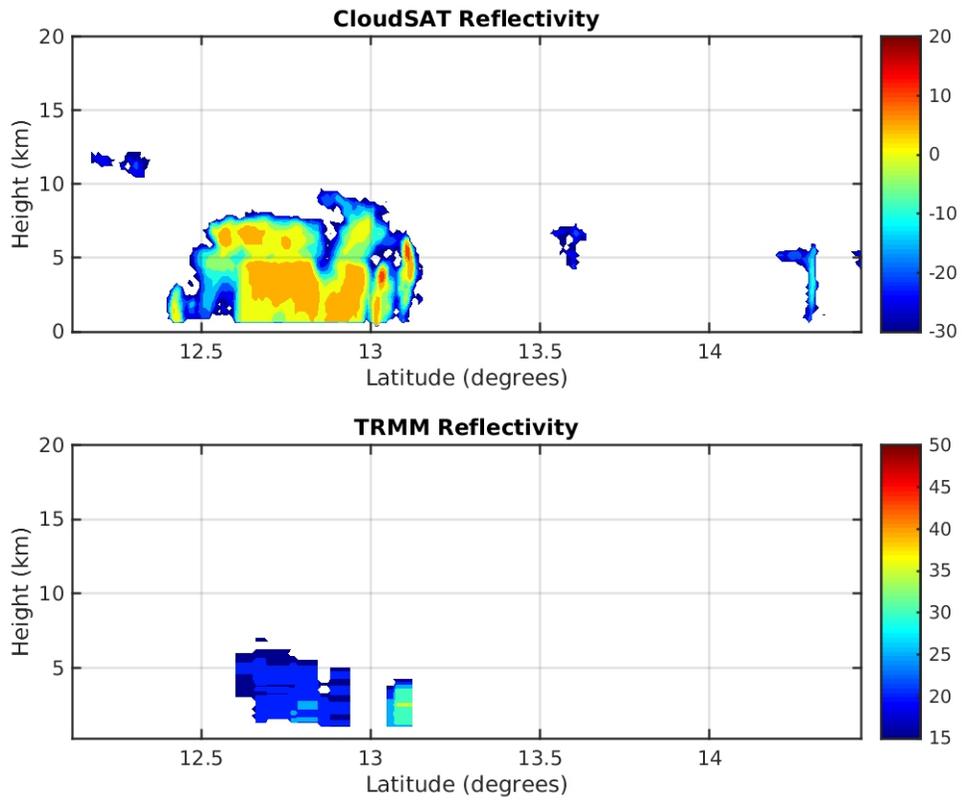


This is an example of cluster one; deep convective cloud type.

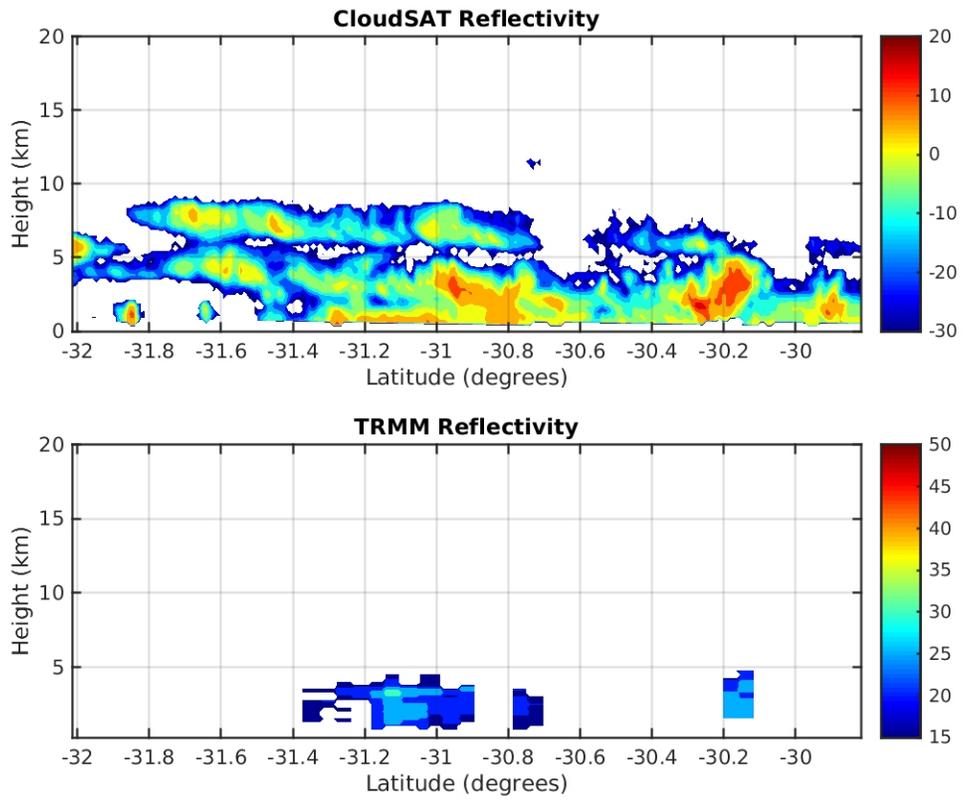
## Appendix B– Cluster 2 examples



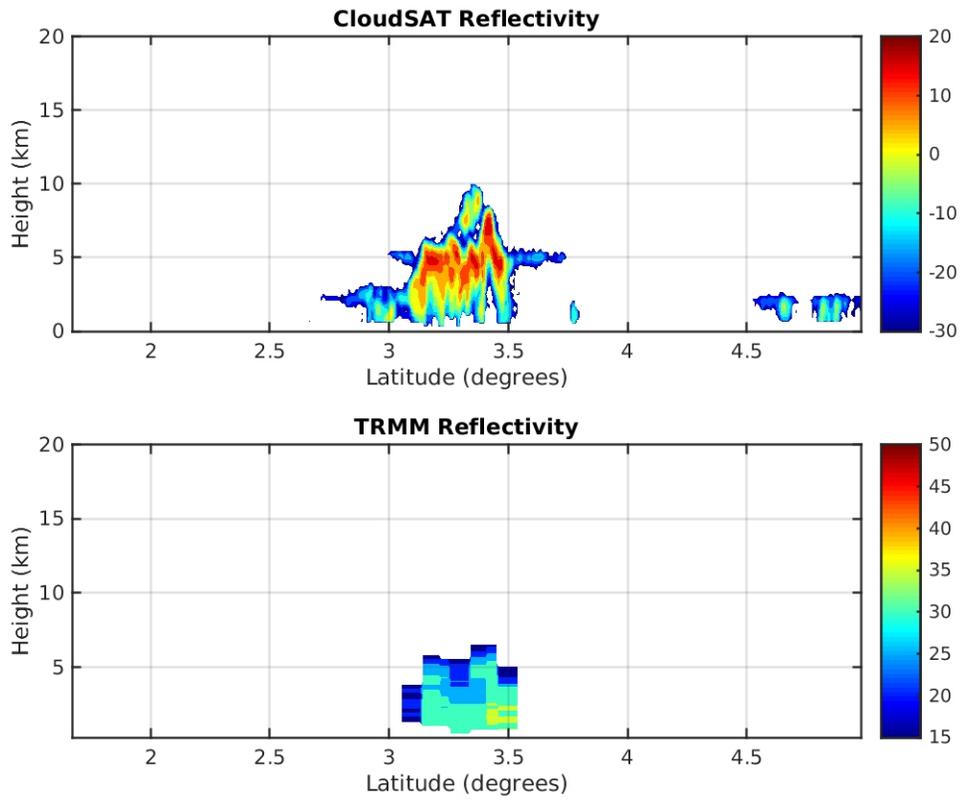
This is an example of cluster two; cumulus congestus cloud type.



This is an example of cluster two; cumulus congestus cloud type.

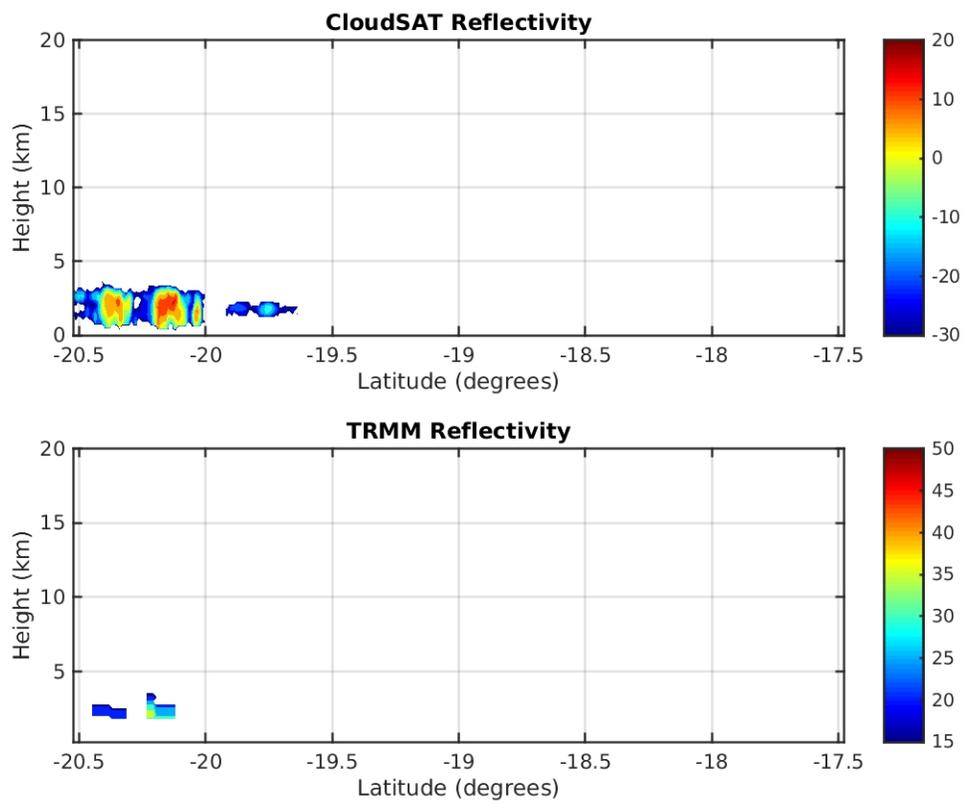


This is an example of cluster two; cumulus congestus cloud type.

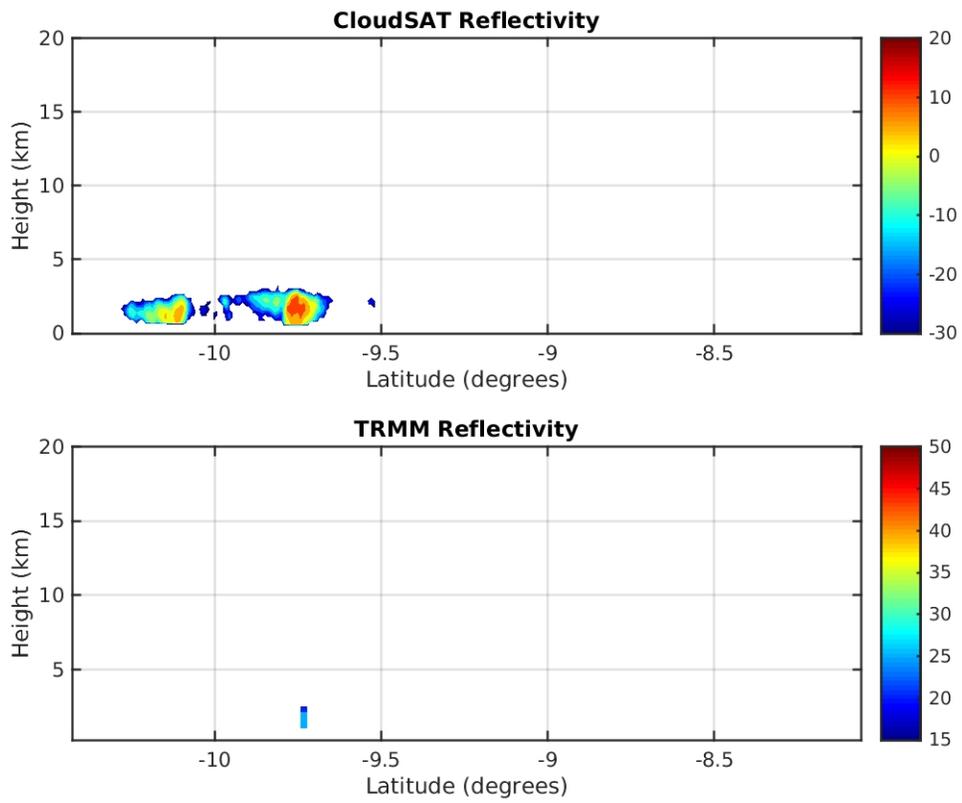


This is an example of cluster two; cumulus congestus cloud type.

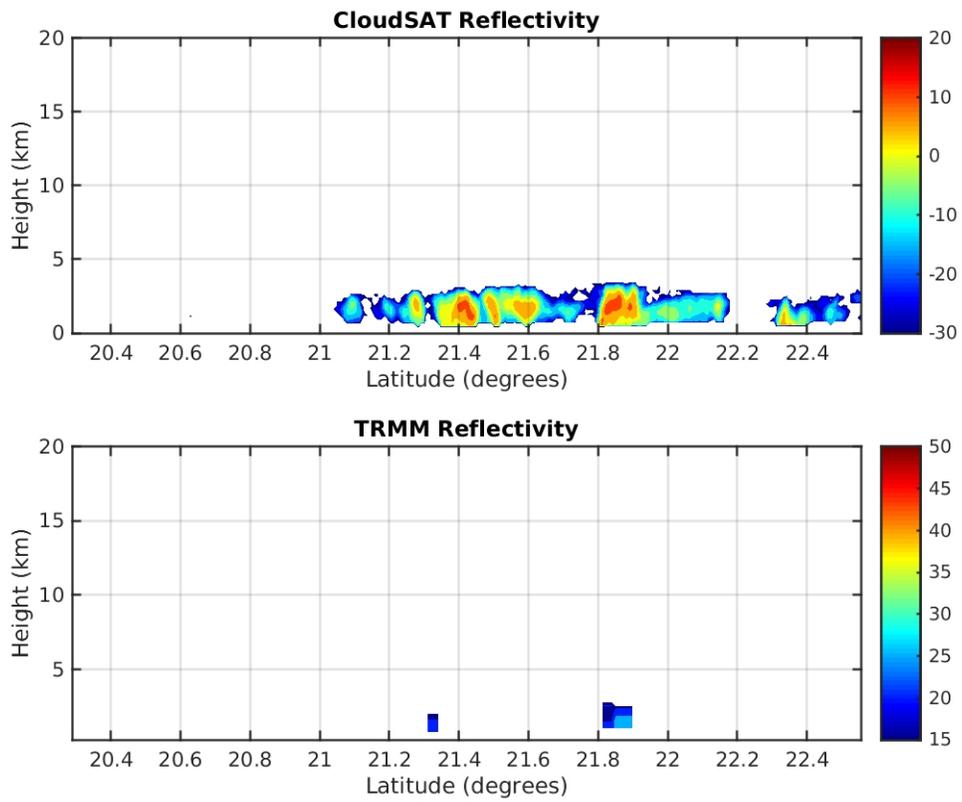
## Appendix C- Cluster 3 examples



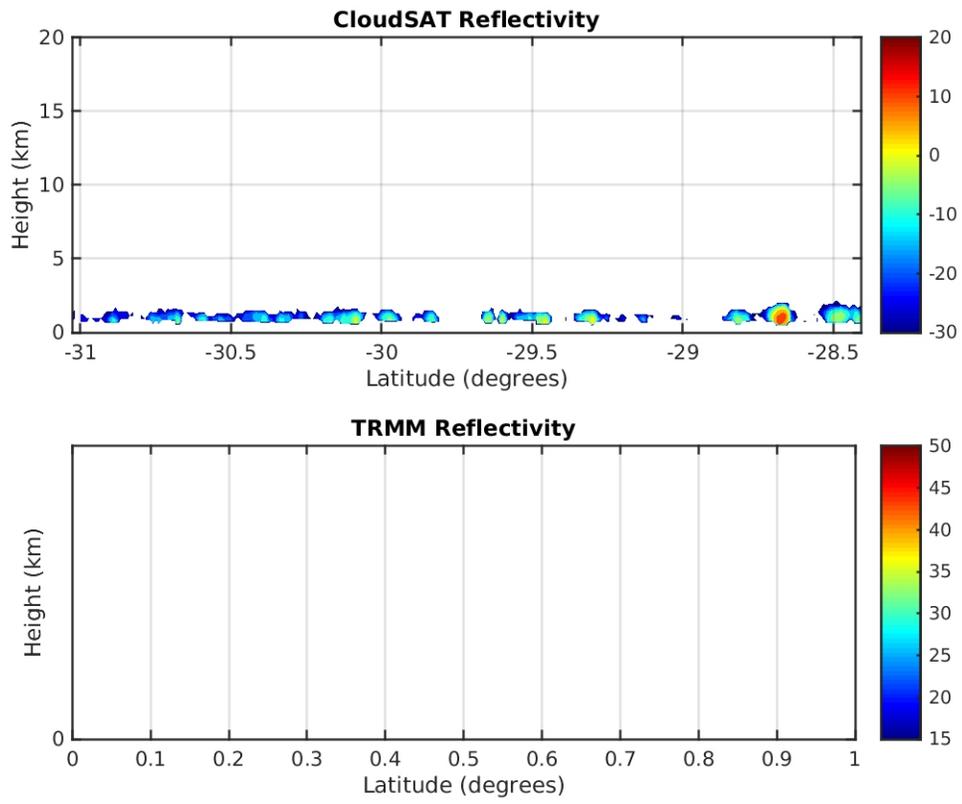
This is an example of cluster three; shallow cloud type.



This is an example of cluster three; shallow cloud type.

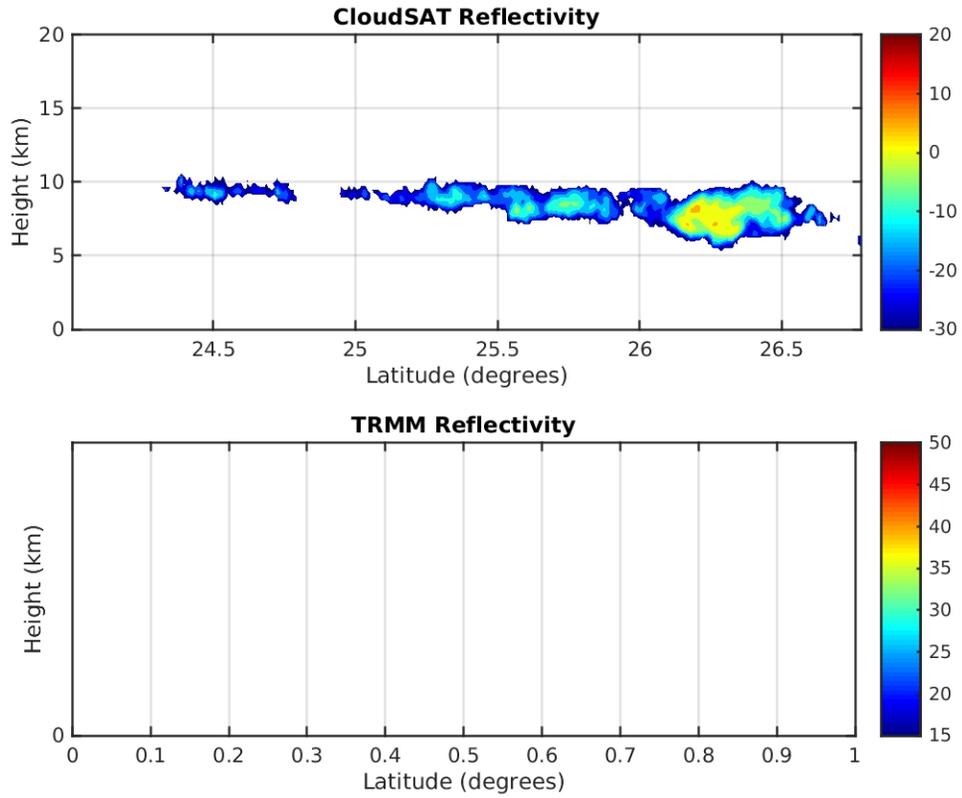


This is an example of cluster three; shallow cloud type.

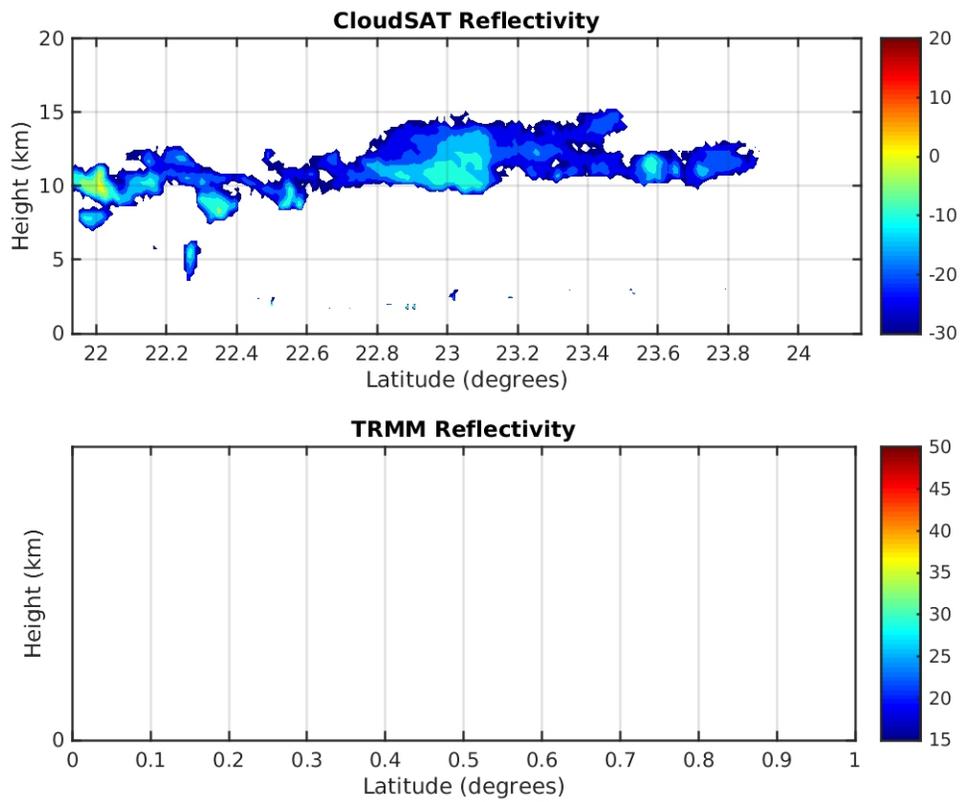


This is an example of cluster three; shallow cloud type.

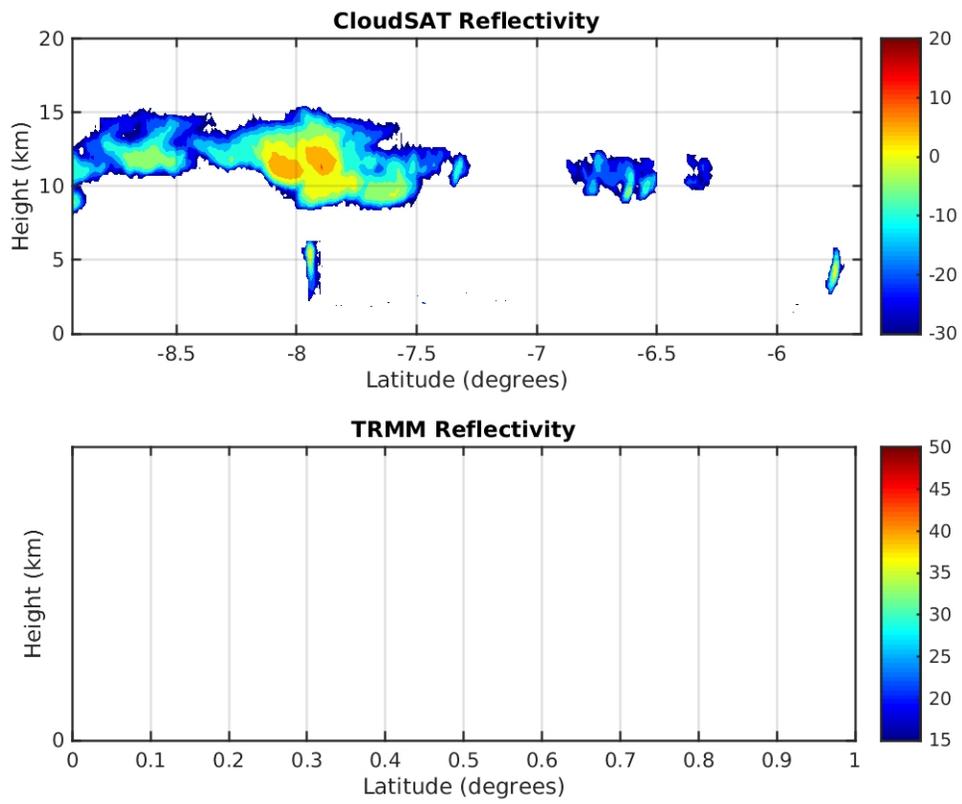
## Appendix D- Cluster 4 examples



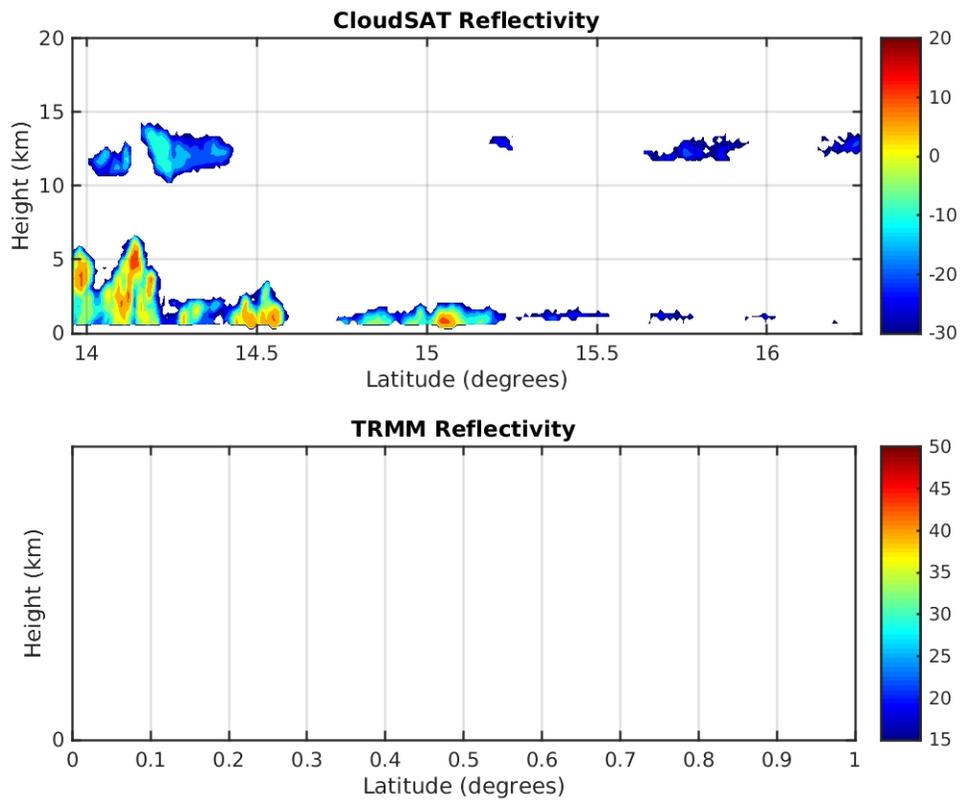
This is an example of cluster four; non-precipitating cloud type.



This is an example of cluster four; non-precipitating cloud type.



This is an example of cluster four; non-precipitating cloud type.



This is an example of cluster four; non-precipitating cloud type.

## Appendix E

Sample code for kmeans analysis.

```
%% This is to find the k-means
clc
clear
close all
%-----
----
%% INDICATE TIME DIFFERENCE... AND SATELLITE...
POV = 'cs'; % either cs, tm, or cstm
numClusters = 4; %

if strcmp(POV, 'cs')==1
    titleName = 'CloudSat';
elseif strcmp(POV, 'tm') ==1
    titleName = 'TRMM';
else
    disp('TITLE ERROR')
end
timeRestrained = 20; % number from 0 to 50
%-----
----

%%
%Loading the CFAD data

lcl_dir = '/Users/ricardoanderson/Desktop/Research/thesis2/';

load
([lcl_dir, 'cfad_', POV, '_', num2str(timeRestrained), '_min_apart30.mat'])

% this will be the normalizing portion of the data

normalized_all_data = [second_csOnly; second_tmOnly];

satellite = normalized_all_data(:,1:storing_count)'; % this is both
CS/TM data.. first half of the 'm'being CS
%subplot is an m x n matrix, define m & n here

%---- cosmetic stuff
m = 3;
n = 2;

% colorbar range
r = 0; %min
s = .25; %max

t = .5;
```

```

% -----

%% ----- K-MEANS HAPPENING HERE -----

opts = statset('Display','final');

[idx, c] = kmeans(satellite, numClusters, 'Distance', 'sqEuclidean',
'Replicates', 15, 'Options', opts, 'Display', 'iter');

size_idx = sum(double(~isnan(idx))); %number points used
[idx2, idx_pos] = sort(idx, 'descend');

%% redoing positions of clusters from least to most frequent which
%%corresponds to the most to the least convective.
tempFind= zeros(numClusters,1);
for kk = 1:numClusters
    tempFind(kk) =length(find(idx==kk));
end %kk
[~, z1] = sort(tempFind, 'ascend');

for i7 = 1:length(z1)
    posToChange = find(idx == z1(i7));
    idx(posToChange) = i7 * 10;
    c1(i7,:) = c(z1(i7),:);
end
idx = idx/10;
%manually fixing since the 3rd and 4th clusters need to swap

if strcmp(POV, 'cs') ==1 % manually organizing based off convection
c3 = c1(3,:);
c4 = c1(4,:);
c1(3,:) = c4;
c1(4,:) = c3;
i33 = find(idx==3);
i44 = find(idx==4);
idx(i44) =3;
idx(i33) = 4;
end

%plotting standards
ysx = (2:2:20);

%%

for i= 1:numClusters
    currentCluster = c1(i,:);

    sCluster = sum(currentCluster);

    %% This is to find the amount of times in Percentage that this
cluster
    %% number is assigned to a particular CFAD situation..
    positions_i = find(idx == i);
    amount_i = length(positions_i);
    percentage_i = (amount_i/size_idx) * 100;

```

```

cluster_cs = currentCluster(1, 1:360);
cluster_tm = currentCluster(1, 361:720);

cluster_cs = reshape(cluster_cs, 20, 18);
cluster_tm = reshape(cluster_tm, 20, 18);
%
%   cluster_cs = flipud(cluster_cs);
%   cluster_tm = flipud(cluster_tm);

figure(1)
subplot (m,n,i)
contourf(cluster_cs, 'LineStyle', 'none')
caxis([r s])
%   imagesc(cluster_cs, [r s])
%   imagesc(cluster_tm)
set(gca, 'XTick', [0:2:18])
set(gca, 'XTickLabel', [-35:10:55])

set(gca, 'YTick', [1:2:20])
set(gca, 'YTickLabel', ysx)
grid on

xlabel('Reflectivity (dbZ)', 'FontSize', 16)
ylabel('Height (km)', 'FontSize', 16)
title(['Cluster ' num2str(i) ' For CloudSAT '
(num2str(percentage_i)) '%' ], 'FontSize', 16)

hhh = colorbar('horiz');
set(hhh, 'Position', [0.12 .05 .8 .03]);
xlabel(hhh, 'Relative Frequency of Occurrence (%)', 'FontSize', 14)

figure(2)
subplot (m,n,i)
contourf(cluster_tm, 'LineStyle', 'none')
caxis([r t])
%   imagesc(cluster_tm, [r t])
%   imagesc(cluster_tm)
set(gca, 'XTick', [0:2:18])
set(gca, 'XTickLabel', [-35:10:55])

set(gca, 'YTick', [1:2:20])
set(gca, 'YTickLabel', ysx)
grid on
xlabel('Reflectivity (dbZ)', 'FontSize', 16)
ylabel('Height (km)', 'FontSize', 16)

title(['Cluster ' num2str(i) ' For TRMM ' (num2str(percentage_i))
'% ' ], 'FontSize', 16)
hhh = colorbar('horiz');
set(hhh, 'Position', [0.12 .05 .8 .03]);
xlabel(hhh, 'Relative Frequency of Occurrence (%)', 'FontSize', 14)

end

```

```
figure(1)
file_name_to_save1 = ['kmeans_', num2str(timeRestrained), '_'
POV, '_cldsats_k_is_', num2str(numClusters), '30NSlevel'];
orient tall
print ('-djpeg99', sprintf('%s.jpeg', file_name_to_save1))

figure(2)
file_name_to_save2 = ['kmeans_', num2str(timeRestrained), '_'
POV, '_trmm_k_is_', num2str(numClusters), '30NSlevel'];
orient tall
print ('-djpeg99', sprintf('%s.jpeg', file_name_to_save2))
close all

filename =
['k_eq_', num2str(numClusters), '_', POV, '_only_1_percent30NS_level'];
save(filename)
```

## Bibliography

- Anderberg, M. R. (1973), Cluster Analysis for Applications, 359 pp., Elsevier, New York. Bony
- Boccippio, Dennis J., Walter a. Petersen, and Daniel J. Cecil. 2005. "The Tropical Convective Spectrum. Part I: Archetypal Vertical Structures." *Journal of Climate* 18(14): 2744–69.
- Elsaesser, Gregory S. et al. 2010. "Observed Self-Similarity of Precipitation Regimes over the Tropical Oceans." *Journal of Climate* 23: 2686–98.
- Jakob, Christian. 2003. "Objective Identification of Cloud Regimes in the Tropical Western Pacific." *Geophysical Research Letters* 30(21): 1–4.
- Johnson, R. H., T. M. Rickenbach, S. A. Rutledge, P. E. Ciesielski, and W. H. Schubert (1999), Trimodal characteristics of tropical convection, *J. Clim.*, 12, 2397 – 2418.
- Kummerow, Christian et al. 1998. "The Tropical Rainfall Measuring Mission (TRMM) Sensor Package." *Journal of Atmospheric and Oceanic Technology* 15(3): 809–17.
- Luo, Zhengzhao, G. Y. Liu, Graeme L. Stephens, and Richard H. Johnson. 2009. "Terminal versus Transient Cumulus Congestus: A CloudSat Perspective." *Geophysical Research Letters* 36(5): 4–7.
- Mace, Gerald G., Roger Marchand, Qiquing Zhang, and Graeme Stephens. 2007. "Global Hydrometeor Occurrence as Observed by CloudSat: Initial Observations from Summer 2006." *Geophysical Research Letters* 34.
- Rossow, William B., and Robert a. Schiffer. 1999. "Advances in Understanding Clouds from ISCCP." *Bulletin of the American Meteorological Society* 80: 2261–87.
- Rossow, William B., George Tselioudis, Allyson Polak, and Christian Jakob. 2005. "Tropical Climate Described as a Distribution of Weather States Indicated by Distinct Mesoscale Cloud Property Mixtures." *Geophysical Research Letters* 32(October): 1–4.
- Simpson, J, Kummerow, C, Meneghini, R, Hou, R, Adler, R, Huffman, G, Barkstrom, B, Wielicki, B, Goodman, S, Kozu, T, Wang, S, Krishnamurti, T, Ferrier, B. 1999. "The Tropical Rainfall Measuring Mission (TRMM) Progress Report."

- Stephens, Graeme L. et al. 2002. "The CloudSat Mission and the A-Train: A New Dimension of Space-Based Observations of Clouds and Precipitation." *Bulletin of the American Meteorological Society* 83(12): 1771–90+1742.
- Tan, Jackson, and Christian Jakob. 2013. "A Three-Hourly Data Set of the State of Tropical Convection Based on Cloud Regimes." *Geophysical Research Letters* 40(April): 1415–19.
- Wielicki, B. A., R. D. Cess, M. D. King, D. A. Randall, and E. F. Harrison, 1995: Mission to planet Earth: Role of clouds and radiation in climate. *Bull. Amer. Meteor. Soc.*, 76, 2125–2153.
- Yuter, Sandra E., and Robert a. Houze. 1995. "Three-Dimensional Kinematic and Microphysical Evolution of Florida Cumulonimbus. Part II: Frequency Distributions of Vertical Velocity, Reflectivity, and Differential Reflectivity." *Monthly Weather Review* 123: 1941–63.
- Zhang, Yujin, Steve Klein, Gerald G. Mace, and Jim Boyle. 2007. "Cluster Analysis of Tropical Clouds Using CloudSat Data." *Geophysical Research Letters* 34: 1–6.