JGR Atmospheres

RESEARCH ARTICLE
10.1029/2024JD043098

Tropical Cloud Classification With a Clustering Analysis Applied to CloudSat and CALIPSO Observations

Dea T. Octarina[1] **and Hirohiko Masunaga**[2]

[1]Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan, [2]Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Japan

Key Points:

- Cluster analysis provides a robust cloud classification based on the CALIPSO and Cloudsat measurements
- Increasing the number of clusters beyond five only causes redundant high-cloud clusters
- Meteorological interpretation is crucial when assessing the clustering solution without fully relying on statistical evaluations

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

D. T. Octarina,
dea.tania.octarina.f3@s.mail.nagoya-u.ac.jp

Citation:

Octarina, D. T., & Masunaga, H. (2025). Tropical cloud classification with a clustering analysis applied to CloudSat and CALIPSO observations. *Journal of Geophysical Research: Atmospheres*, *130*, e2024JD043098. https://doi.org/10.1029/2024JD043098

Received 4 DEC 2024
Accepted 9 JUL 2025

Author Contributions:

Conceptualization: Dea T. Octarina, Hirohiko Masunaga
Formal analysis: Dea T. Octarina
Investigation: Dea T. Octarina
Methodology: Dea T. Octarina, Hirohiko Masunaga
Resources: Hirohiko Masunaga
Supervision: Hirohiko Masunaga
Visualization: Dea T. Octarina
Writing – original draft: Dea T. Octarina
Writing – review & editing: Hirohiko Masunaga

Abstract Tropical cloud fraction profiles from combined CALIPSO and CloudSat observations are classified based on a K-means cluster analysis. The Cloud Profiling Radar and Cloud-Aerosol Lidar with Orthogonal Polarization carried by these satellites together retrieve detailed vertical structure of clouds, allowing for an objective decomposition of tropical clouds into distinct regimes. The identified cloud regimes are overall robust except that an increasing number of redundant high-cloud clusters are produced as the specified number of clusters is increased. This redundancy arises because geometrically thin clouds with slightly different altitudes can be mathematically distant from one another in the clustering procedure despite their morphological similarity, suggesting that an excessive number of clusters only generates unrealistic subclasses of high clouds. Objective evaluation metrics indicate that the most appropriate number of clusters is four, where the clusters may be labeled as sparse multilayer clouds, low cloud, high cloud, and deep convective regimes. A five-cluster classification is also deemed as reasonable, where a fifth regime of congestus clouds joins the other four regimes, practically equivalent to the four-cluster solution.

Plain Language Summary Clouds have essential radiative impacts on Earth's climate system. Tropical anvil clouds are especially a vital element of climate feedback in response to a temperature perturbation to the climate system. To this end, recognizing the roles of various cloud types and their impact on the future environment is essential. The vertical structure of tropical clouds observed by satellites is classified into groups based on their similarities using the statistical technique called cluster analysis. Using this technique, cloud profiles can be divided into distinct categories such as low cloud, high cloud, deep convection, and sparse thin clouds. Clustering analysis only produces unrealistic multiple high-cloud groups when it is attempted to generate more and more cloud groups. While meteorological and mathematical assessments suggest that the four groups are the most appropriate grouping, adding a fifth group offers another reasonable alternative including a 5 or 6-km thick cloud called congestus as well. This study suggests that a full reliance on statistical evaluation can be misleading, and careful interpretation based on atmospheric science can help the assessment of cluster analysis performance.

1. Introduction

Clouds play a crucial role in balancing the global energy budget. Their geographical distribution, especially in the tropics, closely matches the pattern of energy transport in global atmospheric circulations. An energy surplus in the tropics is balanced by upward energy transport into the upper troposphere through deep convective clouds concentrated over the Pacific warm pool, Indian Oceans, and tropical continents, forming the intertropical convergence zone (ITCZ). These towering clouds in ITCZ are typically accompanied by broad high clouds, ascribed mainly to the air mass detrainment from the convective clouds in the upper troposphere. In areas away from the convective region, lower clouds are more prevalent, trapped below an inversion layer where downward motions find spaces between these shallow cumulus clouds, as depicted in Simpson (1992). This conventional picture was later revised by Johnson et al. (1999), who highlighted the abundant population of cumulus congestus clouds also prevailing over the tropics. As such, a variety of clouds with different thicknesses and top heights are spread with distinct regionalities in the tropics, impacting the global climate through their different radiative properties and characteristics. That is why distinguishing clouds with attention to their respective roles and impacts on the environment is fundamental in climate studies.

Clouds have essential radiative impacts on Earth's climate system. Tropical anvil clouds are especially a vital element of climate feedback in response to a temperature perturbation to the climate system. Anvil clouds can

give a warming effect by trapping outgoing longwave radiation and a cooling effect by reducing incoming solar radiation. Past studies have drawn mixed conclusions about how anvil clouds interact with surface temperature. Some said that anvil cloud cover could increase with increasing surface temperature (Ohno & Satoh, 2018; Ramanathan & Collins, 1991). Others argued that warmer surface temperatures would reduce anvil cloud fraction instead (Bony et al., 2016; Lindzen et al., 2001). Another related hypothesis is that the top temperature of tropical anvil clouds is insensitive to surface temperature changes (Hartmann & Larson, 2002; Zelinka & Hartmann, 2010). As a tool to test such hypotheses against observations, we develop an analysis method to classify cloud types based on cloud vertical profiles so that the cloud physical properties and their radiative effects are studied separately for anvil and other clouds. The future climate cannot be captured using observations, but the underlying process explaining these hypotheses can still be investigated by observing how the current climate system works at a process level. This method is aimed to serve as a valuable tool for unraveling the intricate processes that drive cloud changes in an increasingly warmer climate. To this end, this study applies a clustering analysis to spaceborne radar and lidar measurements of in-cloud vertical structure.

Attempts have been made previously to devise statistical methods of categorizing cloud types in an objective manner. In these studies, satellite remote sensing offers a primary tool for analyzing clouds observed from passive and/or active sensors. Passive sensors can provide data over an extended period of time capable of capturing cloud variability over decades. For example, Rossow et al. (2005) employ 21.5 years of data from the International Satellite Cloud Climatology Project (ISCCP) D1 data to break down tropical clouds into six weather states by applying the K-means clustering to the joint frequency distribution of cloud top pressure and optical thickness. Tselioudis et al. (2013) expanded this result into the global domain, while Chéruy and Aires (2009) applied a similar method on a regional scale. Oreopoulos et al. (2014) and Cho et al. (2021) derived cloud regimes from the Moderate Resolution Imaging Spectroradiometer (MODIS) data using the same clustering method. Gordon et al. (2005) also used the long-term ISCCP D1 data to classify midlatitude cloud regimes, using the cloud top, cloud reflectivity, and cloud top pressure information.

Passive instruments, however, are not an optimal tool for profiling in-cloud structure and properly detecting overlapped clouds. Active sensors such as radars and lidars, on the other hand, can retrieve more detailed information, such as clouds' vertical structures, forms, and microphysical properties. Although they are limited in temporal and spatial coverage, previous studies have proved the utility of active remote sensing to identify cloud regimes through similar clustering methods. Zhang et al. (2007) applied K-means clustering to the joint frequency distributions of radar reflectivity and pressure levels. The reflectivity in the vertical dimension was binned into 42 independent elements of the joint histogram. Zhang et al. (2010) improved this approach by combining data from CloudSat and CALIPSO, and clustering was performed on fewer elements than the previous work to avoid uncertainties caused by distances between the elements. Luo et al. (2017) further elaborated a similar method by adding the Tropical Rainfall Measuring Mission data to associate tropical cloud regimes with precipitation regimes. Most recently, Lee et al. (2025) developed a clustering method for cloud vertical morphology that introduces the possibility to connect the spatiotemporally sparse active sensor observation with the gridded passive sensor data.

Existing cloud classifications through a clustering scheme rely only on statistically pre-processed cloud information such as a joint frequency distribution or a normalized index. In this study, a more straightforward approach is employed. Using a combination of CloudSat and CALIPSO data, a cluster analysis is performed on the vertical profiles of cloud fraction on a footprint-by-footprint basis. To classify tropical clouds into regimes, the high vertical resolution native of CloudSat measurements is retained for fully exploiting the original information content and to investigate the potential complexities that may arise from maintaining high cloud profile resolution during the clustering process. The primary emphasis of this paper should be on the intricacies and methodologies involved in the clustering process itself, rather than merely discussing the resulting cloud classifications. As will become clear later, the native vertical resolution of Cloudsat/CALIPSO measurements gives rise to technical challenges that would not be encountered for a coarser resolution. This work carefully explores a strategy to overcome such difficulties without compromising to lower the vertical resolution. A K-means clustering method is used as in previous studies. Here, the clustering analysis result is evaluated using existing statistical tools, without blindly relying on them, to find the optimal number of cloud regimes. Then, the characteristics of each cloud regime are discussed, including their global distributions and cloud radiative properties.

2. Data and Methodology

2.1. CCCM Data

The vertical profiles of cloud fraction and irradiance from merged CALIPSO and CloudSat data provided in the CALIPSO-CloudSat-CERES-MODIS (CCCM) Release D1 product (NASA/LARC/SD/ASDC, 2011) are used in this study. The product is an integrated CALIPSO Cloud-Aerosol Lidar with Orthogonal Polarization, CloudSat Cloud Profiling Radar, Clouds and the Earth's Radiant Energy System (CERES), and the MODIS data (Kato et al., 2010). The cloud vertical structure is derived from CloudSat and CALIPSO measurements as a volumetric cloud fraction. A given layer is identified as overcast when CALIPSO and/or CloudSat detected a cloud. The CloudSat and CALIPSO cloud masks are adjusted to the CloudSat vertical resolution with the horizontal resolution first merged into a 1 × 1-km grid box. The horizontal resolution is then adjusted to match the size of a CERES footprint (20 km), allowing for up to 50 merged profiles within this footprint. The cloud fraction is calculated as a volumetric fraction, that is, the fraction of cloudy CALIPSO bins to the total number of bins in every CERES footprint. These cloud profiles are also ingested in the radiative transfer model used to produce the radiative flux profile in the CCCM product. The volumetric cloud fraction is essential for the radiative transfer model calculation as it provides accurate information about cloud boundaries. A slight change in this cloud structure information might alter radiative profiles within the atmosphere and clouds. Here, the irradiance profiles are calculated using a combination of CALIPSO, CloudSat, and MODIS cloud and aerosol properties in the radiative transfer model, as explained in Kato et al. (2011). The use of MODIS cloud and aerosol properties in CCCM modestly improves the retrieval accuracy as an additional observational constraint to the CALIPSO and CloudSat-derived cloud and aerosol profiles (Ham et al., 2022), whereas the vertical structure of cloud properties used for radiation estimates relies exclusively on the active sensor observations. The original vertical resolution of cloud parameters in the CALIPSO product is 30 m below 8.2 km and 60 m above 8.2 km. While in CloudSat, it is 480 m. After merging, the cloud fraction and irradiance profile vertical resolution decreases with altitude, ranging from 120 m near the surface to 5,000 m near the top of the atmosphere (TOA).

CCCM cloud occurrences exhibit overall similar features to those in CloudSat 2B-GEOPROF-LIDAR, particularly over the ocean tropics, where the occurrence differences are minimal, except at altitudes below 1 km (Ham et al., 2017). CCCM is, however, unique in that the variables are collocated to CERES footprints for ease of comparison with CERES measurements. This can be an advantage for applications targeted on atmospheric radiation budget. We used CCCM data from 2007 until 2010 over the tropics (20°S–20°N). A narrow tropical zone was chosen to avoid the influence of extra-tropical disturbances in our analysis.

2.2. Clustering of Tropical Cloud Fraction Profiles

In this study, a cluster analysis (Anderberg, 1973) is applied to classify cloud vertical structures in the tropics into separate regimes. A number of clusters are identified so every cluster is mathematically separable as much as possible from one another in parameter space (i.e., the vertical profile of cloud fraction in the present case). Each cloud fraction profile is then assigned to the most appropriate cluster.

The improved version of K-means clustering, the K-means++ algorithm (Arthur & Vassilvitskii, 2007), is implemented to analyze 4-year cloud fraction profiles from CCCM data. The cloud fraction profiles over land and completely clear-sky profiles (i.e., columns with no cloud detected at any level) are excluded from the clustering process. The study area is limited to over the oceans. While the inclusion of clouds over land has been confirmed not to essentially alter the identified clusters, over-land profiles are excluded to avoid complications in interpreting the results due to variable surface altitudes. The K-means clustering method defines clusters based on the distance between cluster centroids and the proximity of data to the centroid. First, k clusters are determined as the initial condition. Then, the K-means++ algorithm applies its stochastic approach by supervising the initial selection of k centroids with probability instead of randomly selecting them, and then assigns each cloud profile to an adequate cluster based on the distance to the centroids. Cluster centroids are then updated repeatedly until an optimal solution is obtained. The algorithm, in short, consists of two-fold iterations: assigning each observation to its closest center and calculating new cluster centers. The iterations would continue until the cluster assignment does not change anymore or the number of iterations achieves a specified upper limit.

2.3. Statistical Evaluation Tools

The performance of our clustering analysis is assessed by the sum of point-to-centroid distances within each cluster. The K-means method calculates the squared Euclidean distance as a distance metric between one observation and its closest centroid. The sum of this distance within one cluster represents the proximity of the cluster members to its centroid. When the sum of distances is small, the member's distribution within a cluster is denser, indicating that the grouping is of better quality. As a mathematical metric, this measure accounts for non-dimensional distance in cloud-fraction space (0–1) and should not be confused with physical distance scale. The frequency of occurrence, or the cluster population as a fractional percentage of members in each cluster, serves as another key factor measuring the robustness of clustering.

Different metrics have been proposed to find out the optimal number of clusters for a given sample of data. In this study, this assessment is made using two evaluation indices, the Calinski-Harabasz index (Caliński & Harabasz, 1974) and the Davies–Bouldin index (Davies & Bouldin, 1979). The Calinski-Harabasz index is called the variance ratio criterion as it compares the between-cluster variance (SSB) and within-cluster variance (SSW) as shown in Equation 1.

$$\mathrm{VRC}_k = \frac{\mathrm{SS}_B}{\mathrm{SS}_W} \times \frac{(N-k)}{(k-1)}, \tag{1}$$

where k is the number of clusters and N is the number of observations. A proper criterion for cluster analysis is considered here to have a large between-cluster variance and a small within-cluster variance, or resultingly VRCk should be maximized. In contrast with this index, which measures the discrepancy between groups, the Davies–Bouldin index measures the similarity. The Davies–Bouldin index computes the ratio of distances within a cluster and between clusters (Equations 2 and 3).

$$R_{ij} = \frac{S_i + S_j}{M_{ij}}, \tag{2}$$

$$\overline{R} = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\{R_{ij}\}. \tag{3}$$

Here R_{ij} is a normalized measure of separation, while S_i and S_j are the dispersion of clusters i and j, respectively, and M_{ij} is the distance between the centroids of the ith and jth clusters. As such, $\overline{R}$ represents an averaged similarity measure. To obtain an appropriate clustering solution, the Davies–Bouldin index should be minimized.

3. Result and Discussion

3.1. Clustering Evaluation

The K-means clustering is applied to the 4-year CCCM cloud fraction profiles over the tropics. Four clusters are set in the first attempt of the clustering. The result shows distinct cloud regimes as indicated by Figure 1. Figures 1a and 1b are an example snapshot of a sequence of cloud profiles with their decomposition into four different clusters by the K-means. An inspection of the composite profiles of the four clusters (Figures 1c–1f) allows us to interpret each clustering solution in light of conventional meteorological terminology: low clouds (Cluster 1), deep convective clouds (Cluster 2), sparse multilayer clouds (Cluster 3), and high clouds (Cluster 4).

One of the well-known uncertainties intrinsic to the K-means clustering arises from its stochastic nature in selecting the initial cluster cores. The K-means clustering could produce entirely different results depending on the initial guess. This issue is addressed by conducting multiple repetitions of K-means clustering. A minimum of seven repetitions is carried out without adjusting the initial guesses. Each repetition resulted in four to eight cluster solutions, which are illustrated in Figures S2–S8 in Supporting Information S1. The 4, 6, 7, and 8-cluster solutions each have two representations, while the 5-cluster solution has three representations. Overall, the K-means repetitions converge into two or three consistent representatives, regardless of the stochastically chosen initial guess. This consistency ensures the robustness of the current classification. Multiple representations from different initial guesses are compared in terms of the sum of point-to-centroid distances to determine which

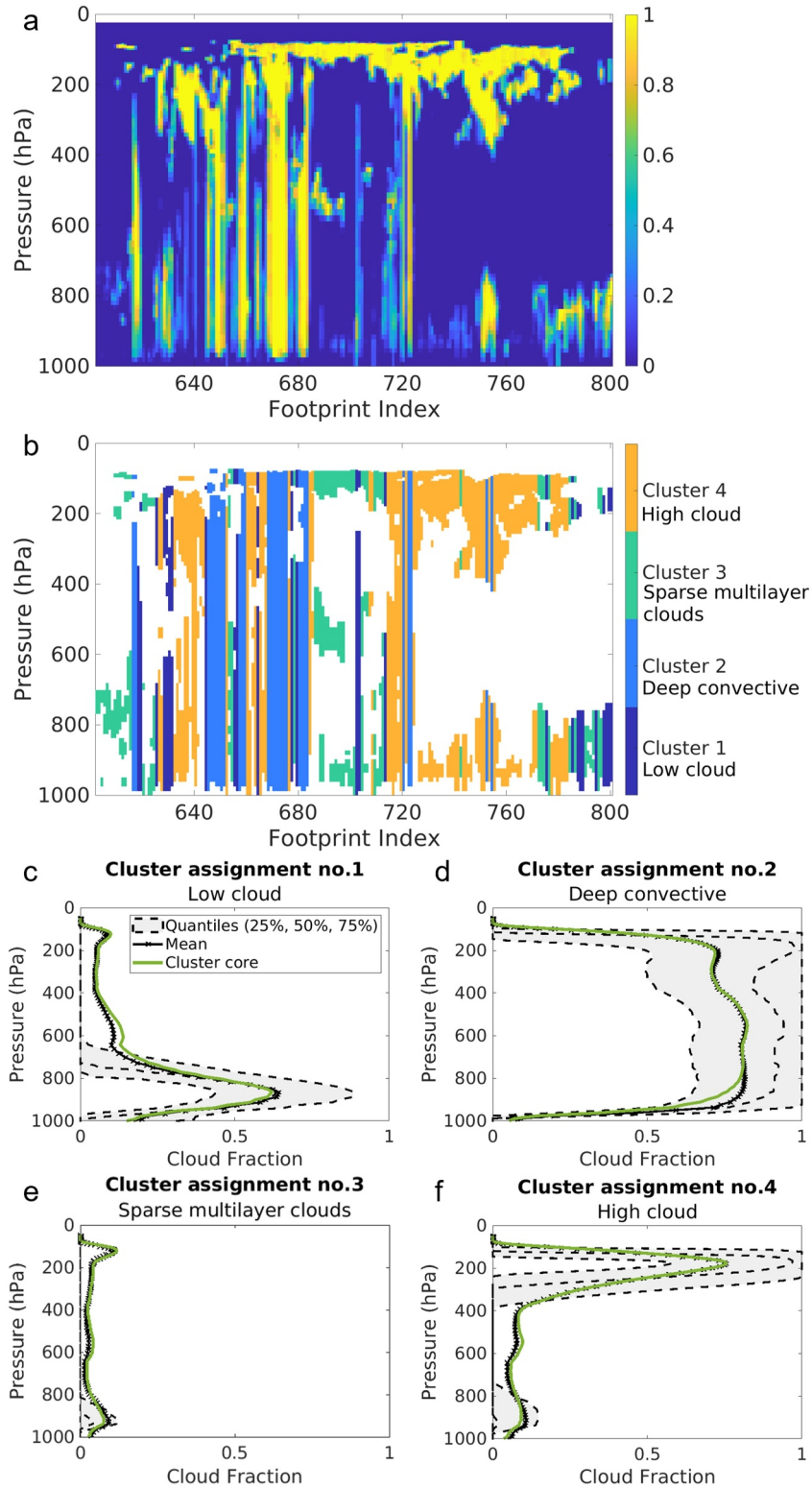

Figure 1. Four clusters solution of K-means derived from 4-year cloud fraction profiles in CCCM data (year 2007–2010). (a) Vertical cross-section of a sequence of cloud fraction profiles from 1 January 2007 footprint (footprint index 601–800). (b) Decomposition of (a) into four clusters. (c)–(f) Composite cloud fraction profiles for each cluster (solid) and their statistical spread at 25%, 50%, and 75% quartiles (dashed and gray area).

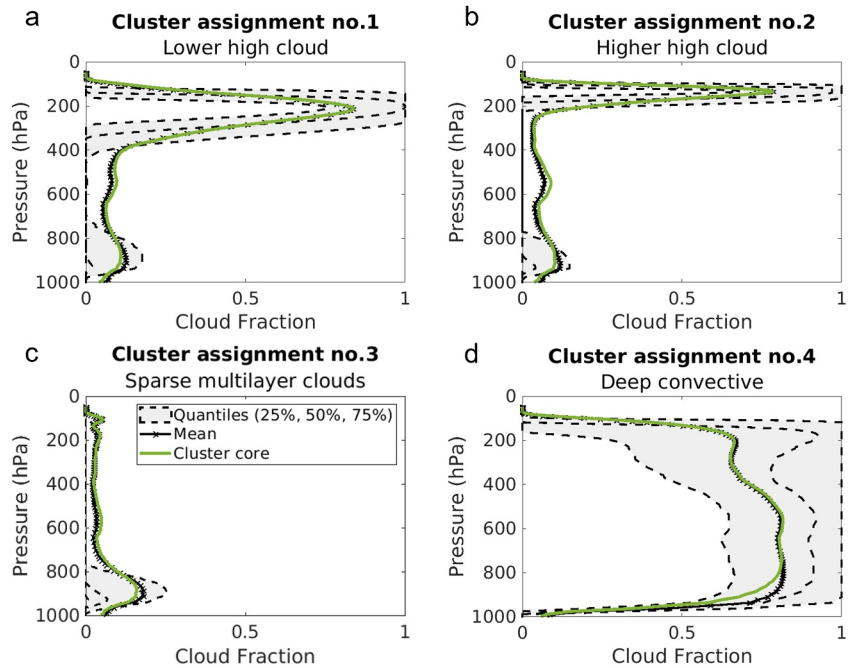

Figure 2. Same as Figures 1c–1f but for the second representation of the four clusters solution.

representation is statistically more robust than the other. A lower sum of point-to-centroid distances indicates a better clustering because it reflects the compactness of the groups (Section 2.2).

Two representations of four-cluster solutions are shown as Rep. 1 in Figure 1 and Rep. 2 in Figure 2. The most noticeable difference is that the second representation does not include any low cloud cluster as in Rep. 1 but instead creates another high-cloud cluster at a slightly higher altitude. The doubling of the high-cloud cluster is understandable in a mathematical sense because the distance to the centroid of the high cloud cluster in Rep. 1 is quite large (Table 1). The K-means offers an alternative solution in Rep. 2 to optimize this issue by splitting the high-cloud cluster into two subgroups, each of which has a much smaller sum of distances than the single high-cloud cluster in Rep. 1.

Such duplication of high-cloud clusters becomes more evident as we attempt to increase the number of clusters from four to eight (Figure 3). As the number of clusters increases, the K-means clustering tends to produce more high-cloud clusters, whereas the other three cloud regimes of low clouds, deep convective clouds, and sparse multilayer clouds are never replicated. The members of the newly produced high cloud clusters are decomposed from the old high cloud, low cloud, and sparse multilayer cloud clusters. This is reflected in Figure 3, where peaks of cloud fraction composites around 100 hPa in low cloud and sparse multilayer clouds profiles are reduced, and the quantile spreads in high cloud profiles become narrower as the clusters increase. A new type of cluster, to be later called the congestus cloud regime, appears for the numbers of clusters of seven and eight. The sum of distances and frequency of occurrence for the four to eight-cluster solution is presented in Table 2. These high cloud clusters have comparable occurrences with the low cloud clusters but yet produce the largest sum of

Table 1
The Sum of Point-to-Centroid Distances of the Two Representations of Four-Cluster Solutions

	Sparse multilayer clouds	Low cloud	High cloud	Deep convective	Total
Sum of distances ($\times 10^6$)					
Rep. 1	10.7	4.8	12.1	4.0	31.6
	Sparse multilayer clouds	Lower high cloud	Higher high cloud	Deep convective	
Rep. 2	11.9	7.8	6.7	4.5	30.9

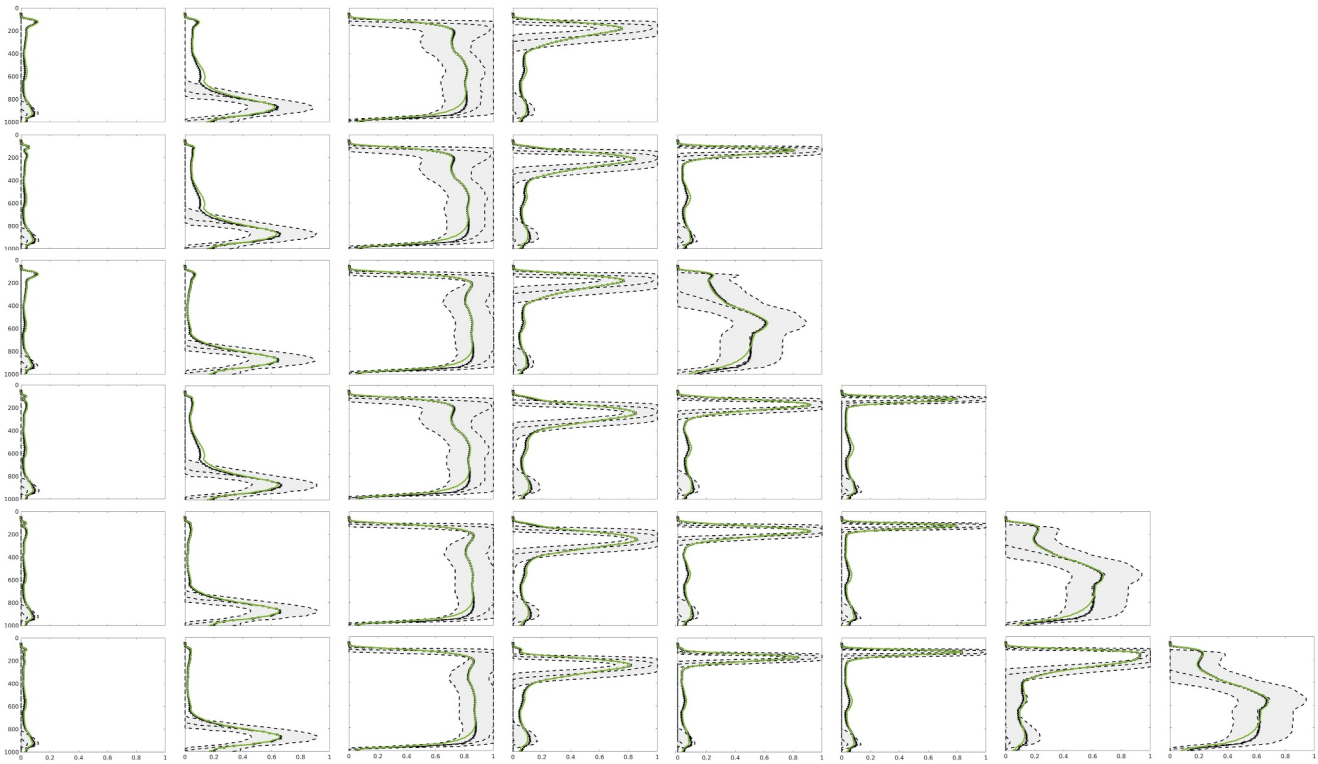

Figure 3. Same with Figures 1c–1f but for the four-to eight-cluster solutions.

distances to the centroid (4 CLS and 5 CLS statistics in Table 2). When the number of clusters increases, the K-means algorithm systematically improves the grouping performance in that the sum of distances to the centroid is reduced in every cluster producing more compact clusters. As mentioned before, however, these multiple high-cloud clusters seem redundant in the meteorological context. This problem is studied further below.

Table 2
Sum of Point-to-Centroid Distances and Frequencies of Occurrence of Four- to Eight-Cluster Solutions

# CLS	Sparse multilayer clouds	Low cloud	Deep convective	Congestus	High cloud 1	High cloud 2	High cloud 3	High cloud 4
Sum of distances (Occurrence-weighted sum of distance) ($\times 10^6$)								
4	10.7 (0.2)	4.8 (0.3)	4.0 (0.8)		12.1 (0.6)			
5*	10.2 (0.2)	3.0 (0.2)	2.3 (0.6)	3.4 (0.9)	11.3 (0.6)			
5	7.0 (0.1)	4.1 (0.3)	3.8 (0.8)		7.4 (0.6)	6.1 (0.4)		
6	5.6 (0.1)	3.8 (0.3)	3.6 (0.7)		4.8 (0.5)	4.8 (0.5)	3.9 (0.4)	
7	5.3 (0.1)	2.5 (0.2)	2.1 (0.6)	2.4 (0.8)	4.6 (0.5)	4.6 (0.4)	3.6 (0.3)	
8	4.5 (0.1)	2.4 (0.2)	1.9 (0.5)	2.4 (0.8)	3.4 (0.4)	2.8 (0.3)	3.6 (0.3)	3.1 (0.6)
Frequency of occurrence (%)								
4	61.7	13.7	5.1		19.4			
5*	60.5	12.9	3.9	3.9	18.8			
5	55.5	12.7	5.0		13.1	13.7		
6	52.8	12.2	4.9		8.7	10.6	10.8	
7	51.7	11.7	3.7	3.1	8.7	10.4	10.6	
8	49.7	11.4	3.6	3.1	7.6	8.8	10.3	5.5

Note. The Occurrence-weighted Sum Point-to-Centroid Distances are in Parentheses. 5* CLS is the five-cluster solution which includes the congestus cloud cluster.

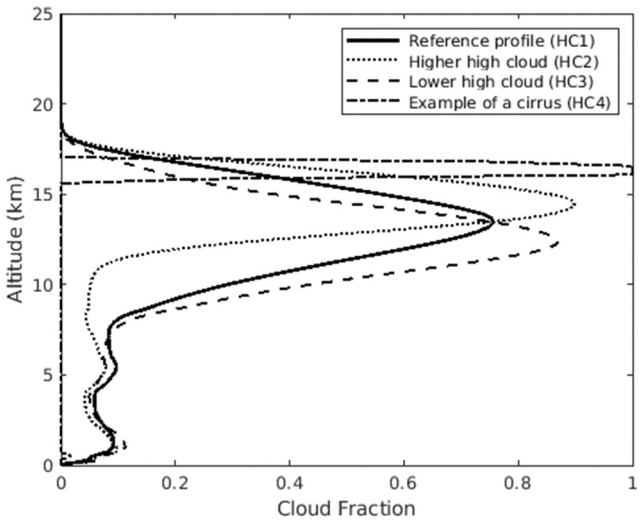

Figure 4. Schematic of high cloud profiles with different levels.

Figure 4 is a schematic figure comparing different high cloud profiles. The solid line (HC1) is the reference profile of a high cloud cluster from the four-cluster solution in Figure 1f. High clouds are divided into slightly higher and lower subgroups in the five-cluster solution in Figure 3. These doubled high cloud clusters are labeled as HC2 for the higher component and HC3 for the lower counterpart. The squared Euclidean distances between HC1-HC2 and HC1-HC3 are 2.35 and 0.94, respectively, explaining a considerable statistical gap that allowed the K-means clustering to separate HC2 and HC3 into different clusters. The last profile in the figure is an example of a thinner cirrus cloud (HC4), a member of the sparse multilayer cloud cluster. The distances between HC4 and the three other high cloud profiles of HC1, HC2, and HC3, are 10.9, 9.7, and 14, respectively. This shows that high cloud profiles are inclined to generate large squared Euclidean distances from one another, although these clusters all represent high clouds in a general sense of the word. Geometrically thin clouds having a delta-function-like cloud profile give rise to a large distance in between even though the profiles are practically similar except for a subtle difference in altitude.

An illustrative comparison of the distances produced by high cloud, low cloud, and deep convective cloud profiles is shown in Figure 5. It is demonstrated again that an instantaneous high and low cloud profiles can produce a large distance from each corresponding cluster centroid even for a slight level shift of the cloud fraction maximum. However, variability in cloud top height is much less pronounced for low clouds than high clouds since low clouds are typically bounded by the earth's surface below and the inversion layer above. As a result, low clouds tend to be fixed around similar levels and have little room for splitting into separate clusters with different altitudes unlike high clouds.

Besides the level shift, the high cloud cluster members can also generate a significant distance to the centroid when cloud fraction is double-peaked in the profile, which could also happen with deep-convective clouds. In the four-cluster solution (Figure 1), the deep-convective cluster, on average, gives a significant distance but is almost four times smaller in occurrence (8.83 mean distance, 5.51% occurrence) than the high cloud cluster (7.26 mean distance, 19.36% occurrence). The more frequent occurrence makes the high cloud cluster produce a much larger sum of distances than the deep convective cluster. It has also become evident that when the sum of distances is divided by the frequency of occurrences (Table 2, numbers in parentheses), the deep convective cluster generates the largest distance. Due to their infrequent occurrence, the K-Means algorithm chooses not to further subdivide these clusters, unlike the high cloud clusters.

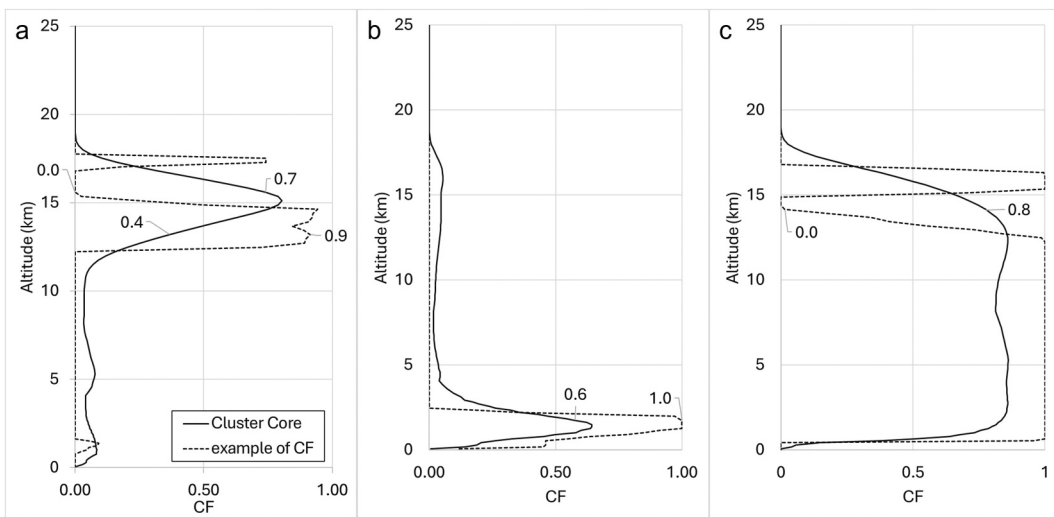

Figure 5. Cluster centroids (solid) and an example of cloud fraction profiles (dashed) for high cloud cluster (a), low cloud cluster (b), and deep convective cluster (c).

Table 3
Evaluation Indices of the Four- to Eight-Cluster Solutions

# CLS	Calinski Harabasz index $(\times 10^6)$	Davies Bouldin index
4	1.61	1.44
5	1.54	1.47
5*	1.31	1.58
6	1.39	1.53
7	1.26	1.56
8	1.18	1.54

Note. 5* CLS is the five-cluster solution which includes the congestus cloud cluster.

In summary, it is found that the large sum of distances within a high cloud cluster is ascribed to the geometrical thinness, relatively rich variability in altitude, and the abundance, which are all intrinsic to tropical high clouds. This could lead to an unrealistic multiplication of high-cloud clusters, which becomes more and more problematic as the total number of clusters is increased.

The multiple high cloud clusters are statistically independent in the K-means clustering but physically and visually alike. This investigation suggests that interpreting clustering results cannot merely rely on statistical evaluation but requires careful interpretation in the meteorological context as well. The first representation of the four-cluster solution (Figure 1) is adopted in the analysis that follows, ruling out the second representation (Figure 2) being redundant with duplicated high-cloud clusters.

The clustering results are evaluated in an objective manner using two mathematical indices mentioned in Section 2.2 (Table 3) in order to support the previous investigation. It is remembered that a higher value indicates a better clustering for the Calinski-Harabasz index, while for the Davies–Bouldin index, a smaller number indicates a better clustering solution. Both indices in Table 3 imply that the best number of clusters is four, backing up the argument above. Meanwhile, the congestus cloud cluster that emerges in the five, six, seven, and eight cluster solutions in Figures S2–S8 in Supporting Information S1 appears to serve as a physically plausible option as well. Johnson et al. (1999) revealed that congestus clouds are among the prominent cloud types that occur in tropical areas, along with shallow cumulus and cumulonimbus. In fact, among the possible representations of the five-cluster solution that we have scanned, there exists a realization that contains a congestus cloud regime (Figure S6 in Supporting Information S1) in addition to other four clusters practically identical to the four-cluster solution presented in Figure 1, namely the sparse multilayer clouds, low cloud, deep convective, and high cloud clusters.

This five-cluster representation is shown in Figure 6, with its statistics presented in Table 2 as 5* CLS. The occurrence of the congestus clouds (3.94%) appears to be comparable with the deep convective one (3.89%). The 5* CLS has a larger total sum of distance (3.03 E+7) compared with another five-cluster representation (Figures 3 and 5 CLS in Table 2 with a total sum of distance 2.82E+7). This may seem to suggest at a glance that 5 CLS is a

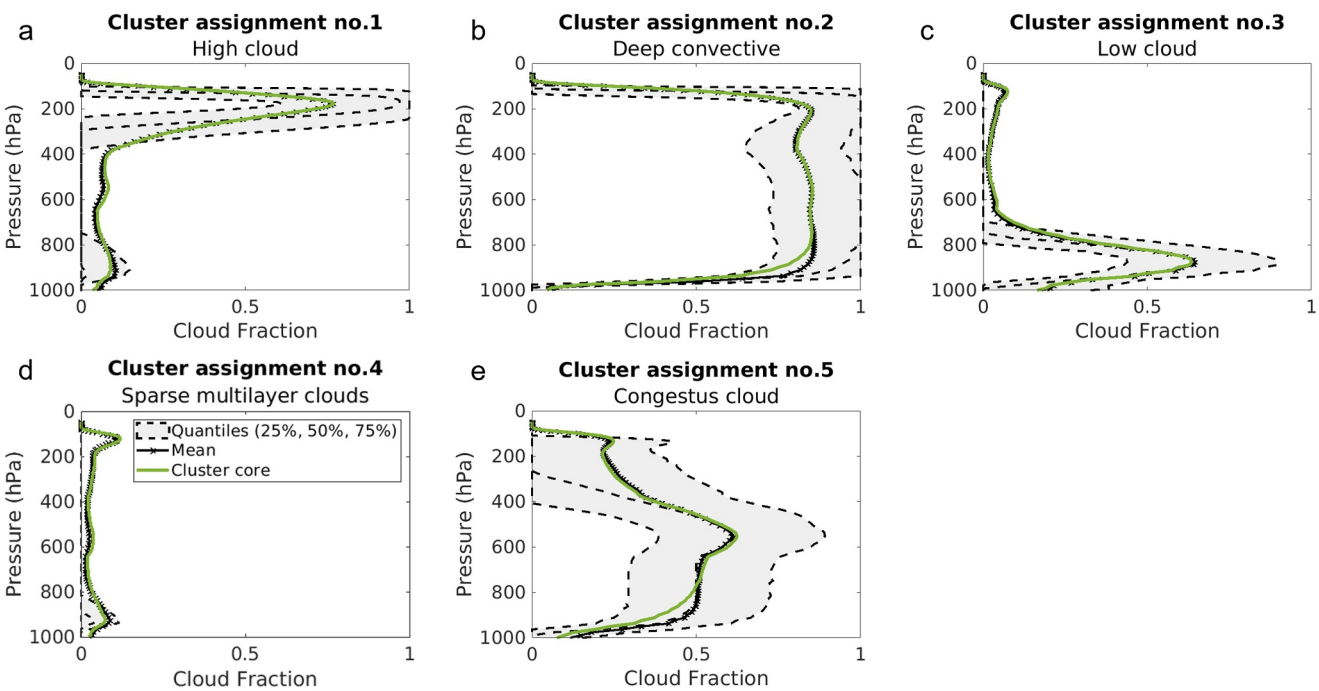

Figure 6. Same with Figures 1c–1f but for the five-cluster solution that includes congestus cloud cluster.

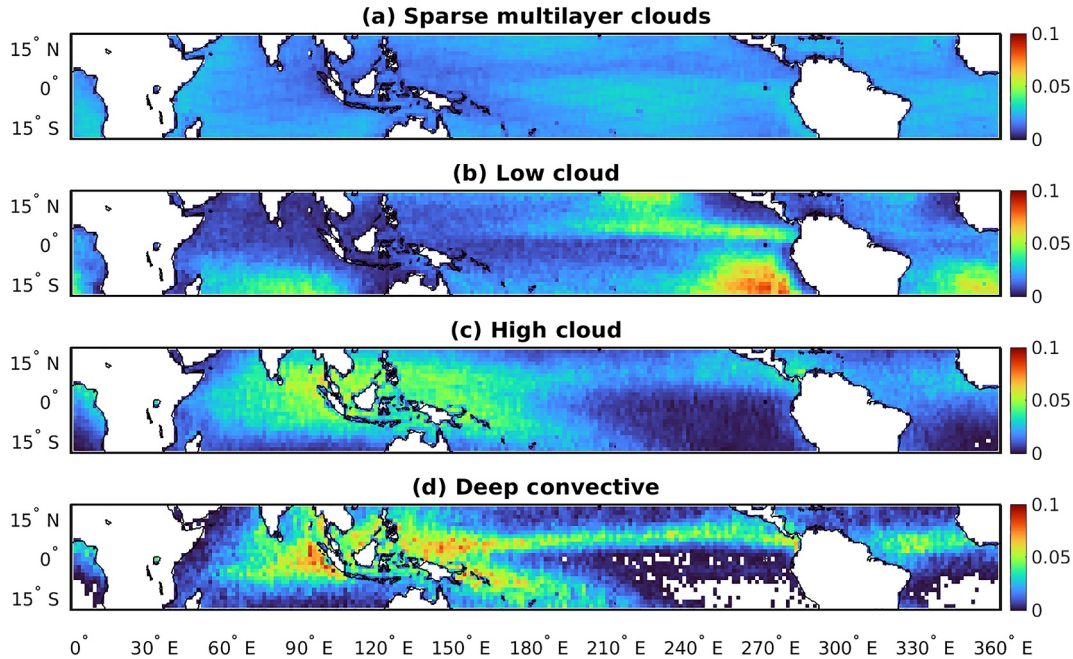

Figure 7. Members frequency of occurrence within ~1.5° grid for the four-cluster solution (%); the sparse multilayer clouds (a), low cloud (b), high cloud (c), and deep convective cloud (d) clusters.

better solution than 5* CLS. The 5 CLS solution resembles the 4 CLS cluster other than a second high-cloud cluster (see Figure 3 and High cloud 2 of 5 CLS in Table 2). The 5 CLS solution has a somewhat better performance (i.e., a smaller sum of distance) only to create separate high cloud profiles, which is artificial as previously discussed. The low cloud and deep convective clusters, on the other hand, show a smaller sum of distances in the 5* CLS solution than the 5 CLS solution. The sum of distances of the 5* CLS for the sparse multilayer clouds and high cloud clusters also reduces from the corresponding values for the four-cluster solution, which indicates that adding a congestus cloud cluster also improves the grouping of other clusters.

Based on the results so far, two clustering solutions are concluded to be most preferable. The four-cluster solution, which consists of sparse multilayer clouds, low cloud, high cloud, and deep convective cluster, is proven to be the best solution based on a statistical evaluation. On the other hand, the five-cluster solution including a congestus-cloud regime as well offers another possibility from meteorological perspectives. It is noted that we still retain the four-cluster version in this paper despite its incompleteness because the aim of this study is to demonstrate the technical limitation of clustering methods as well as the benefits. In the next section, the regime characteristics of these two clustering solutions are explored further.

3.2. Regime Characteristics

3.2.1. Spatial Distribution

The cluster members' spatial distributions are depicted in Figures 7 and 8 for the four and five-cluster solutions, respectively. The colors represent the frequency of occurrence obtained by dividing the tropical domain into 27 × 241 cells so the geographic coordinate system automatically creates each grid to have a size of around 1.5° × 1.5°. The frequency of occurrence is calculated within this grid size. The values of each grid cell are summed to 100% when integrated over the entire domain within each cloud type. The results are similar between the four and five clusters. The sparse multilayer cloud cluster member is broadly distributed everywhere across tropical latitudes rather than localized in any specific region. Areas of a high low-cloud occurrence are concentrated in the eastern Pacific and the southern Indian and Atlantic oceans. The prominent occurrence in the southeastern Pacific and Atlantic oceans is accounted for by stratocumulus clouds, known as the semi-permanent subtropical marine stratocumulus sheets (Wood, 2012). Stratocumulus clouds in these regions are related to the subtropical subsidence. The subsidence brings about a dry, stable airmass above the boundary layer, confining the

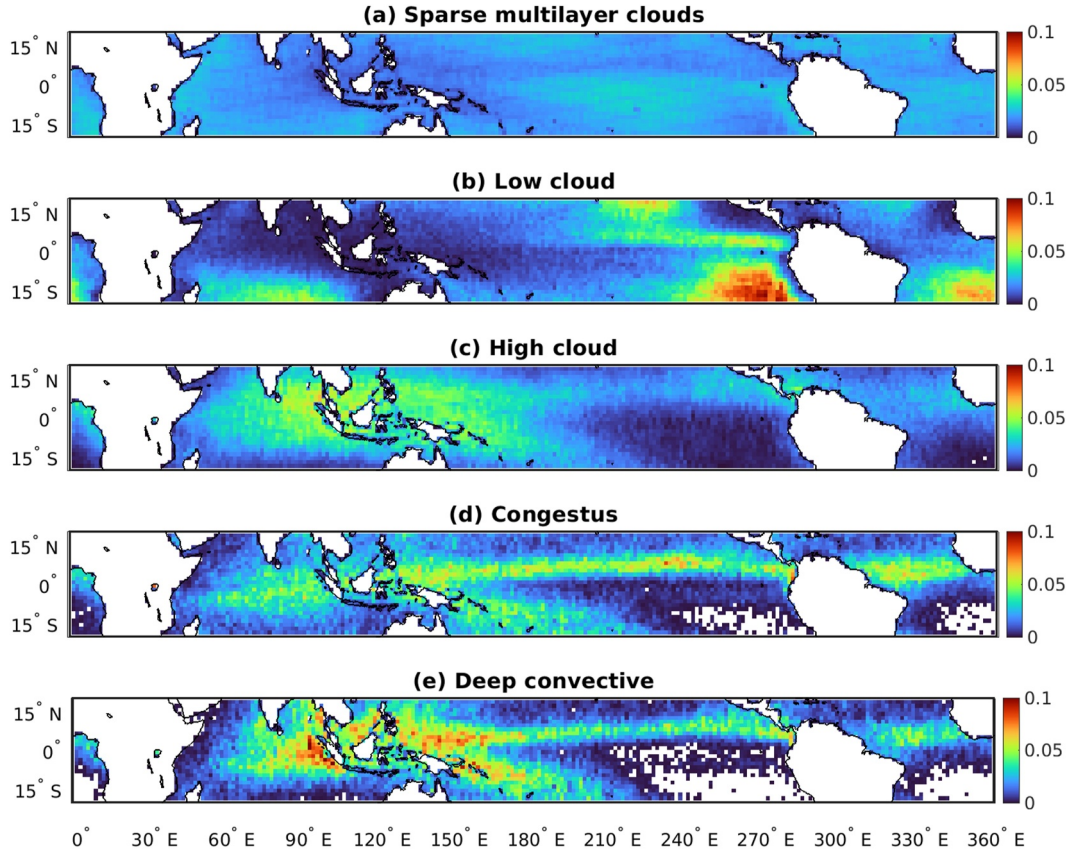

Figure 8. Same as Figure 7, but for the five cluster solution.

clouds within the boundary layer. This semi-permanent stratocumulus also exists over the northeast Pacific to the west of North America, although only a glimpse of it is seen in these figures. The stratocumulus in the northeast Pacific is also flanked by low clouds in the eastern Pacific ITCZ. This is probably related to the study by Wood and Hartmann (2006), which highlighted the mesoscale spatial variability of marine low clouds forming Mesoscale Cellular Convection (MCC) organization. These low clouds around 10°N east of the Pacific are shallow cumuli described as the "disorganized" type of MCC, mostly occurring away from the North and South American coastlines when the marine boundary layer deepens as the ocean surface warms toward the ITCZ. Low clouds south of the Indian Ocean are also oceanic stratocumuli, commonly found in the mid to high latitudes in association with outbreaks of cold air.

Deep convective clouds mainly occur in the Pacific ITCZs, Asian monsoon regions, and the Atlantic ITCZ. Concentrated around the maritime continent, the deep convective clouds developing in the warm, moist environment over the west Pacific drive easterly and westerly low-level inflows from the Pacific and Indian Oceans. The high cloud distribution overall follows the deep convective cloud distribution but extends somewhat more broadly. These high clouds can be anvils accompanying the deep convective clouds and *in situ* cirrus (Luo & Rossow, 2004). Anvil clouds are propagated by the detrainment from deep convection, then decay into cirrus after 6 hr, and they are most populated over the Indian and Pacific warm pools. However, not all cirrus clouds are associated with the high cloud cluster in this study. Thin cirrus away from the convective region are most likely to belong to the sparse multilayer cloud cluster. The high cloud cluster consists of thicker clouds, extending vertically from around 400 to 100 hPa in the vicinity of the convective regions.

The congestus cloud distribution is a category unique to the five-cluster solution (Figure 7d). The congestus distribution resembles the deep convective pattern. Congestus clouds prefer a relatively unstable environment as deep convection do, but the 0°C stable layer limits the growth of congestus clouds (Johnson et al., 1999). Congestus clouds appear more frequently in the northeast of the Pacific ITCZ than around the west Pacific and

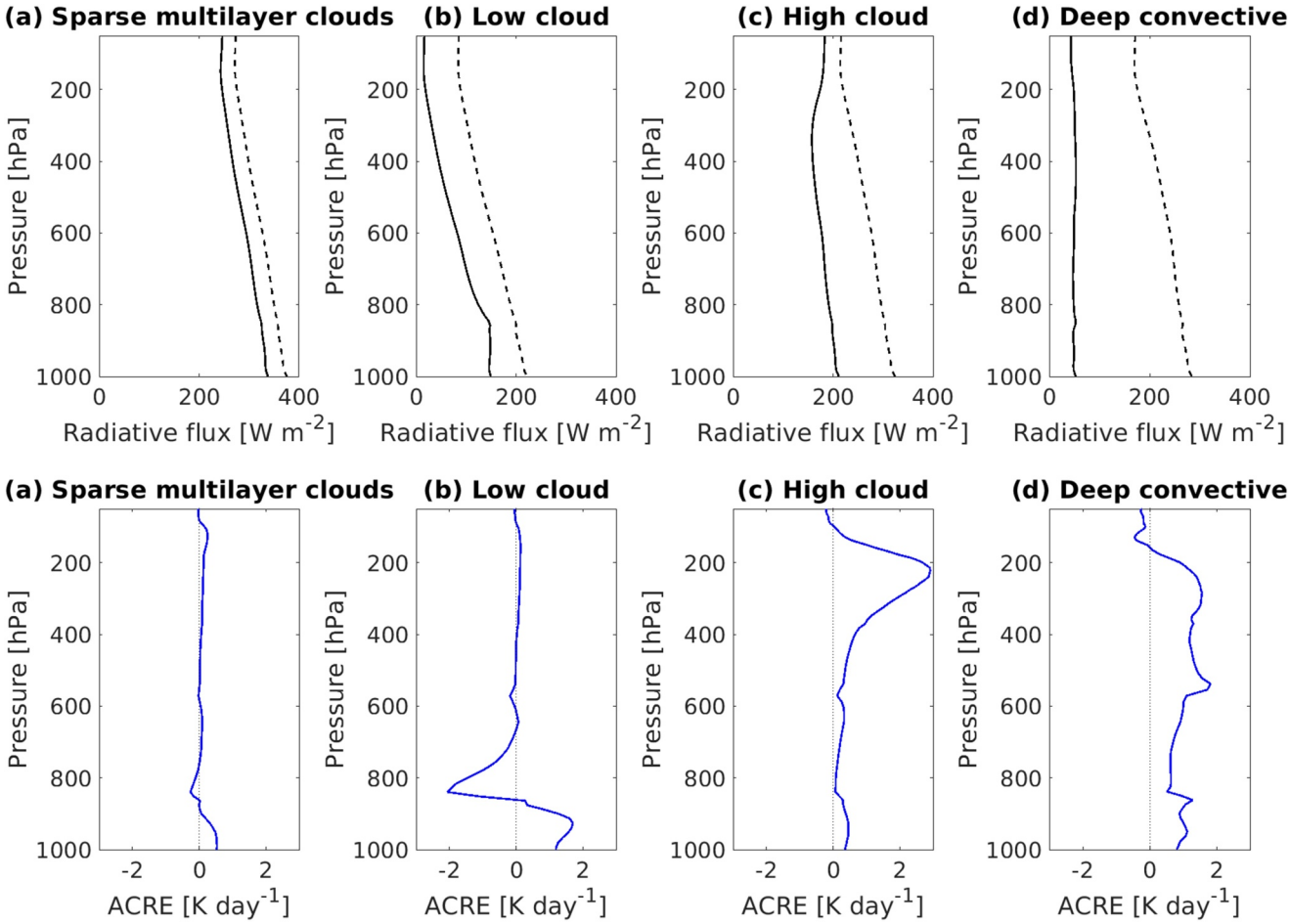

Figure 9. Composite of net radiative flux [W m^{-2}] profiles (upper) considering clouds and aerosols (solid black lines) and for clear-sky condition (dashed lines), and Atmospheric Cloud Radiative Effect [K day-1] profiles (lower) for the four-cluster solution.

Maritime continent. This is probably because the northeast Pacific area has a less warm SST than the west Pacific, causing a weak low-level forcing and hence preventing the congestus clouds from reaching a higher altitude.

3.2.2. Radiative Properties and Cloud Radiative Heating

Figures 9 and 10 depict the radiative properties for the four and five-cluster solutions. The upper rows in both figures are the composite net (shortwave plus longwave) flux profiles. The solid black lines are the profiles that include the cloud and aerosol impacts, while the dashed lines are those under the clear-sky condition. The lower rows are the cloud radiative heating or the Atmospheric Cloud Radiative Effect (ACRE) calculated as the difference between the all- and clear-sky heating rate, which includes the aerosol impact. The profiles with aerosols are almost indistinguishable from those without (not shown here), suggesting that aerosols have only a very minor influence on cloud radiative properties for the present analysis. Tables 4 and 5 list the net fluxes at TOA and bottom (BOA) for the all-sky and clear-sky conditions and their differences. The differences here (all-sky minus clear-sky net radiative flux) represent the clouds' cooling effect when negative.

The net radiation flux generally decreases with altitude except where very intense convection exists (Figure 10e and Table 5). The sparse multilayer clouds profiles in the four and five-cluster solutions are similar except a slight difference, as confirmed by Tables 4 and 5. The net radiative flux at BOA is around 340–343 W m^{-2} under the clouds and aerosol influence and around 378–381 W m^{-2} under the clear-sky conditions. At the TOA, the net radiative flux is reduced to around 254–259 W m^{-2} and 281–285 W m^{-2} with the cloud and aerosol influence and without, respectively. The sparse multilayer clouds ACRE profiles in the four and five clusters closely reflect the corresponding cloud fraction profiles (Figures 1e and 6d). The cloud fraction profiles depict the heights of thin

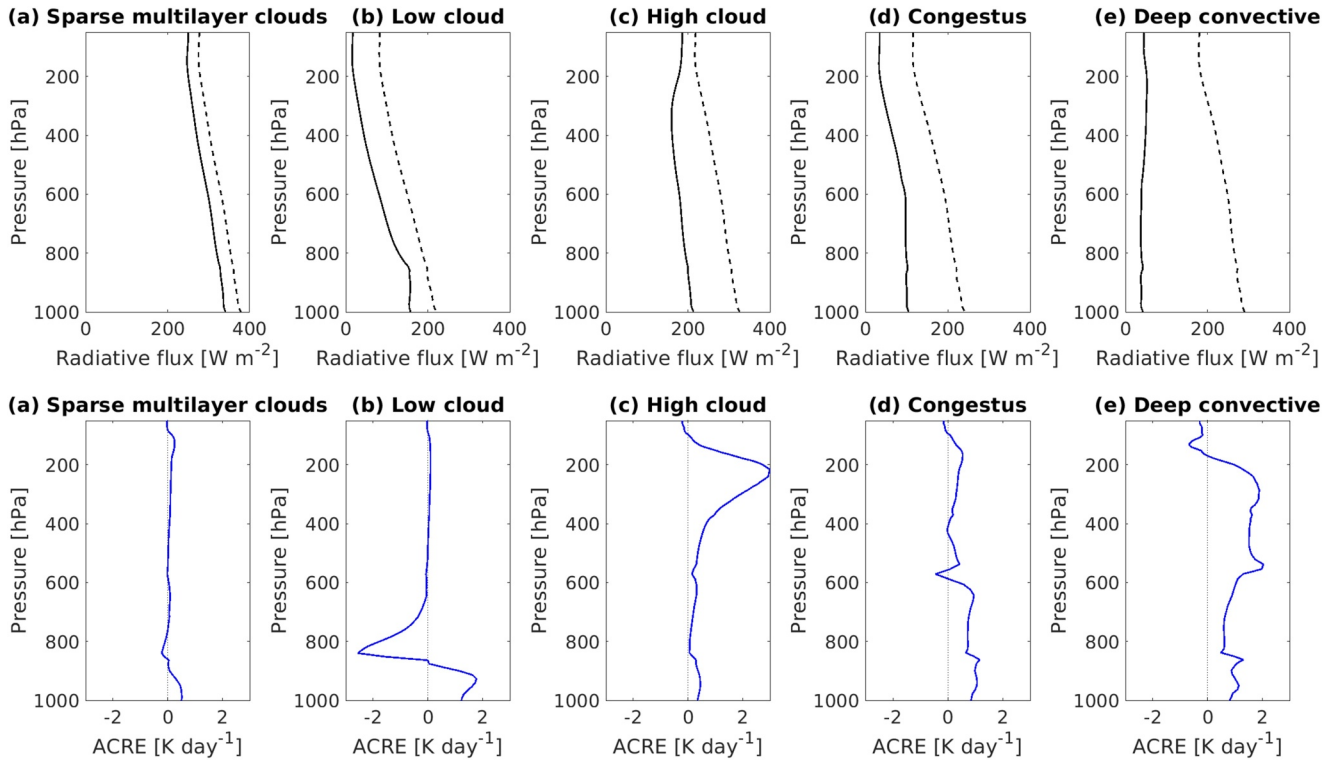

Figure 10. Same as Figure 9, but for results of the five-cluster solution.

clouds below 800 hPa and above 200 hPa, capped by the trade wind inversion layer and the tropopause. The cloud top cooling and in-cloud warming are visible in the ACRE profile.

As shown in Figures 7b and 8b, a majority of low clouds are found over subtropical oceans, where the incoming solar radiation is climatologically nearly balanced with outgoing longwave radiation. This explains why the all-sky net TOA flux for low clouds is as small as about 18.8 W m^{-2} in the four cluster results and 19.5 W m^{-2} in the five cluster results. Low clouds give a warming effect near the surface, paired by an intense cooling of −2 K day-1 near the inversion layer. The intense composite cooling reflects a cloud top cooling from boundary-layer clouds capped by the trade inversion. These clouds have a nearly constant top height bound from above and a hardly variable thickness limited also from below by the surface. Consequently, the pair of in-cloud warming and cloud top cooling remains conspicuous in the composite ACRE without being averaged out.

The net radiation flux for the high cloud cluster shows a distinct cloud impact around 200 hPa (Figures 9c and 10c). The clouds' radiative impact also affects the BOA flux, with a flux difference almost reaching 113 W m^{-2} (Tables 4 and 5). Like the low clouds, the high clouds show a warming effect inside and below the cloud layer, although reduced sharply toward the cloud top, this time above 200 hPa. The warming effect is stronger than the other cloud types, reaching almost 3 K day-1. However, the cloud top cooling is not as intense as low clouds since

Table 4
Composite of Net Radiative Flux [W m^{-2}] at the Top of the Atmosphere and BOA in the All-Sky and Clear-Sky Conditions and Their Differences From the Four-Cluster Solution

	TOA			BOA		
	All-sky	Clear-sky	Differences	All-sky	Clear-sky	Differences
Sparse multilayer clouds	254.0	280.9	−26.9	339.3	377.6	−38.4
Low cloud	18.8	87.2	−68.4	152.4	228.4	−76.0
Deep convective	46.9	175.0	−128.1	54.6	287.7	−233.0
High cloud	187.6	221.6	−34.0	213.1	326.1	−113.0

Table 5
Same as Table 4, But for the Five-Cluster Solution

	TOA			BOA		
	All-sky	Clear-sky	Differences	All-sky	Clear-sky	Differences
Sparse multilayer clouds	258.5	285.0	−26.5	342.7	380.7	−38.1
Low cloud	19.5	85.0	−65.5	161.2	227.9	−66.6
Deep convective	48.4	184.0	−135.6	44.2	294.5	−250.3
Congestus cloud	37.3	118.8	−81.5	106.9	245.2	−138.4
High cloud	190.9	224.4	−33.4	215.5	328.3	−112.8

high clouds are more variable in height and thickness than low clouds. The cloud top cooling in one high cloud might be compensated by the warming by another cloud at a slightly higher level, producing less significant cooling than the low cloud composite and, thus, a deeper vertical extent of warming. As such, the radiative heating profiles shown here are qualitatively different from what would be expected from an individual snapshot. It is noted, however, that the mean profile with a pair of cooling and warming averaged out is still useful for climatological applications where the statistical ensemble of many cloud profiles matters more than individual snapshots.

The presence of deep convective clouds reduces the clear-sky flux most dramatically at all levels to the point where there are almost indiscernible differences between the TOA and BOA fluxes (BOA minus TOA fluxes ~ 8 W m^{-2} on the four clusters and ~ -4 W m^{-2} on the five clusters). The congestus profile gives a similar vertically constant structure of all-sky flux from 600 hPa to the surface (Figure 10d, upper) with a BOA flux difference of around 138.4 W m^{-2} (Table 5). The deep convective clouds have the most significant radiative impact, reaching 233 W m^{-2} in the four clusters and 250.3 W m^{-2} in the five clusters at BOA.

The ACRE profiles for both congestus (Figure 10d, lower) and deep convective (Figures 9d and 10e, lower) depict three local maxima, presumably corresponding to the three inversion layers referred to by Johnson et al. (1999): the trade wind inversion around 850 hPa, the 0°C level around 550 hPa, and the tropopause near 100 hPa. The first two stable layers can also be seen in the high cloud's ACRE in both cluster solutions but with a weaker intensity. A subtle sign of the 0°C freezing layer also appears in the low-cloud ACRE profiles of the four clusters but disappears in the five-cluster solution as the congestus group was added.

The 0°C layer gives opposite radiative impacts between the congestus and the deep convective regimes. The stable 0°C layer is produced by the melting of frozen hydrometeors precipitating through this layer, which occurs mostly inside or in proximity to the precipitating system (Johnson et al., 1996). The sharp cooling in the congestus ACRE profile suggests a cloud top cooling from congestus clouds. The stable layer prevents congestus clouds from growing further upward and rather promotes cloud detrainment. As a result, cloud fractions are maximized at or slightly above this layer, generating cloud top radiative cooling through infrared emission and longwave heating beneath.

As has been highlighted in previous studies, CloudSat and CALIPSO observations can effectively capture multilayer clouds (Hang et al., 2019; Tselioudis et al., 2013). The ACRE profiles here, as well as the composite cloud fraction profiles in Figures 1 and 6, implicitly indicate patterns of multilayer clouds. For example, the sparse multilayer clouds ACRE profiles are positive near the surface and above 200 hPa. Additionally, the high cloud ACRE profiles suggest the presence of thin low clouds below 800 hPa, while the congestus ACRE profiles indicate overlying high clouds around 200 hPa. Although the ACRE profiles do not clearly depict this, the low cloud composite cloud fraction profiles (Figures 1c and 6c) also reveal the existence of thin high clouds above them.

Overall, this discussion demonstrates that the clustering results from this study are of potential utility for analyzing the cloud radiative effects.

4. Conclusions

A methodology was developed to objectively classify tropical clouds into distinguishable regimes. Based on a K-means clustering analysis, cloud fraction profiles observed over 4 years in the tropics were decomposed into groups. In order to find the "best number" of clusters, several k numbers from four to eight clusters were tested. The clustering solutions were evaluated using statistical measures and tools: the sum of point-to-centroid distance, frequency of occurrence, Calinski-Harabasz index, and Davies Bouldien index. Uncertainties in the K-means clustering were found to be caused by two factors of the stochastic nature in selecting the initial cluster centroids and of technical challenges in determining the high cloud cluster. To tackle the first problem, the K-means algorithm is run repeatedly with different initial guesses to find the most proper clustering solution. There are two distinct representations of four clusters produced by this procedure. The first representation consists of the sparse multilayer clouds, low cloud, high cloud, and deep convective clusters. The second representation excludes the low cloud cluster and instead divides the high cloud cluster into two groups having slightly different altitudes. The evaluation of the second representation using the sum of point-to-centroid distances and frequency of occurrence suggests the second source of uncertainty, or an artificial multiplication of the high cloud cluster. The main reason for this problem is that the high cloud cluster tends to produce a large sum of point-to-centroid distance compared to other cloud types. This is because geometrically thin clouds with even slightly different altitudes can be mathematically distant from one another despite the morphological similarity. High clouds are a frequently occurring component of tropical clouds, making duplicated high-cloud clusters each constitute an appreciable size of samples. This allows the K-means clustering to create more and more redundant high-cloud clusters as the number of clusters is increased. Based on our evaluation, producing more than one high-cloud cluster is unnecessary. The high cloud clusters generated this way are statistically distant from one another but physically and visually very similar. It is suggested that interpreting clustering results cannot merely rely on automated statistical evaluation and requires careful interpretations from the meteorological points of view.

Based on the two evaluation indices, the first representation of the four clusters including the low-cloud cluster instead of a second high-cloud cluster is found to be the most optimum solution. On the other hand, a congestus cloud cluster that appears in a certain representation of five-cluster solutions offers another physically plausible classification. It is concluded that two clustering solutions, the four and five clusters with the latter including the congestus cloud cluster, are accepted as the optimal choices. The spatial distributions and the radiative properties of corresponding cluster pairs from the two clustering solutions are consistent with each other. Adding a congestus cloud cluster does not disturb other regimes' spatial distribution and radiative properties. The four-cluster solution is the best statistically, but the five-cluster solutions provide another useful interpretation of the tropical cloud regimes.

The cluster analysis results from this study will have a range of applications explored in future studies. We are particularly interested in investigating the physical processes governing the theories explaining the response of tropical anvil clouds to changing climate, as it requires a proper method to identify anvil clouds, deep convective clouds, and clear sky areas from observation data.

Data Availability Statement

The CCCM Release D1 data product is open for public and can be found in the EARTHDATA Atmospheric Science Data Center of NASA (NASA/LARC/SD/ASDC, 2011) (https://doi.org/10.5067/AQUA/CERES/CCCM-FM3-MODIS-CAL-CS_L2.RELD1).

Acknowledgments
This research is part of a doctoral study supported by the Niterra scholarship fund for overseas students.

References

Anderberg, M. R. (1973). *Cluster analysis for applications*. Academic Press.

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Soda '07: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 1027–1035). New Orleans, Louisiana: Society for Industrial and Applied Mathematics.

Bony, S., Stevens, B., Coppin, D., Becker, T., Reed, K. A., Voigt, A., & Medeiros, B. (2016). Thermodynamic control of anvil cloud amount. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(32), 8927–8932. https://doi.org/10.1073/pnas.1601472113

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Chéruy, F., & Aires, F. (2009). Cluster analysis of cloud properties over the southern european mediterranean area in observations and a model. *Monthly Weather Review*, *137*(10), 3161–3176. https://doi.org/10.1175/2009mwr2882.1

Cho, N., Tan, J., & Oreopoulos, L. (2021). Classifying planetary cloudiness with an updated set of MODIS cloud regimes. *Journal of Applied Meteorology and Climatology*, *60*(7), 981–997. https://doi.org/10.1175/jamc-d-20-0247.1

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Gordon, N. D., Norris, J. R., Weaver, C. P., & Klein, S. A. (2005). Cluster analysis of cloud regimes and characteristic dynamics of midlatitude synoptic systems in observations and a model. *Journal of Geophysical Research*, *110*(D15), D15S17. https://doi.org/10.1029/2004jd005027

Ham, S.-H., Kato, S., Rose, F. G., Sun-Mack, S., Chen, Y., Miller, W. F., & Scott, R. C. (2022). Combining cloud properties from CALIPSO, CloudSat, and MODIS for top-of-atmosphere (TOA) shortwave broadband irradiance computations: Impact of cloud vertical profiles. *Journal of Applied Meteorology and Climatology*, *61*(10), 1449–1471. https://doi.org/10.1175/jamc-d-21-0260.1

Ham, S.-H., Kato, S., Rose, F. G., Winker, D., L'Ecuyer, T., Mace, G. G., et al. (2017). Cloud occurrences and cloud radiative effects (CREs) from CERES-CALIPSO-CloudSat-MODIS (CCCM) and CloudSat radar-lidar (RL) products. *Journal of Geophysical Research: Atmospheres*, *122*(16), 8852–8884. https://doi.org/10.1002/2017jd026725

Hang, Y., L'Ecuyer, T. S., Henderson, D. S., Matus, A. V., & Wang, Z. (2019). Reassessing the effect of cloud type on earth's energy balance in the age of active spaceborne observations. Part II: Atmospheric heating. *Journal of Climate*, *32*(19), 6219–6236. https://doi.org/10.1175/jcli-d-18-0754.1

Hartmann, D. L., & Larson, K. (2002). An important constraint on tropical cloud - Climate feedback. *Geophysical Research Letters*, *29*(20), 12–1–12–4. https://doi.org/10.1029/2002gl015835

Johnson, R. H., Ciesielski, P. E., & Hart, K. A. (1996). Tropical inversions near the 0°C level. *Journal of the Atmospheric Sciences*, *53*(13), 1838–1855. https://doi.org/10.1175/1520-0469(1996)053<1838:tintl>2.0.CO;2

Johnson, R. H., Rickenbach, T. M., Rutledge, S. A., Ciesielski, P. E., & Schubert, W. H. (1999). Trimodal characteristics of tropical convection. *Journal of Climate*, *12*(8), 2397–2418. https://doi.org/10.1175/1520-0442(1999)012<2397:tcotc>2.0.CO;2

Kato, S., Rose, F. G., Sun-Mack, S., Miller, W. L., Chen, Y., Rutan, D. A., et al. (2011). Improvements of top-of-atmosphere and surface irradiance computations with CALIPSO-CloudSat-and MODIS-Derived cloud and aerosol properties. *Journal of Geophysical Research*, *116*(D19), D19209. https://doi.org/10.1029/2011jd016050

Kato, S., Sun-Mack, S., Miller, W. F., Rose, F. G., Chen, Y., Minnis, P., & Wielicki, B. A. (2010). Relationships among cloud occurrence frequency, overlap, and effective thickness derived from CALIPSO and CloudSat merged cloud vertical profiles. *Journal of Geophysical Research*, *115*(D4), D00H28. https://doi.org/10.1029/2009jd012277

Lee, D., Oreopoulos, L., & Cho, N. (2025). Regimes of cloud vertical structure from active observations. *Journal of Geophysical Research: Atmospheres*, *130*(1), e2024JD041716. https://doi.org/10.1029/2024jd041716

Lindzen, R. S., Chou, M.-D., & Hou, A. Y. (2001). Does the Earth have an adaptive infrared Iris? *Bulletin of the American Meteorological Society*, *82*(3), 417–432. https://doi.org/10.1175/1520-0477(2001)082<0417:DTEHAA>2.3.CO;2

Luo, Z., & Rossow, W. B. (2004). Characterizing tropical cirrus life cycle, evolution, and interaction with upper-tropospheric water vapor using lagrangian trajectory analysis of satellite observations. *Journal of Climate*, *17*(23), 4541–4563. https://doi.org/10.1175/3222.1

Luo, Z. J., Anderson, R. C., Rossow, W. B., & Takahashi, H. (2017). Tropical cloud and precipitation regimes as seen from near-simultaneous TRMM, CloudSat, and CALIPSO observations and comparison with ISCCP. *Journal of Geophysical Research: Atmospheres*, *122*(11), 5988–6003. https://doi.org/10.1002/2017jd026569

NASA/LARC/SD/ASDC. (2011). CERES A-Train integrated CALIPSO, CloudSat, CERES, and MODIS (CCCM) merged release D1 [Dataset]. *NASA Langley Atmospheric Science Data Center DAAC*. https://doi.org/10.5067/AQUA/CERES/CCCM-FM3-MODIS-CAL-CS_L2.RELD1

Ohno, T., & Satoh, M. (2018). Roles of cloud microphysics on cloud responses to sea surface temperatures in radiative-convective equilibrium experiments using a high-resolution global nonhydrostatic model. *Journal of Advances in Modeling Earth Systems*, *10*(8), 1970–1989. https://doi.org/10.1029/2018ms001386

Oreopoulos, L., Cho, N., Lee, D., Kato, S., & Huffman, G. J. (2014). An examination of the nature of global MODIS cloud regimes. *Journal of Geophysical Research: Atmospheres*, *119*(13), 8362–8383. https://doi.org/10.1002/2013jd021409

Ramanathan, V., & Collins, W. (1991). Thermodynamic regulation of ocean warming by cirrus clouds deduced from observations of the 1987 El Niño. *Nature*, *351*(6321), 27–32. https://doi.org/10.1038/351027a0

Rossow, W. B., Tselioudis, G., Polak, A. A., & Jakob, C. (2005). Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures. *Geophysical Research Letters*, *32*(21), L21812. https://doi.org/10.1029/2005gl024584

Simpson, J. (1992). Global circulation and tropical cloud activity. In *The global role of tropical rainfall* (pp. 77–92). Deepak Publishing.

Tselioudis, G., Rossow, W., Zhang, Y., & Konsta, D. (2013). Global weather states and their properties from passive and active satellite cloud retrievals. *Journal of Climate*, *26*(19), 7734–7746. https://doi.org/10.1175/jcli-d-13-00024.1

Wood, R. (2012). Stratocumulus clouds. *Monthly Weather Review*, *140*(8), 2373–2423. https://doi.org/10.1175/mwr-d-11-00121.1

Wood, R., & Hartmann, D. L. (2006). Spatial variability of liquid water path in marine low cloud: The importance of mesoscale cellular convection. *Journal of Climate*, *19*(9), 1748–1764. https://doi.org/10.1175/jcli3702.1

Zelinka, M. D., & Hartmann, D. L. (2010). Why is longwave cloud feedback positive? *Journal of Geophysical Research*, *115*(D16), D16117. https://doi.org/10.1029/2010jd013817

Zhang, Y., Klein, S., Mace, G. G., & Boyle, J. (2007). Cluster analysis of tropical clouds using CloudSat data. *Geophysical Research Letters*, *34*(12), L12813. https://doi.org/10.1029/2007gl029336

Zhang, Y., Klein, S. A., Boyle, J., & Mace, G. G. (2010). Evaluation of tropical cloud and precipitation statistics of community atmosphere model version 3 using CloudSat and CALIPSO data. *Journal of Geophysical Research*, *115*(D12), D12205. https://doi.org/10.1029/2009jd012006