

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

DISSERTATION

PROPERTIES OF THE TROPICAL HYDROLOGIC CYCLE AS ANALYZED
THROUGH 3-DIMENSIONAL *K*-MEANS CLUSTER ANALYSIS

Submitted by

Matthew Alan Rogers

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2008

UMI Number: 3332703

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 3332703

Copyright 2008 by ProQuest LLC.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.


ProQuest LLC
789 E. Eisenhower Parkway
PO Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

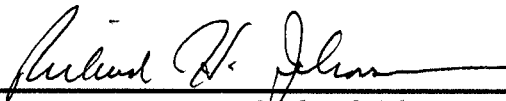
April 11th, 2008

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY MATTHEW ALAN ROGERS ENTITLED PROPERTIES OF THE TROPICAL HYDROLOGIC CYCLE AS ANALYZED THROUGH 3-DIMENSIONAL K-MEANS CLUSTER ANALYSIS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

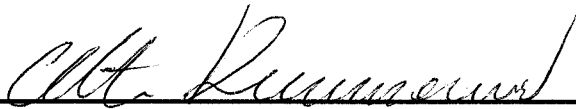
Committee on Graduate Work



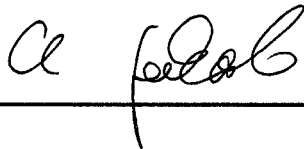
Graeme Stephens, Advisor



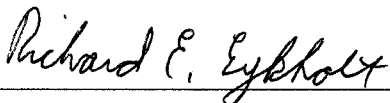
Richard Johnson, Department Head and Committee Member



Christian Kummerow, Committee Member



Christian Jakob, Committee Member



Richard Eykholt, Department of Physics, Outside Committee Member

ABSTRACT OF DISSERTATION
PROPERTIES OF THE TROPICAL HYDROLOGIC CYCLE AS ANALYZED
THROUGH 3-DIMENSIONAL *K*-MEANS CLUSTER ANALYSIS

As the primary locations of deep convective activity and unrestrained tropical wave dynamics, the tropical West Pacific and East Indian oceans are among the most important regions in the tropics. Given that most of the region consists of unpopulated expanses of ocean, observations of tropical atmospheric properties in this important region is exceptionally difficult. Only with the help of satellite observations are we capable of gleaning valuable data from this region, and our utilization of advanced analysis techniques allows us to gain more from these observations than would otherwise be possible.

In that vein, this dissertation reports on the use of a unique statistical technique, long known to other fields of research, as applied to a combined-instrument satellite observation dataset over the warm pool region of the tropical West Pacific ocean. The statistical technique, known as *k*-means cluster analysis, is used to delineate self-similar populations of cloud type, hereafter referred to as cloud *regimes*, from frequency-distribution histograms of cloud-top height, cloud optical thickness, and rainfall amount. We will show that four primary cloud regimes exist in the tropical region discussed, that the four regimes vary primarily through differences in convective activity, and that these four cloud regimes exist in a coherent temporal structure that explains the long-observed variability in convective activity seen in the tropics.

Combining this regime information with satellite observations, along with reanalysis data, we then examine the individual properties of each cloud regime. These observations give us the means to understand the forcings behind cloud regime change in the region. We confirm the structural properties of these regimes using analysis from a cloud-resolving model, and apply our new understanding of the mechanism behind this large-scale forcing to the governance of the tropical hydrologic cycle as a whole. The insights gained from this analysis have benefits to both the fields of atmospheric remote sensing, and of cloud- and climate modeling of the tropical atmosphere.

Applications of this technique are of particular interest to researchers developing retrieval algorithms for latent heat profiles using active sensors such as the cloud-profiling radar aboard CloudSat.

Matthew Alan Rogers
Department of Atmospheric Science
Colorado State University
Fort Collins, Colorado 80523
Summer 2008

“A cloud does not know why it moves in just such a direction and at such a speed...It feels an impulsion...this is the place to go now. But the sky knows the reasons and the patterns behind all clouds, and you will know, too, when you lift yourself high enough to see beyond horizons.”

-- Richard Bach, American author

Acknowledgements

The 2-D *k*-means cluster analysis code which served as the basis for the development of the 3-D code used in this research was graciously provided by Drs. Christian and Doerte Jakob, respectively of Monash University and the BMRC in Melbourne, Australia. MODIS Level-2 cloud observations were obtained courtesy of the NASA GSFC DAAC data gateway. AMSR/E observations of rainfall were provided by the NSIDC DAAC data gateway. CERES observations of OLR were provided courtesy of the NASA Langley ASDC data gateway. NCEP/NCAR reanalysis products were provided courtesy of the Earth System Research Laboratory of NOAA. RAMS model runs were kindly provided by Dr. Sue van den Heever. Additionally, I would like to thank my committee, consisting of my advisor, Prof. Graeme Stephens, along with Prof. and Department Head Richard Johnson and Prof. Christian Kummerow, Prof. Christian Jakob of Monash University, and Prof. Richard Eykholt of the Department of Physics at Colorado State University. Members of the Stephens Research Group have provided great insights into this research, especially those of Dr. Lyle Pakula and John Haynes. Ian Wittmeyer and Natalie Tourville have been instrumental in providing systems support for the computers used in this research. Finally, I would like to thank my family, especially my wife Karin, and all our friends, for supporting me along this long and winding path.

Contents

Abstract	iii
Acknowledgements	vi
Table of Contents	vii
List of Figures	xi
List of Tables	xiii
Chapter 1: Introduction and Motivation	
1.1 Background	1
1.2 Motivation	6
1.3 Outline of Research	8
Chapter 2: Cloud Clustering as a Diagnostic Tool	
2.1 The k -means algorithm	10
2.2 The Method of Jakob and Tselioudis	13
2.3 k -means Cluster Analysis using MODIS Level-2 Cloud Product Data	
2.3.1 Background and Dataset Description	20
2.3.2 Spatial and Temporal Selection, and discussion on choosing k	23
2.3.3 Cluster Results – MODIS-determined regimes	26
2.3.4 Bulk Properties of Regimes, and Comparison to Jakob et al.	32
2.4 Discussion	36

Chapter 3: Towards Development of a Three-Dimensional Cluster Analysis

3.1 Motivation	39
3.2 The AMSR/E Precipitation Product	39
3.3 Selection of a Suitable 3D Histogram	40
3.4 Selection of value for k in 3D Histogram Cluster Analysis	47
3.4.1 Results for $k=3$	55
3.4.2 Results for $k=4$	60
3.4.3 Results for $k=5$	65
3.4.4 Results for $k=6$	71
3.4.5 Discussion on selection of k	75
3.5 Discussion	75

Chapter 4: Properties of the Tropical West Pacific from 3-Dimensional Cluster Analysis

4.1 Introduction	78
4.2 Description of 3-D Regimes	79
4.2.1 Cluster 3 – Trade Cumulus and Cirrus Regime	80
4.2.2 Cluster 1- ‘Enhanced’ Cumulus Regime	83
4.2.3 Cluster 2 – Isolated Convection	85
4.2.4 Cluster 4 – Organized Deep Convection	87
4.3 Temporal Characteristics of 3D Regimes	89
4.4 External Observations of 3D Regimes	92
4.4.1 Radiative Characteristics	94
4.4.2 Thermodynamic Characteristics	98

4.4.2.a 850mb zonal wind	99
4.4.2.b SST	100
4.4.2.c TCWV	102
4.4.2.d Lifted Index	104
4.4.2.e Vertical Velocity	106
4.5 Summary of Results and Discussion	110
Chapter 5: Cluster Analysis of Modeled Results and Comparison to Observations	
5.1 Introduction	112
5.2 RAMS simulation of Tropical West Pacific convection	113
5.3 Cluster analysis of simulated Tropical West Pacific convection	116
5.4 Properties of individual cluster and comparison to observations	
5.4.1 Cluster 1 – Trade Cumulus and Cirrus Regime	123
5.4.2 Cluster 2 – Enhanced Cumulus Regime	125
5.4.3 Cluster 3 – Singular Convection Regime	127
5.4.4 Cluster 4 – Organized Convection Regime	129
5.5 Discussion	131
Chapter 6: Towards a CloudSat Latent Heating Product Using Cluster Analysis	
6.1 Introduction	136
6.2 Identifying Latent Heating Properties through Cluster Analysis	136
6.3 Discussion	144

Chapter 7: Conclusion and Future Research

7.1 Review of Research	145
7.2 Future Work	147
7.3 Conclusions	148

Bibliography	151
---------------------	------------

List of Figures

1.1	Schematic describing the interaction of the ITCZ with trade cumulus	3
1.2	Diagram from Johnson et al. (1999) of trimodal characteristics	4
1.3	CTP- τ histograms of centroids from Jakob and Tselioudis (2003)	5
2.1	CTP- τ histograms of centroids from Jakob et al. (2005)	15
2.2	Time series of the occurrence of TWP regimes from Jakob et al. (2005)	18
2.3	ISCCP histogram from MODIS observations of cloud properties	24
2.4a	Comparison of the choice of k for MODIS cluster analyses	27
2.4b	Comparison of the choice of k for MODIS cluster analyses	28
2.4c	Comparison of the choice of k for MODIS cluster analyses	29
2.4d	Comparison of the choice of k for MODIS cluster analyses	30
2.5	Clusters for choice of $k=4$ for MODIS analyses	31
2.6	CloudSat profile of multiple cloud layers	37
3.1	21 June 2002 observation of MODIS and AMSR/E cloud properties	41
3.2	Histograms of precipitation amount sorted by cluster	46
3.3	3D histogram of MODIS and AMSR/E observations from 21 June case	48
3.4a	Cluster 1 for $k=3$ case	56
3.4b	Cluster 2 for $k=3$ case	57
3.4c	Cluster 3 for $k=3$ case	58
3.5a	Clusters 1 and 2 for $k=4$ case	61
3.5b	Clusters 3 and 4 for $k=4$ case	61

3.6a	Clusters 1 and 2 for $k=5$	66
3.6b	Clusters 3 and 4 for $k=5$	67
3.6c	Cluster 5 for $k=5$	68
3.7a	Clusters 1 and 2 for $k=6$	72
3.7b	Clusters 3 and 4 for $k=6$	73
3.7c	Clusters 5 and 6 for $k=6$	74
4.1	3D histogram of trade cumulus regime	81
4.2	As in 4.1 but for enhanced cumulus regime	84
4.3	As in 4.2 but for isolated convection regime	86
4.4	As in 4.3 but for organized deep convective regime	88
4.5	Evolution with time of the 3D cloud regimes	91
4.6	As in 4.5, but with overlay of sawtooth overlay	91
4.7	Cluster trend versus CERES OLR	97
4.8	Cluster trend versus CERES downwelling SW flux at the surface	97
4.9	Cluster trend versus CERES downwelling LW flux at the surface	97
4.10	Cluster trend versus NCEP 850mb zonal wind	101
4.11	Cluster trend versus NCEP SST	101
4.12	Cluster trend versus NCEP TCWV	103
4.13	Regime-averaged vertical profiles of specific humidity from NCEP	103
4.14	Cluster trend versus NCEP lifted index	105
4.15	Cluster trend versus NCEP vertical velocity	107
5.1	RAMS simulation trade cumulus regime	124
5.2	As in 5.1, but for enhanced cumulus regime	126

5.3	As in 5.2, but for isolated convection regime	130
5.4	As in 5.3, but for organized convection regime	121
6.1	Simulated radar reflectivity of RAMS output using QuickBeam	138
6.2	Reflectivity and Latent Heat case study: isolated convection	140
6.3	Regime-average profiles of latent heating in 3D RAMS simulation	141

List of Tables

2.1	Comparison of JT03 cluster values versus MODIS cluster values	34
2.2	Comparison between MODIS/NCEP results and Jakob et al. (2005)	35
3.1	Sorting bin values for MODIS-AMSR/E 3D histogram	46
3.2	Intercluster distances for $k=4$	63
3.3	Intra-cluster distances for $k=4$	63
3.4	Coefficients of pattern correlation between clusters in the $k=4$ case	63
3.5	Average coefficients of pattern correlation between clusters for the $k=5$ case	70
3.6	Average intercluster distances within clusters for the $k=5$ case	70
4.1	CERES observed radiative properties by regime type	98
5.1	Intercluster distances for $k=4$	121
5.2	Intra-cluster distances for $k=4$	121
5.3	Coefficients of pattern correlation between clusters in the $k=4$ case	123
5.4	Summary table comparing observed and modeled regimes	132

Chapter 1

Introduction and Motivation

1.1 Background

The feedbacks associated with the tropical hydrologic cycle are of paramount importance to understanding global climate change. Our relative inability to describe completely the physical mechanisms behind these feedbacks greatly inhibits our ability to analyze (and by extension, predict) important climate signals. It is of some interest, therefore, to develop a hypothesis that explains the physical mechanisms that govern the hydrologic cycle as well as the associated feedbacks.

Development of a hypothesis that adequately explains the hydrologic cycle and its evolution is greatly hampered by the lack of a complete dataset that describes the hydrologic cycle as an organic whole. The history of said hypotheses is therefore based on arguments that rely on the impact of the hydrologic cycle on more observable quantities. As an example, Riehl (1954) computed the yearly mean precipitation of the Earth based on the required latent heat release to balance the effect of radiative cooling of the atmosphere. Later studies tend to confirm, at least to first order, this indirect analysis, although the analysis cannot provide physical mechanisms merely from the inferences drawn therein. Other studies used self-regulating theories of the tropical hydrologic cycle. Manabe and Wetherald (1967) applied their theory of radiative-convective equilibrium, based on hypothetical distributions of relative humidity and using a mechanism of convective adjustment, to compute the change in global temperature due to a doubling of CO₂.

One way of observing the tropical hydrological cycle is in the intraseasonal variation in clouds and weather. Theories regarding the structure of the intertropical convergence zone (ITCZ) and the trade-wind inversions were related to hypotheses of water vapor and latent heat flux transport out of the tropics. A diagram of this system, taken from Simpson (1992) is presented as Figure 1.1. Johnson et al. (1999) modified this structure to include the effects of the freezing-layer inversion (as well as the trade inversion) to describe a tri-modal model of tropical convection, consisting of trade cumulus, cumulus congestus, and cumulonimbus cloud types, as shown in Figure 1.2. The evolution with time of the tropical hydrologic cycle has also been studied. The canonical discovery of Madden and Julian (1971) and summarized by Madden and Julian (1994) of a 40-60 day oscillation (hereafter referred to as the Madden-Julian Oscillation, or MJO) in convection, concurrent with westerly wind bursts, is perhaps one of the best known tropical intraseasonal variations. Nakazawa (1988) analyzed the structure of convection in the framework of the MJO and defined a hierarchical structure of tropical convection in relation to forcing from the MJO. The largest convective cloud clusters, referred to by Nakazawa as ‘super-clusters’ (and also referred to as mesoscale convective systems (MCSs)) are of particular interest, especially with regard to their propagating behavior.

These theories of tropical cloud structure tend to focus on two or three predominant cloud populations – cumuliiform clouds (fair weather and congestus) as well as cumulonimbus. More contemporary studies (e.g. Jakob and Tselioudis (2003), Rossow et al. (2005) and Jakob et al. (2005)) suggest the presence of additional cloud populations. Using the ISCCP dataset and cloud classification scheme, described by

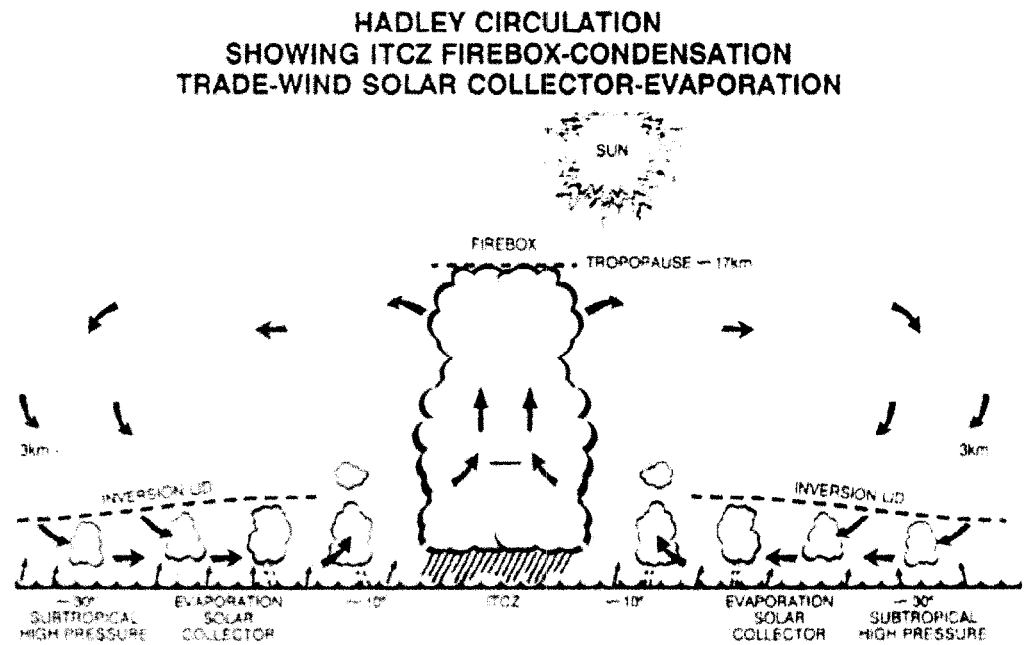


FIG. 1. Schematic north-south slice through the tropical atmosphere showing the towering rainclouds in the ITCZ "firebox" (not to scale). Arrows show the meridional Hadley circulation, whose upper branch transports some of the released heat energy poleward in both hemispheres (Simpson 1992).

Figure 1.1. Schematic describing the interaction of the ITCZ with trade cumulus. Figure taken from Simpson (1992).

Rossow and Schiffer (1991), Jakob and Tselioudis (2003, hereafter JT03) used histograms of ISCCP cloud type using three-hourly observations of cloud top pressure and cloud optical depth in the framework of a nearest-neighbor (or *k-means*) clustering algorithm (Anderberg, 1973). The results of their efforts were distinguishable and unique cloud clusters composed of cloud populations with similar cloud properties. The resulting clusters are considered to be objectively-determined cloud regimes. JT03 find four dominant regimes (Figure 1.3): a convectively suppressed cumulus regime, a convectively suppressed high thin cloud regime, and two convectively active regimes with differing strengths of convection. Another key finding of JT03 is that the

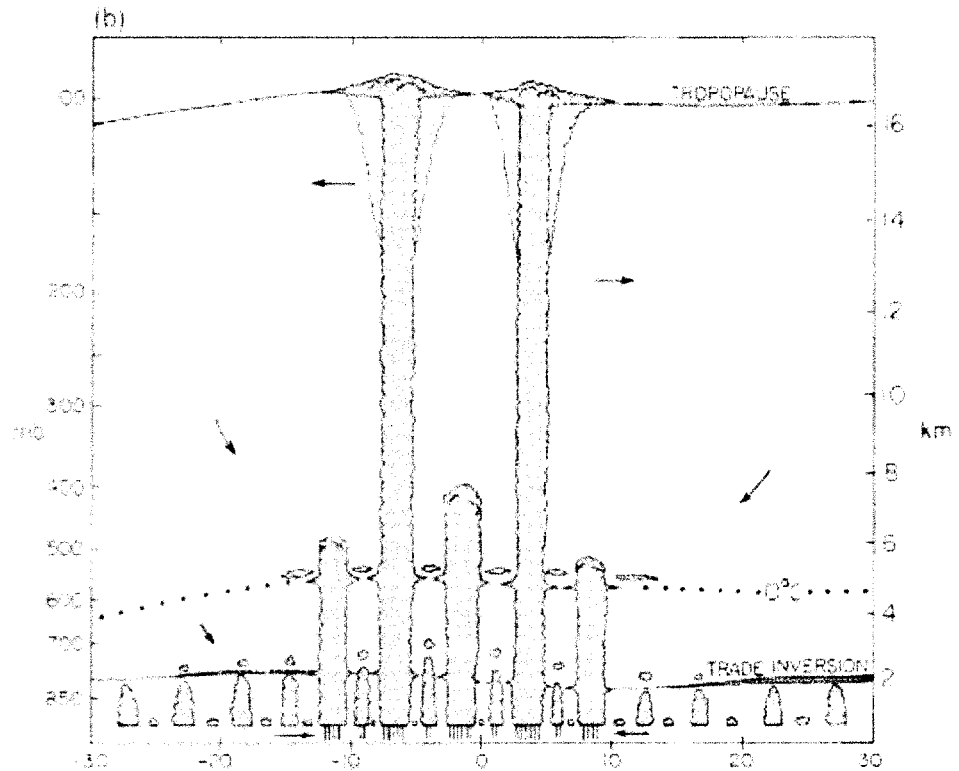


FIG. 13. (Continued) cloud types are indicated: shallow cumulus, cumulus congestus, and cumulonimbus. Within the shallow cumulus classification, there are two subdivisions: forced and active cumulus. Three stable layers are indicated: the trade inversion, the 0°C layer, and the tropopause. Shelf clouds and cloud debris near the trade and 0°C stable layers represent detrainment there. Cirrus anvils occur near the tropopause. Considerable overshooting of the trade and 0°C stable layers occurs in the equatorial trough zone. Arrows indicate meridional circulation. Although double ITCZ is indicated, representing IOP mean, this structure is transient over the warm pool and a single ITCZ often exists.

Figure 1.2. Diagram from Johnson et al. (1999) of trimodal characteristics of tropical convection

occurrence of deep convection in the tropical West Pacific ocean is ‘relatively small and rare compared to the cirrus outflow they produce.’

Further implications on the nature of cloud regimes in the TWP are contained in a hypothesis regarding a self-regulating mechanism of the tropical hydrologic cycle. As described in Stephens et al. (2003), mechanisms associated with the MJO are found to correspond with cyclic processes between sea-surface temperature (SST), upper- and

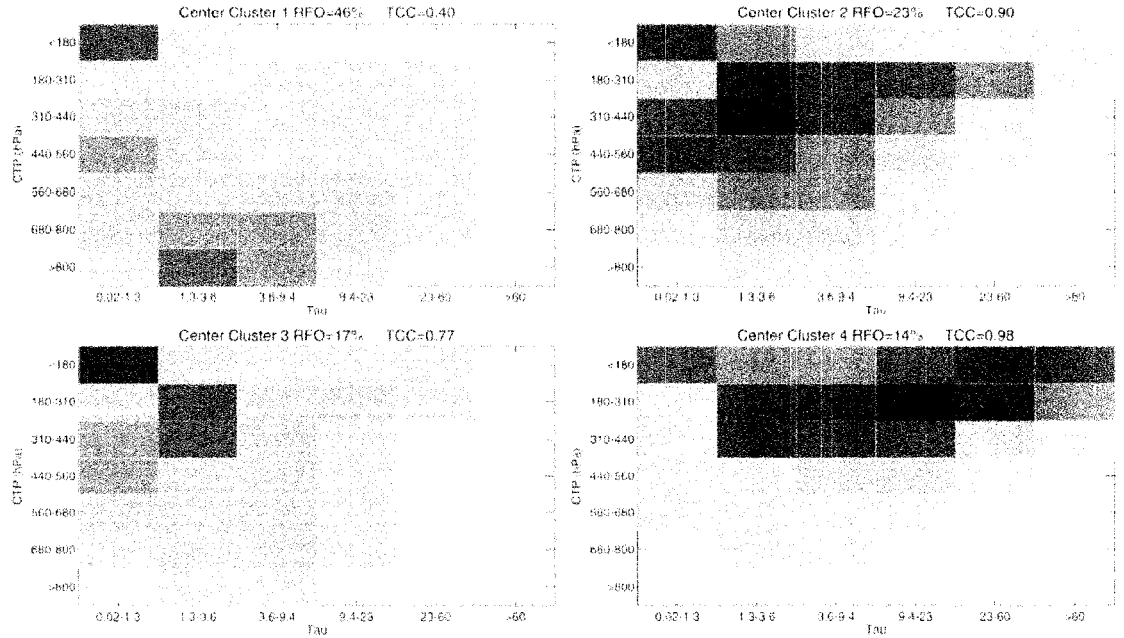


Figure 1.3. CTP- τ histograms of the centroids of a four-cluster analysis using all ISCCP histograms from 1999 in the TWP, from Jakob and Tselioudis (2003).

lower-tropospheric heating, and the adjustment of the tropospheric stability profile caused by the former. Essentially, the hypothesis posits a three-phase cycle governing the magnitude of convection, concurrent with an externally-driven convective triggering, mechanism driven by tropical wave dynamics. Other theories investigate the genesis of this convective triggering, and the various feedbacks between wave dynamics and convective heating profiles.

Other theories tying feedbacks between dynamic and thermodynamic properties of the tropical West Pacific with tropical convection focus on the presence of tropospheric humidity as a controlling feature of the tropical hydrologic cycle. A classic example of a simple feedback mechanism between tropospheric moisture processes and

convection is found in the convective parameterization described by Arawaka and Schubert (1974.) Conditioning of the tropical environment by convection as seen through observations is also a ubiquitous topic (Yanai et al. (1973), Dudhia and Moncrieff (1987)). These theories tend to describe a desiccating influence of convection on the descending regions surrounding the convective cores – studies that include the radiative effects of modification of the tropospheric humidity profiles (Grabowski and Moncrieff (2004)) are used to demonstrate capabilities in improving such tropical features as the MJO.

There exists a common thread among all of the theories described here – they all tie the presence and prevalence of tropical cloudiness to a series of complex and interconnected feedbacks between a series of important atmospheric variables. Tropical clouds are seen essentially as results of the interplay between these variables, while at the same time complicating the analyses due to the difficulty in retrieving parameters in cloudy or semi-cloudy atmospheres, as well as the complex and poorly understood feedbacks due to the physical presence of tropical clouds themselves.

1.2 Motivation

Our proposal is to combine objective, observation-based categorization of the atmosphere with the mechanistic and dynamical analysis of contemporary theories on the tropical hydrologic cycle to tease out as much information as possible about each of the phases of the hydrologic cycle. This accomplishes two tasks – it gives an objective definition of each phase based on a cogent physical convention, and it greatly facilitates intercomparison between each phase of the hydrologic cycle and the theorized phases that result from theories regarding their governance. Observing the change with time as each

phase of the hydrologic cycle changes to a new phase could allow us to gain insight into possible feedback mechanisms between the phases, and perhaps even work out (to first order, at least) the magnitude of change between the dominant variables responsible for the initiation of change in the hydrologic cycle. What remains for us is to construct the appropriate ‘cogent physical convention’ with which to apply our analysis.

Current approaches to the topic at hand typically approach the topic in the framework of a complex interplay of discrete components whose forcings and feedbacks are indistinguishably wrapped up in the observations. Complicating the issue is the reality that several of the necessary components are difficult, or perhaps impossible to retrieve over the sparsely populated regions of the tropics typically being studied, especially during periods of cloud cover, when the ability of satellite-based instruments to reliably retrieve surface parameters is reduced or perhaps even eliminated. Given the ubiquity of cloud cover in the tropical atmosphere, this is no small issue. Consequently, progress in research on the tropical hydrologic cycle is hampered by the difficulties encountered in confirming or rejecting hypotheses due to uncertainty (or complete lack) of the observations.

Perhaps a different way to approach the topic would be to observe primarily those parameters of the tropical atmosphere that are easiest (or at least, most prominent) to infer (say, cloud properties), and then examine the remaining parameters in a framework of the *easy* parameters, which we have more confidence in. Clouds, which are a direct result of the complicated and poorly-understood thermodynamic mechanisms within the tropical system, are relatively easy to observe – perhaps creating a framework of understanding the tropical system as a function of the clouds it produces would be

fruitful. k -means cluster analysis, predicated on cloud properties may well provide such a framework, assuming we apply it in a manner that accounts for the uncertainties in our retrieval of cloud properties. The framework of cluster analysis based on geostationary ISCCP histograms of cloud top pressure and cloud optical depth is a good start – the ISCCP histogram structure is ideally suited to minimize the effects of retrieval error, due to the relatively large bin sizes of the classification scheme. And the retrievals of cloud-top pressure and cloud optical depth are generally well-behaved – one thing to investigate is whether or not these two parameters alone are sufficient to categorize tropical convection. To first-order, they certainly are, but it may well be that the inclusion of additional cloud properties could improve the categorization necessary to examine other, less-reliable parameters in context of cloud evolution.

One final motivation worth considering is the ability of any method that organizes tropical convection by cloud properties to be combined with other cloud observations of the same region. By necessity, any method we propose will certainly utilize some form of satellite-based retrieval – it would be ideal if we could choose a satellite observing platform that facilitated further research using additional instruments.

1.3. Outline of Research

The upshot of the previous several pages is that we seek an intelligently-designed way to organize the tropical West Pacific in terms of the dominant cloud regimes contained therein, and then interpret observations of atmospheric parameters sorted by cloud regime to better understand the changes between the regimes. With the assumption that the dominant cloud regimes computed by our analysis will correspond with the important phases of the tropical hydrologic cycle (and we'll have to test to make sure that

this is so), we will have then gained understanding on the changing nature of the tropical hydrologic cycle. In keeping with our final motivation, we would like to use cloud observations from the EOS-PM (Aqua) satellite, which is the ‘lead’ satellite of the so-called ‘A-Train’ (Stephens et al. (2004)), which would provide us with a surfeit of relevant atmospheric observations with which to compare. We would combine the satellite-based information with reanalysis products to further enhance our understanding.

Where we go from here is an essentially linear progression – we need to ensure that we can first reproduce the results of JT03 with observations from a new satellite platform, namely Aqua (Chapter 2). Having done so, we need to examine those results, and suggest potential changes, and investigate the utility of additional information. We comprehensively develop a new framework for k -means analysis of TWP cloud properties, and then describe the results of performing that analysis (Chapter 3). Having done this preparatory work, we begin our analysis of the TWP using our regime information, establish the connection between our regimes and the tropical hydrologic cycle, and further develop our understanding of these regimes using external observations, both satellite-based and from reanalysis (Chapter 4). We explore the great utility of cluster analysis in the framework of large-scale cloud resolving models, and investigate some key differences between observations and models (Chapter 5). We perform a feasibility study on the utility of cluster analysis in the development of future retrieval algorithms (Chapter 6). Finally, we summarize the research, discuss future applications, and list the key conclusions of this research. (Chapter 7.)

Chapter 2

Cloud Clustering as a Diagnostic Tool

2.1 The k-means algorithm

Clustering algorithms are used in a wide variety of analyses requiring an intelligent means of sorting between different states of a system. Clustering techniques have also the additional benefit of being able to quantify the *amount* of difference between different unique states within a multi-dimensional system. Other applications of cluster techniques aside from climate research include economics (to quantify and predict trends), and are also incorporated in facial recognition software used by surveillance camera security systems commonly found at airports.

As described by Anderberg (1973), the k-means clustering algorithm is an iterative process that assigns individual observations of a dataset to one of any number of cluster seeds (called *centroids*.) At the beginning of the algorithm, k observations are chosen at random from a dataset containing N observations. These observations become the first class of centroids that are used to sort the remaining $N-k$ observations. The algorithm then proceeds individually through the remaining observations, and computes the Eulerian distance of the individual observation from each of the k centroids. The individual observation is then assigned to the nearest (and therefore, the most similar) centroid (as determined from the previous step) and the centroid is recomputed using a weighted average of the old centroid value combined with the new observation. The algorithm then moves to the next observation, again comparing the individual observation to the new centroids, and repeating until additional iterations do not produce a significant change in the centroids, thus terminating the iteration process. After each individual

observation has been sorted in this manner, the new centroids represent, in sense, an aggregate of the mean properties of several similar observations in the dataset (in fact, if k is chosen to be 1, the process is identical to computing the mean of every observation in the dataset.) We define these resultant centroids as the clusters of archetypal states existing within the system being analyzed. Using facial recognition as an example, this technique is analogous to taking a group of facial photographs, sorting them into subgroups based on roughly similar appearances (choosing $k=2$ in this example might well yield two clusters comprised of male faces and female faces, for example), and then computing an average male and female face that is representative of the bulk characteristics of each subgroup.

The process of choosing a suitable value for k is the only non-objective component of the algorithm, although it is possible to use objective criteria to finalize an appropriate value. To select an appropriate value for k , the clustering algorithm is performed several times with increasing values of k , and then analyzing the resulting clusters against several ‘rules’. Rossow, et al. (2005) describes a set of four such rules that, when met by the clusters, establish a ‘best’ value for k . These rules are:

- 1.) The resulting clusters must have high pattern correlation with themselves if the initial k observations are changed (e.g. picking different ‘seed’ clusters does not change the outcome,)
- 2.) The resulting clusters should have low pattern correlation values with other clusters. This is a particularly useful criteria, as choosing a value of k that is too

large often results in several clusters being near-copies of each other, with only minute differences.

3.) The distribution of cluster frequency over space and time should also have low correlations with each other, and

4.) The distance between the resulting clusters should be larger than the intra-cluster dispersion, signifying confidence that observations ‘belong’ to a particular cluster.

The criteria set by Rossow, et al. (2005) create an extremely useful objective framework for determining the ‘best’ value for k , although the criteria themselves are, of course, subjectively selected. The algorithm itself requires no predetermined set of criteria to select a value of k , and other possible mechanisms to objectively determine the ‘best’ value of k may yet be described. The original analysis of JT03 do not specify a particular algorithm for determining k – they instead choose $k=4$ based on the ‘visual significance’ of choosing four clusters over a larger number. (Rossow et al. (2005) using a much different set of observations found a ‘best’ value for k of six.)

It is also possible to *begin* with archetypal clusters at the start of the process, computed from some prior analysis, and then compare the individual observations to the archetypal clusters to compute the difference between each individual observation and any of the clusters, *without* modifying the clusters themselves. Returning again to our example of facial recognition, this is analogous to processing through a stack of

photographs and comparing each photograph to photographs of known criminals, and triggering a possible match when the similarity (as expressed through an Eulerian distance) between an individual photograph and a photograph of a criminal exceeds some set value. Although this method is not technically a clustering algorithm *per se*, it uses several of the same techniques. As applied to cloud observations, this research will use both methods – the former to determine dominant cloud regimes of the tropical West Pacific, and the latter to assess the evolution in time of cloud properties with respect to the dominant regimes.

2.2 The Method of Jakob and Tselioudis

In order to apply the k-means algorithm to a dataset of cloud observations, one must first devise a framework in which to describe the cloud properties in a manner that is easily understood by the algorithm. The algorithm as described would perform poorly if temporal and spatial properties of tropical cloud systems were included in the information used to ‘cluster’ a dataset, due to the great variability in these properties. What is desired is a framework that contains only the essential *physical* properties of cloud systems, separated from their temporal and spatial characteristics. This necessitates choosing temporal and spatial scales of interest, and then within those scales of interest, choosing a suitable method to analyze the cloud properties in a purely physical manner.

Jakob and Tselioudis (2003, hereafter referred to as JT03) used the k-means algorithm to analyze ISCCP D1 (Tselioudis et al. 2000) cloud observations from a region within the TWP (specifically, 130° to 170° degrees East longitude, 10° South to 10° North latitude) over a period of one year (calendar year 1999.) These observations are three-

hourly histograms of cloud optical depth (τ) and cloud-top pressure (CTP), following the ISCCP cloud classification scheme described by Rossow and Schiffer (1991), and are taken from a combination of polar-orbiting and geostationary satellites. The ISCCP cloud classification scheme described by Rossow and Schiffer (1991) is based on observations of cloud radiative properties, and as such, permits discrimination between cloud systems based entirely on observed physical properties – making them particularly useful in the framework of the k-means algorithm, as these properties occur independently of the cloud system’s physical location or timescale.

Using these histograms as the individual observations, the k-means algorithm as described in section 2.1 is applied. When applied to the ISCCP D1 dataset, JT03 resolved four clusters, each of which JT03 suggested was representative of a distinct cloud regime. These regimes are presented as Figure 2.1.

As is described in JT03, the regimes found are: a shallow cumulus regime (Cluster 1), an isolated cirrus regime (Cluster 3), a convective/thick cirrus regime (Cluster 2), and a deep and “probably organized” convective regime (Cluster 4.) The two convective regimes populate the right-hand column of Figure 2.1, while the so-called ‘suppressed’ regimes populate the left hand column. The shading in Figure 2.1 describes the normalized occurrence of a particular grouping of τ and CTP – a darker shade corresponds to an increased occurrence of that particular CTP- τ pairing. As an example, Cluster 3 in Figure 2.1 demonstrates more frequent occurrences of high, thin cloud pixels that are interpreted as cirrus clouds.

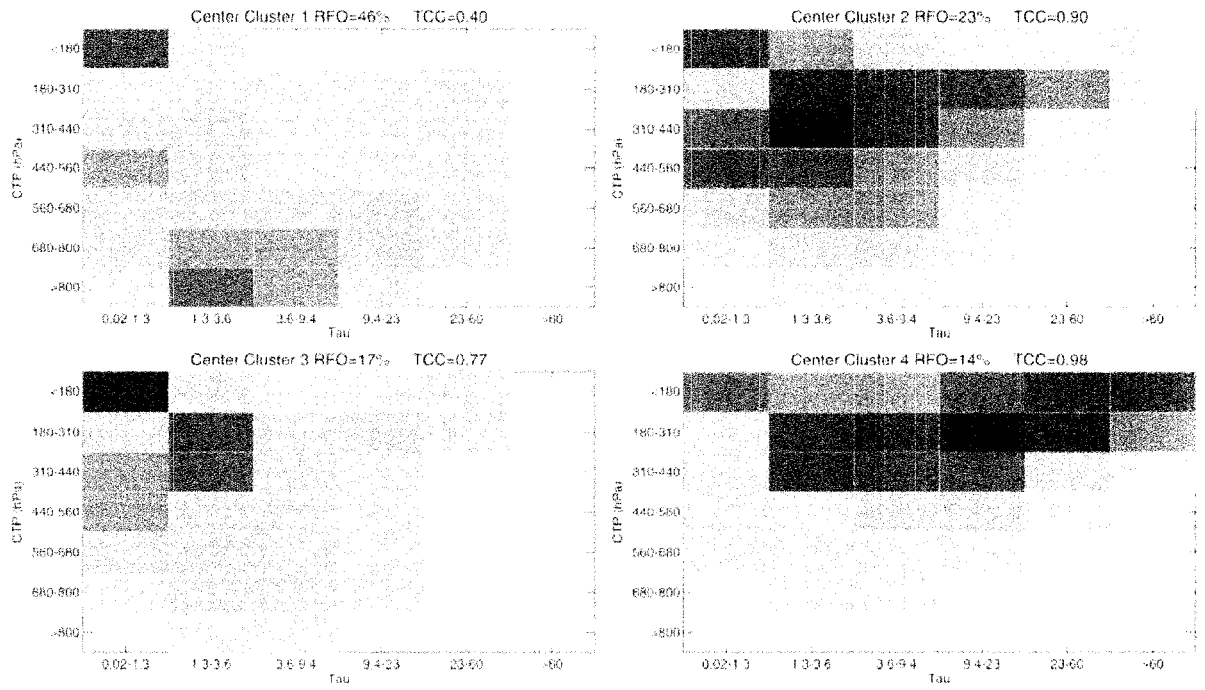


Figure 2.1: CTP- τ histograms of the cloud regimes described by JT03 using ISCCP D1 data of the TWP in 1999. Used with permission from Jakob and Tselioudis (2003)

Each regime is also described by a relative frequency of occurrence (RFO) and a total cloud-cover amount (TCC.) The RFO is the ratio of the number of observations that ‘belong’ to a certain cluster to the total number of observations, expressed as a percentage. As an example, Cluster 1 in Figure 2.1 has an RFO of 46% meaning that 46% of the observations in the dataset identified with the suppressed shallow-cumulus regime. The total cloud cover (TCC) is readily found by summing the individual histograms to get the total occurrence of cloudy pixels in the histogram, again expressed as a percentage.

An interpretation of JT03 is a picture of the atmosphere as a series of repetitions through ‘modes’ in the tropical atmosphere, with each successive mode described by its own peculiar cloud properties. A useful outcome of such an interpretation would be to link each mode (and eventually, the progression through each) with the determining dynamic and thermodynamic mechanisms – such an analysis would have great benefit to the modeling community, for example, in grappling with the complex feedbacks between heat, moisture, and radiation between cloud systems as represented in models. In order to pursue this goal, one must take a close look at the individual properties of each of the modes – this is one of the goals of this research project, as well as other projects. We’ll describe some of the latter first in brief, and then return to the former in *much* more detail.

In an effort to better quantify the radiative and thermodynamic characteristics associated with each regime, Jakob et al. (2005) used the regime analysis of JT03 to compute regime-averaged properties using a variety of observations, including satellite observations of OLR and surface and radiosonde observations taken from the

Atmospheric Radiation Measurement Program (ARM) site located on Manus Island, as well as limited use of reanalysis products to infer other fields that are more difficult to measure directly. The observations listed are sorted and averaged by regime. As in JT03, 3-hourly ISCCP histograms of cloud-top pressure and cloud optical depth (the ISCCP D1 data product) for the region 10°N-10°S, 130°-170°E are used for the cluster analysis - for this study, the observations are extended in time to include the period 1999-2000 (JT03 used only data from 1999.) Also as in JT03, the analysis results in four clusters with properties nearly identical to the clusters described in JT03.

One of the two important findings of Jakob et al. (2005) that we'll discuss here is the ability of the clustering technique to describe, at a glance, the presence of a certain cloud regime as a function of time. Figure 2.2, a reproduction of Figure 2 from Jakob et al. (2005), demonstrates this particular feature. The figure is a representation of cluster trends for the month of May 2000 – time runs along the x -axis, while cluster type is represented on the y -axis, in four rows. A shaded line in a row for a certain day indicates the presence of that row's cluster on that day – looking at Figure 2.2, then, we see that the month started with a preponderance of deep convection through the 8th of May, transitioning to a period of cirrus and low cloud through the 16th, and ending with a period of short-timescale variability in cloud type. It is possible (and very useful!) to extend these types of plots over longer time periods to analyze larger-scale trends in time of cloud regimes, and we will return to several similar analyses shortly.

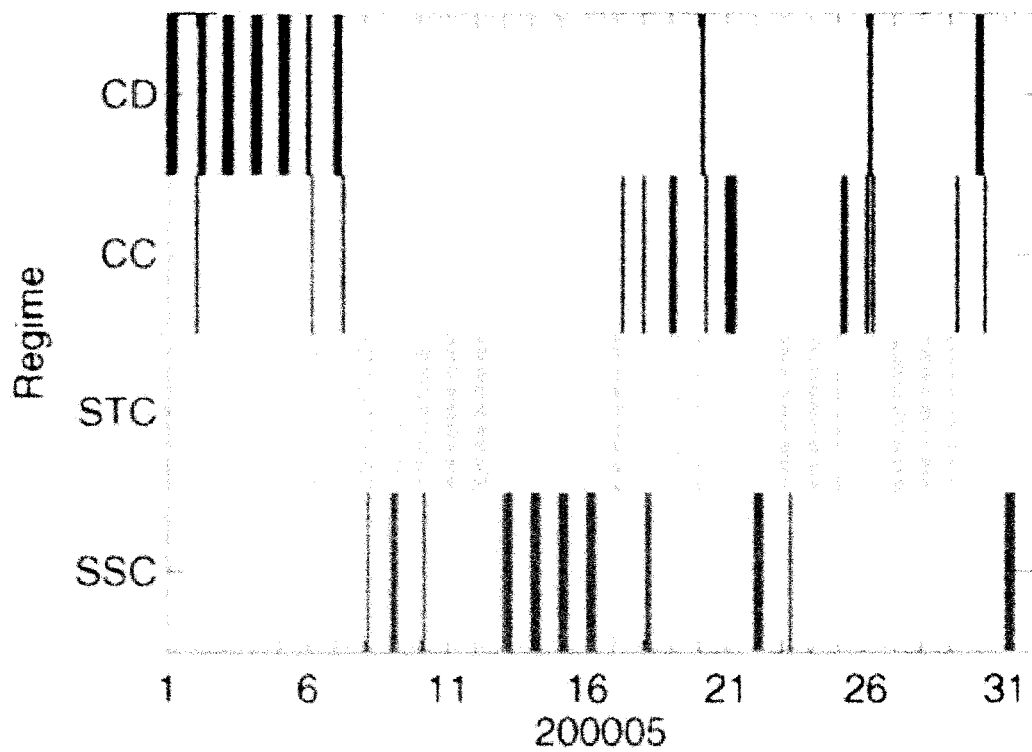


Figure 2.2. Time series of the occurrence of the TWP cloud regimes around Manus Island in May 2000. Reproduced from Figure 2. in Jakob, et al. 2005.

The other important finding (or rather, *findings*) of Jakob et al. (2005) discussed here are the properties of the regimes themselves. An important objective of the clustering technique as applied here is to find unique regimes of the tropical atmosphere. After performing the cluster analysis, it would be ideal if the properties of each unique regime found (especially those properties *not* used in the clustering technique) reinforce the uniqueness of said regime in a coherent manner – for example, the respective OLR for each regime should agree with commonly-accepted values for similar systems. Such reinforcement not only provides additional confidence in the technique, but also allows for quantitative computation of mean physical properties of relevant subsets of the atmospheric system - a particularly useful tool for the researcher.

A detailed description of all the results of Jakob et al. (2005) would unnecessarily burden the familiar reader, but some of the most relevant features of each regime as described are listed here:

- 1.) The shallow-cloud convectively suppressed regime is characterized by low topped clouds with low- to medium optical depths, relatively high OLR (~ 280 W/m²), total column water vapor (TCWV) of ~ 5.0 cm, and approximate mean average vertical velocities (ω) on the order of -0.01 Pa/sec
- 2.) The thin-cirrus convectively suppressed regime is characterized by thin cirrus clouds, lower OLR values (~ 250 W/m²), TCWVs of ~ 5.3 cm, and $\omega \sim -0.02$ Pa/sec.

3.) A convectively active regime consisting of ‘high-top, medium- τ cirrus’ and ‘likely deep convective’ clouds. Some ambiguity about ‘member clouds’ is expressed in this particular regime, which we will address in Chapter 3. The relevant for this regime include a median OLR of 210 W/m^2 , total column water vapor for around 5.5 cm, and $\omega \sim -0.040 \text{ Pa/sec}$

4.) A convectively active regime consisting of high-topped clouds with substantial optical thickness, ascribed to a classic tropical deep convection profile. Median OLRs for this regime are below 170 W/m^2 , albedos are greater than 50%, and total column water vapor for the regime exceed 6 cm, with $\omega \sim -0.045 \text{ Pa/sec}$

The properties for the regimes described are comfortably familiar to researchers, and establish the utility of the clustering algorithm in distinguishing different physical states of the atmosphere based solely on objective analysis of atmospheric observations. In the next section, we will apply the clustering technique to a different set of observations than those used in the JT03 study and compare the results with earlier work.

2.3 k-means Cluster Analysis Using MODIS Level-2 Cloud Product Data

2.3.1 – Background and dataset description

For the most part, the cloud properties used to sort cloud regimes in cluster analysis are retrievable by a number of different spaceborne instruments. For the purposes of this study, we choose observations from the MODIS Level-2 Cloud Product dataset. The *Moderate Resolution Imaging Spectroradiometer* (MODIS) instrument, one of several instruments being flown on the EOS-series of missions, is a high-resolution

36-band radiometer covering the spectral range from the visible through the thermal infrared. Several cloud properties (including cloud optical depth and cloud-top pressure) may be computed using MODIS radiances in the framework of a retrieval algorithm described by Platnick, et al. (2003). Again, of particular interest to this research are the retrievals of cloud-top pressure and cloud optical depth – the ingredients of the ISCCP histograms used in previous cluster analyses. By using MODIS data to create ISCCP-like histograms, it should be possible to perform the same cluster analysis, and compare the results to the earlier ISCCP-based analysis.

There are several relevant differences between the MODIS observations used in this research and the ISCCP D1 dataset used by JT03. The most obvious difference is the orbital characteristics of the instrument itself – the geostationary ISCCP D1 dataset provides constant coverage of the region of interest at around 5km resolution (more or less – for a detailed description on ISCCP pixel size and sampling issues, see Rossow and Schiffer (1991.)), whereas the EOS-series satellites provides a high-resolution (1-5 km) snapshot of the region once per day. The benefit of using higher-resolution (spatially, at least) MODIS data is somewhat offset by the fact that the data will neglect certain features of the system, most notably, the diurnal cycle in convection. It will be shown that the cloud regimes so found (and notably, the bulk properties of these regimes) do not significantly differ from previous cloud regimes using ISCCP data – while the MODIS-based cluster analysis does not explicitly represent the progression in cloud morphology inherent to the diurnal cycle of convection within each regime, it appears that it can identify the same regime-averaged properties found using ISCCP-based analysis.

Other differences between the MODIS and ISCCP observations are the retrievals used to determine the relevant parameters. As an example, cloud-top pressure in ISCCP is determined from inferred height via cloud-top temperature in the IR, whereas in MODIS it is determined through the CO₂ slicing technique (Platnick et al. 2003.) The effects of slightly different retrievals (with concomitant slightly different retrieval error, etc.) is on the clustering algorithm is expected to be small, since the observations themselves are first passed through the framework of the ISCCP cloud classification scheme – the histogram boxes in the classification scheme cover ranges that are exceed the likely differences between retrievals, which may only be a few percent, perhaps. The only notable exception would be for extremely thin clouds (i.e. cirrus) – Jin (1997) compares HIRS and ISCCP observations and notes a significant difference in thin cirrus cloud amount due to retrieval techniques.

The observations used for the purpose of generating ISCCP-style histograms in this research are observations of cloud-top pressure and cloud optical depth taken from the MODIS Level-2 Cloud Product dataset. As discussed in Platnick et al. (2003), the cloud-top pressure retrieval utilizes a CO₂-slicing technique described by Menzel et al. (1983) and Wylie and Menzel (1999) which utilizes several partial-absorption bands in the 15- μ m CO₂ absorption region, as seen by the MODIS instrument. Cloud-top pressure is reported in the Level-2 database in units of millibars, and has a pixel spatial resolution of 5x5 km. Retrievals of cloud optical thickness in the Level-2 Cloud Product are accomplished using a sophisticated algorithm that include decision-making and discrimination by water phase of sensed pixels, then using specific library calculations (based on the outcome of the prior step) from observations of several frequencies in the

visible and NIR, as well as 1.6-, 2.1-, and 3.7 μ m band. A more detailed description may be found in Platnick et al. (2003). Pixel spatial resolution for cloud optical depth in the cloud product data is 1x1 km. For the purposes of generating ISCCP histograms, the 5-km resolution CTP data is oversampled to match the resolution of the 1-km resolution cloud optical depth data; this is preferred to smoothing the optical depth information to match the coarser CTP data.

MODIS data are obtained from the GSFC DAAC service in .hdf format – the relevant parameters are read in and used to populate a two-dimensional array based on the ISCCP cloud classification scheme, and then stored as ISCCP histograms for later use in the cluster algorithm. A sample of MODIS cloud-top pressure, cloud optical depth, and the resulting ISCCP histogram computed from the two are provided as Figure 2.3.

2.3.2 – Spatial and temporal selection, and discussion on choosing k

One other issue that must be considered is the spatial coverage of MODIS granules, as well as some key differences between the histogram construction between the ISCCP-based and MODIS-based cluster analysis. As described in JT03, the area of observation using ISCCP data cover a much greater area than does the typical MODIS granule – a region of the tropical west Pacific from 130°-170° E longitude, and 10°S-10°N latitude. The orbital characteristics of the EOS satellites allow for observations covering the latitude range of JT03, but do not scan far enough longitudinally to cover the entire 40°x20° range visible by ISCCP in a single data granule. Complicating the issue is the construction of the ISCCP histograms themselves – the histograms used in JT03 are computed individually from 2.5°x2.5°-degree subdomains within the larger 40°x20° observation area. Consistent reconstruction of these subdomains in the MODIS dataset is

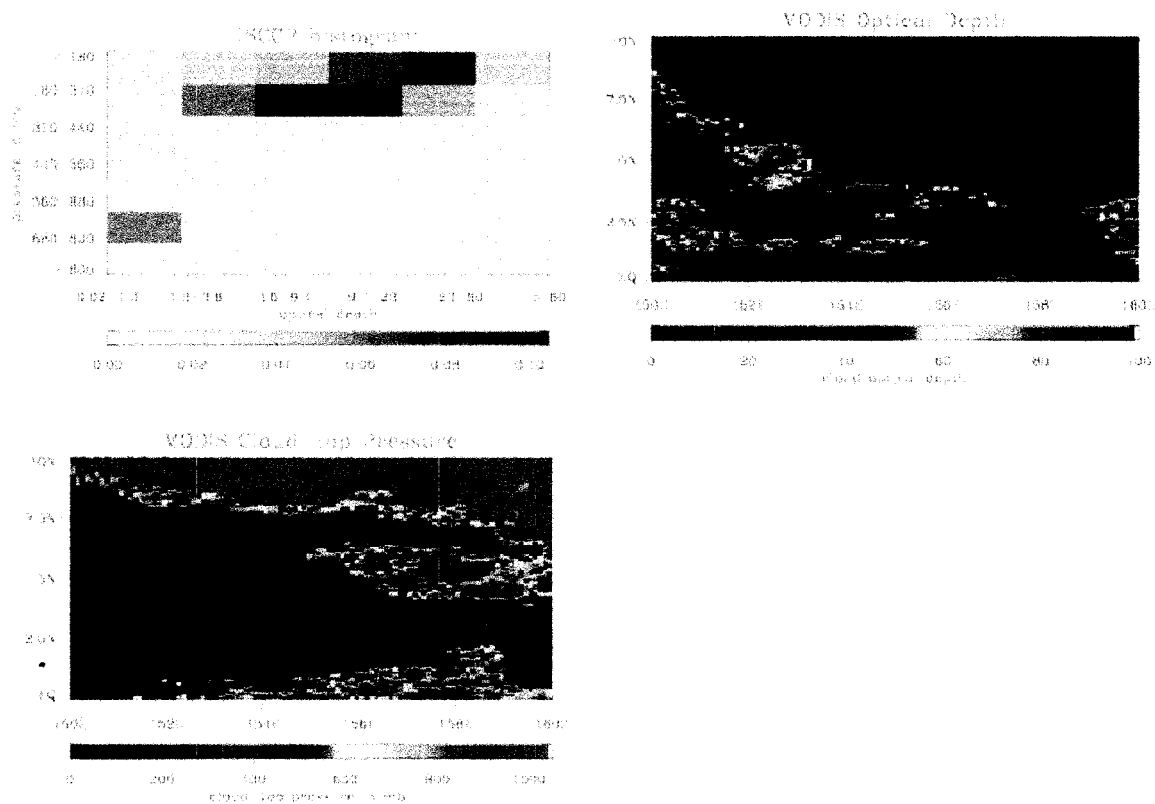


Figure 2.3 – sample ISCCP histogram generated from MODIS Level-2 Cloud Product observations of cloud optical depth and cloud-top pressure. In the cloud-top pressure scene, we see a relatively broad area of high-altitude cloud top pressures in the southwest corner. Cells of high cloud optical depth in the same region suggest the presence of deep convection with attendant cirrus anvil. The generated ISCCP histogram demonstrates the abundance of high-optical depth and high-altitude CTP, as well as a smattering of lower, thinner clouds (as may be seen in the northeastern corner of the scene.)

hampered by the day-to-day orbital variations of the center of the MODIS swath. Given these differences between the satellite platforms, we must make some compromises to accommodate the smaller swath area of a MODIS granule. For this study, we choose a spatial scale that is somewhat smaller than the domain chosen by JT03 in order to ensure complete satellite coverage of the domain. The choice was made to compute single histograms using a region from 150°-160°E longitude, and 0°-10°N latitude, subsetting the region of interest covered by JT03. The histograms used in our MODIS-based analysis are from a region roughly sixteen times larger than those of the ISCCP-based analysis – one histogram from the MODIS analysis would be composed of sixteen adjacent histograms from the ISCCP analysis. For a typical tropical MCS, both the ISCCP- and the MODIS-based histograms are sufficiently large enough to contain all the relevant cloud structures, the largest of which were expected to be the so-called ‘supercluster’ (Mapes and Houze (1993)).

MODIS Level-2 Cloud Product observations from the EOS-PM (Aqua) satellite covering the region of interest described above were collected for the period 01 June 2002 through 31 May 2003, representing a full year of coverage of the region. Days with missing data or incomplete coverage are rejected to ensure quality control on the data.

Since the purpose of this initial cluster analysis using MODIS data is to compare the results to those of JT03, we choose to utilize the same approach as JT03 in determining the value of k to use in the cluster analysis. This criteria, as is discussed in JT03, is one of ‘visual significance’ of the computed clusters, resulting in the selection $k=4$. (More rigorous criteria (along the lines of Rossow et al. 2005) are used in later computations.) Figure 2.4a-d shows the result of running the cluster analysis choosing

values of 3, 4, 5, and 6 for k . The $k=3$ case is shown as Figure 2.4a, continuing through with the $k=6$ shown as Figure 2.4d. Using the ‘visual significance’ criteria, we reject the value $k=6$, as clusters 2 and 6 are quite similar to one another (both clusters roughly contain mid-valued optical depths and high-altitude CTPs.) Similarly, we reject the value $k=3$, as cluster 3 in this run appears to be a conglomerate of clusters 3 and 4 of the $k=4$ run, both of which appear to be significant. Thus, a case for using either $k=4$ or $k=5$ could be made using the ‘visual significance’ criteria; to facilitate comparisons between these results and those of JT03, we choose the value $k=4$ for this section.

It should be noted that the rejection of $k=6$ based on visual criteria alone is perhaps unnecessarily subjective – using the more rigorous criteria of Rossow et al. (2005) results in an optimal value of $k=6$, which is also the value used in recent research by Prof. Christian Jakob (personal communication.) The choice to use $k=4$ over $k=5$ for this section is essentially a matter of convenience, especially considering the temptation to explore the difference between the two convective regimes represented by clusters 2 and 5 in the $k=5$ run (reference Figure 2.4.) In Chapter 3, we will go into more depth regarding possible physical explanations between these weak-convective regimes.

2.3.3 – Cluster Results – MODIS-determined regimes

The k-means clustering algorithm was provided in the form of a FORTRAN code, courtesy of Prof. Christian Jakob. The code was ported to IDL and modified to read MODIS data from EOS-HDF files. After running the analysis to choose k described previously, the code was set to use $k=4$ and then re-run to compute the regimes, along with their respective frequencies and cloud covers. The cluster results are shown as Figure 2.5.

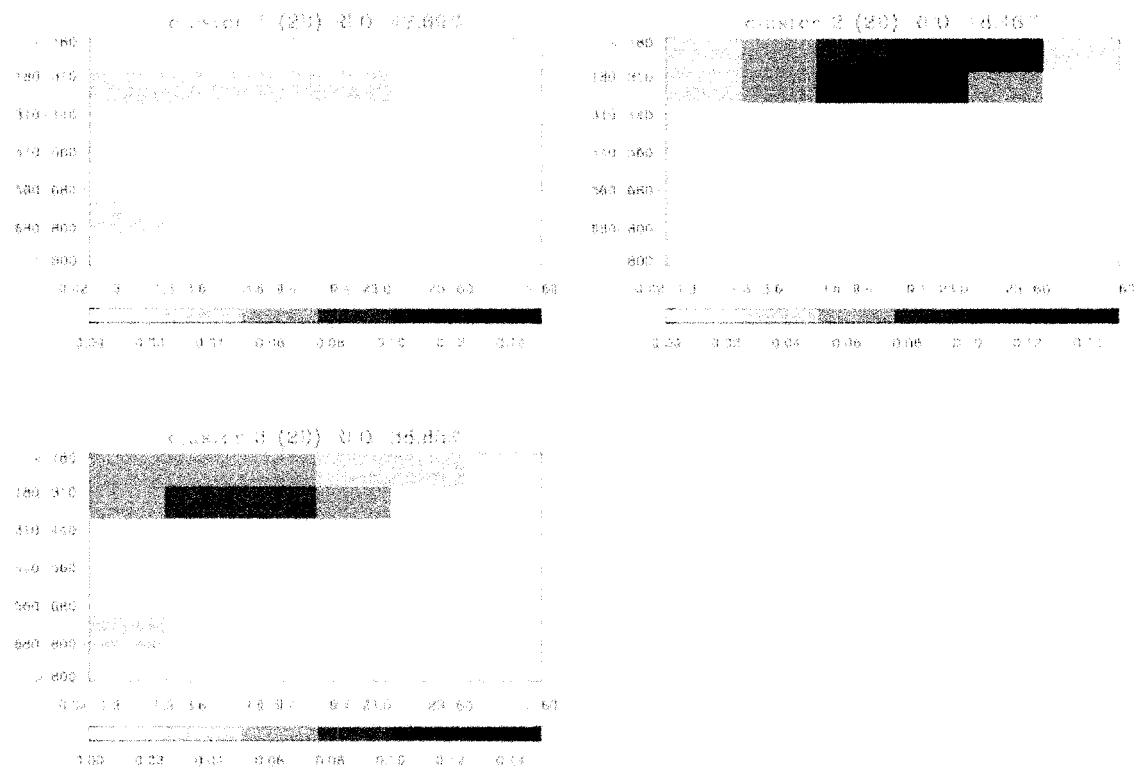


Figure 2.4a – part one of four comparisons of choices for the value of k for the MODIS cluster analyses. This figure represents the choice of $k=3$. The relative frequency of occurrence (RFO) of each centroid is listed over each centroid.

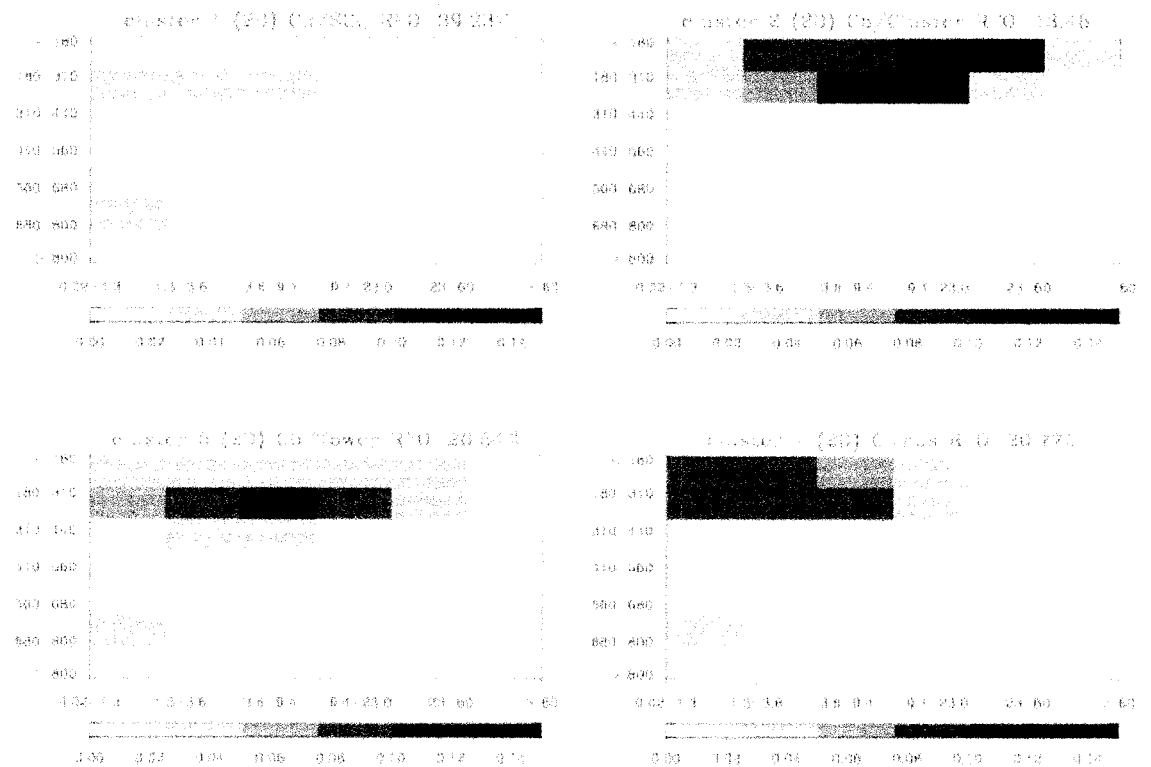


Figure 2.4b – part two of four comparisons of choices for the value of k for the MODIS cluster analyses. This figure represents the choice of $k=3$. The relative frequency of occurrence (RFO) of each centroid is listed over each centroid.

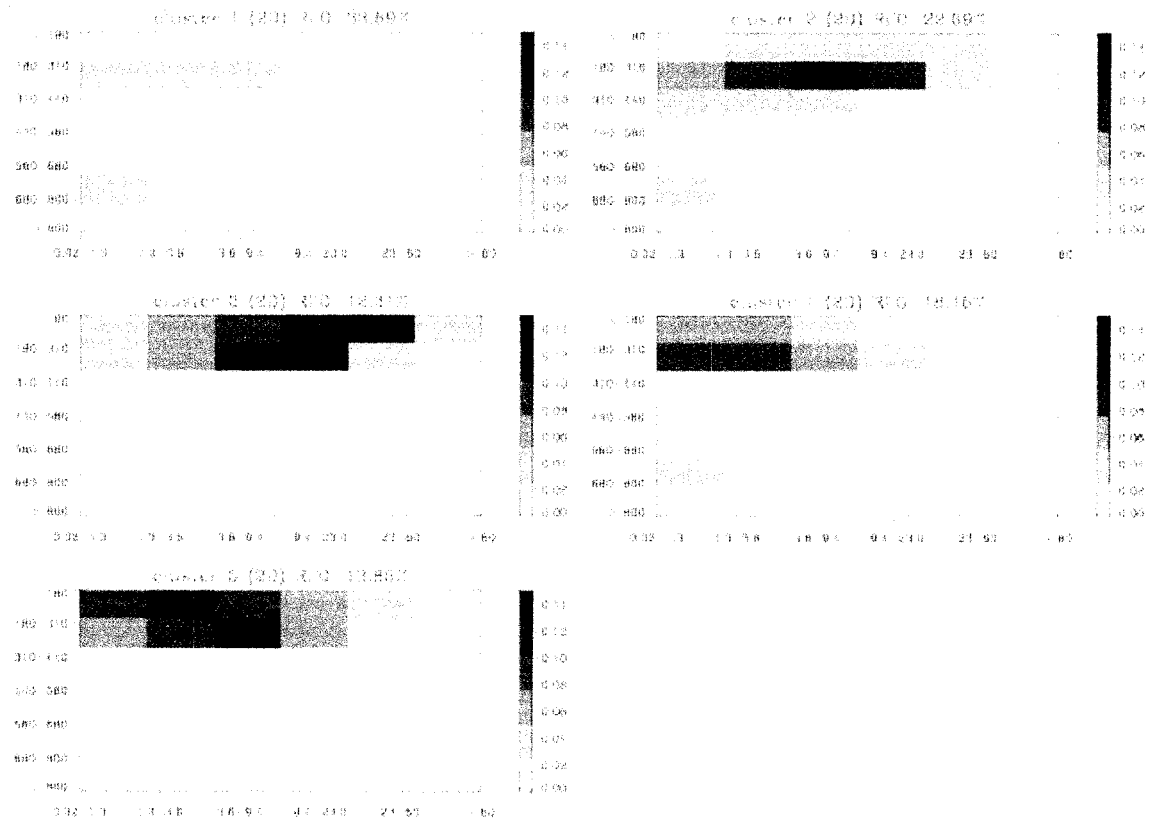


Figure 2.4c – part three of four comparisons of choices for the value of k for the MODIS cluster analyses. This figure represents the choice of $k=5$. The relative frequency of occurrence (RFO) of each centroid is listed over each centroid.

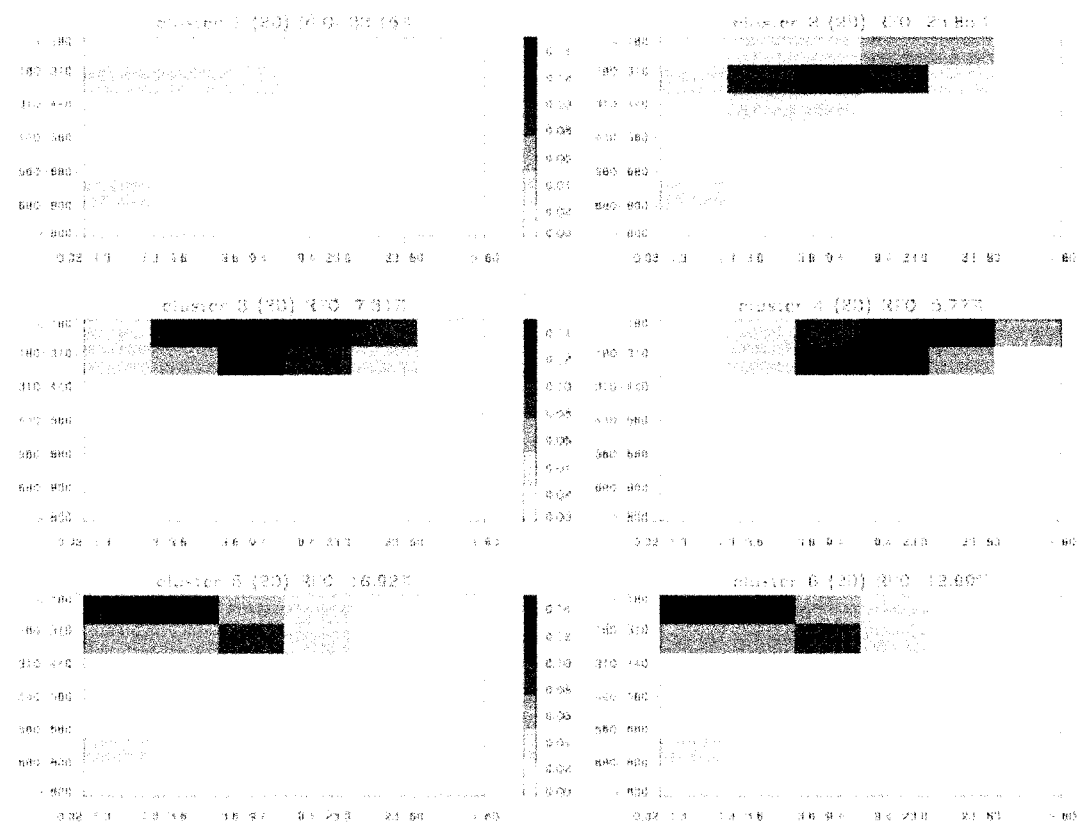


Figure 2.4d – part four of four comparisons of choices for the value of k for the MODIS cluster analyses. This figure represents the choice of $k=6$. The relative frequency of occurrence (RFO) of each centroid is listed over each centroid.

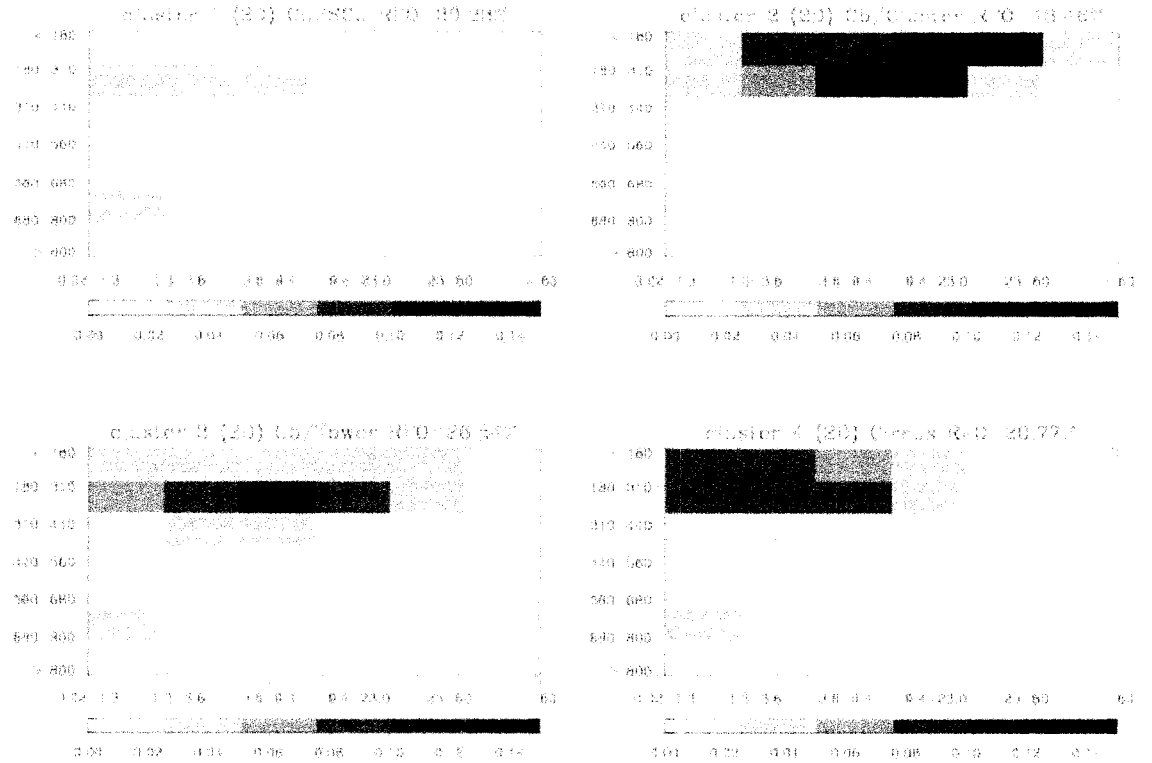


Figure 2.5 – clusters returned from the analysis for choice of $k=4$. Compare to the results of Figure 2.1

Examining Figure 2.5 demonstrate some familiar results – we have two clusters that seem to represent isolated cirrus and shallow cumulus (clusters 4 and 1, respectively) and two convective regimes, one of deep convection (cluster 2) and one of convection/thick cirrus (cluster 3.) The computed frequencies of occurrence (RFOs) of each regime, along with the computed total cloud cover (TCC) for each regime are presented, along with the corresponding values from JT03, in Table 2.1. Perhaps the most remarkable result of the MODIS analysis is that comparable regimes, determined independently, via differing observations systems *and for different years*, have roughly the same total cloud cover amount, as defined by fractional area of cloud cover in the respective regions of interest. The frequency of occurrence vary somewhat more – this is to be expected, since there would be little reason to expect that the weather for two separate years in a region would occur with the same frequency, although the variability is within several percent for each regime. The largest discrepancy between the two comparisons is the cloud cover by isolated cirrus – as mentioned earlier, this was not unexpected, given the difficulties in detecting thin cirrus with different retrieval techniques.

2.3.4 – Bulk Properties of Regimes, and Comparison to Jakob, et al. (2005)

We have seen good agreement in comparative cluster structure between cluster analysis using MODIS observations compared to cluster analysis using ISCCP data, despite the differences in histogram scale sizes between the two analyses. Given that the histogram scaling for both analyses are of sufficient size to encompass the largest cloud features, this is not necessarily a surprise – essentially, both analyses agree that the fundamental characteristics of clouds in the region of interest are similar independent of

observing platform, provided that the histograms used to define the regimes are chosen with the appropriate amount of care. What remains to be tested is whether regime-averaged properties over the region of interest and sorted by MODIS-based regimes are comparable to the regime-averaged properties using ISCCP-based regimes, such as those described by Jakob et al. (2005.)

In order to accomplish this, we use NCEP reanalysis (Kalnay et al. (1996)) data products as a convenient method to look at several relevant parameters for comparison. It should be stressed that we are *not* attempting to match the results of Jakob et al. (2005) exactly, as the observations used to create the clusters are not only from different observation platforms, but were also made some years apart. We instead rely on statistical assessments of mean properties of each regime (as did Jakob et al. (2005)); as such, it is expected that the time-averaged mean properties derived from point sources should not be more or less useful than the time-averaged mean properties derived from reanalysis products – especially considering the great amount of effort put in to engineering the reanalysis to represent ‘reality’, at least in terms of reproducing observations. The general physical properties of each regime, when averaged in the manner described, should be relatively insensitive to the choice of measurements, provided the parameters being analyzed in the reanalysis are not entirely dependent on the parameterizations inherent to the reanalysis model wherever possible.

Three variables, representing in an extremely gross sense the radiative, hydrologic, and dynamic properties of each regime, are used to compare the regimes – these are outgoing longwave radiation (OLR), total column water vapor (TCWV), and average vertical velocity for each regime.

Table 2.1 – Comparison of JT03 cluster values versus MODIS cluster values.

	RFO – JT03	RFO – MODIS	TCC – JT03	TCC - MODIS
Shallow Cumulus	46%	39%	0.40	0.41
Isolated Cirrus	23%	27%	0.90	0.81
Convection/Thick Cirrus	14%	13%	0.96	0.96
Deep Convection	17%	21%	0.77	0.75

Although we express some concern over the accuracy of vertical velocity computations from the reanalysis products, general trends should hold (greater vigor of vertical motion during convective regimes, large-scale atmospheric descent in convectively-suppressed regimes, etc.) The general idea is that we can compare the NCEP vertical velocity values from the MODIS-derived cluster analysis with the NCEP vertical velocity values from the ISCCP-derived cluster analysis, and see if *general trends* are in agreement between the two. Similarly, regime-averaged values of OLR and TCWV between the MODIS- and ISCCP-derived regimes should compare favorably if the regimes are indeed alike.

In order to compute the properties of each regime, we first use the MODIS-derived clusters to go back and sort each daily MODIS observation into its proper regime (for example, if the MODIS ISCCP histogram from June 9th matched most closely to the ‘deep convection’ cluster, that day is assigned as a ‘deep convection’ day.) For each day, the three variables being compared, sorted by their proper regime, are then averaged together to get the bulk properties for each regime – this is the essentially the same method used by Jakob et al. (2005.) Table 2.2 compares the results of Jakob et al. (2005).

Table 2.2 – Comparison between MODIS/NCEP results and Jakob et al. (2005)

	OLR – Jakob et al.	OLR – MODIS/NC EP	TCWV – Jakob et al.	TCWV – MODIS/NC EP	ω – Jakob et al.	ω – MODIS/NC EP
Shallow Cu	280 W/m ²	267.9 W/m ²	5.0 cm	5.16 cm	-0.01 Pa/sec	-0.028 Pa/sec
Isolated Ci	250 W/m ²	252.1 W/m ²	5.3 cm	4.97 cm	-0.02 Pa/sec	-0.034 Pa/sec
Convection/ Thick Ci	210 W/m ²	195.5 W/m ²	5.6 cm	5.37 cm	-0.04 Pa/sec	-0.066 Pa/sec
Deep Convection	170 W/m ²	192.6 W/m ²	5.9 cm	5.81 cm	-0.45 Pa/sec	-0.067 Pa/sec

The greatest differences between the two regimes are, as expected, in the computation of vertical velocity. Jakob et al. (2005) actually used *four* reanalysis products in an attempt to evaluate vertical velocity by regime – unfortunately; none of the reanalysis products provided a consistently realistic value for vertical velocity. In fact, only two of the analyses (neither of which were the NCEP reanalysis) described net downward motion during the convectively suppressed regimes. Whether this is a failing of cluster analysis to ‘sort’ the atmosphere into regimes of like vertical motion based on convective activity, or instead a failure of reanalysis products to accurately represent the instantaneous vertical momentum profile of the atmosphere, however, remains to be seen. Regardless, in both the analysis performed by Jakob et al. (2005) and in this analysis, the minimum (albeit still upward) values of vertical velocity occur during the convectively suppressed regimes, while maximum vertical velocities are found during convective regimes.

We see also the general agreement between the MODIS- and ISCCP-derived regimes in the OLR and TCWV fields – values match remarkably well, especially considering that the regimes described were separated by nearly two years. We stress

that we are not ascribing properties to the atmosphere based on the regime analysis (although we may attempt to do so in future chapters) but are rather confirming that it is possible to replicate the results of an ISCCP-derived cluster analysis, both in overall cluster properties as well as mean regime properties based on the clusters themselves, using MODIS data. The great variety of instruments found on the EOS series of satellites, as well as the many new instruments flying in formation (particularly, those satellites in the ‘A-Train’ constellation behind the Aqua satellite) allow for many opportunities to apply new and effectively simultaneous atmospheric observations in concert with cluster observations made using MODIS observations.

2.4 Discussion

There are some limitations to the cluster analysis as described that should be discussed. Perhaps the most salient of these limitations is the satellite retrievals themselves – since both ISCCP and MODIS observations are passive observations, they contain *no* information about the affects of multiple cloud layers. Histograms of cloud type assembled using passive observations of cloud scenes containing multiple cloud layers suffer in that generally, the lower cloud layers are entirely ignored, or could perhaps influence the values of optical depth retrieved, thereby skewing the histogram. Recent satellites utilizing active sensors are making some headway into discovering these ‘hidden clouds’ – Figure 2.6 presents an 8 January, 2008 QuickLook image from the CloudSat mission (described by Stephens et al. 2003) showing a complex of several shallow-convective towers adjacent to a thick layer of upper-level cirrus. Several of the convective towers appear to be developing underneath the cirrus deck, which may be a remnant of earlier convection. The CloudSat instrument, a 94-GHz cloud-profiling radar,

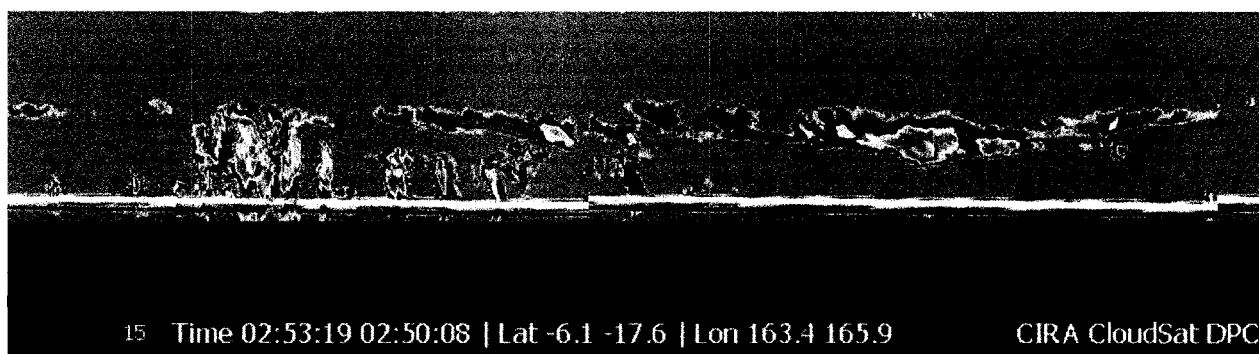


Figure 2.6 – CloudSat profile of multiple cloud layers near the equator in the West Pacific, 8 January 2008. The vertical axis represents height above the ground, and the horizontal axis represents along-track distance. Brighter colors represent greater (uncalibrated) radar reflectivity.

has the ability to penetrate cloud layers and even weak- to moderate-precipitation, and will greatly improve our understanding of the crucially important vertical structure of cloud systems. The utility of directly retrieved vertical cloud structure information used in analyses such as cluster analysis (as opposed to *inferring* vertical properties through variables such as cloud optical depth) is apparent, and there are plans to modify yet again the clustering technique to accommodate these new and valuable observations. We will discuss the use of CloudSat data in cluster analysis techniques in much more detail in forthcoming chapters - for the present, we will move forward with a cluster analysis based on passive-only sensors to facilitate better comparison with past research utilizing cluster analysis.

Another way forward is to rethink the process of how we ‘index’ the atmosphere, at least as it pertains to histograms of cloud type. Using the ISCCP cloud classification scheme makes a good deal of sense, as the sorting scheme used to define the histograms are optimized to produce cloud types of similar radiative impact, but there are other ways to group clouds by their properties, of course. If one imagines the hydrologic cycle in its

simplest form, of the conceptually-simple voyage of water as it evaporates, is transported elsewhere by the ‘dry dynamic’ atmosphere, and is condensed again to fall, only to renew the cycle, then we understand that the presence of water in the atmosphere may be used as a tracer of sorts to mark the goings-on of the atmosphere in the ‘background’. And the properties of clouds are the simplest way to describe the present state of water in the atmosphere. Much of the driving force of the atmosphere is contained in the latent heat release and consumption as water changes phase, and the ultimate driving force of the balance of solar and terrestrial radiation is itself greatly modified by the mere presence of these clouds. Were we to construct a histogram that includes *more* of the relevant features of these clouds, namely their height, relative water amount, and presence of precipitation, we would perhaps make some headway into addressing the issues described above. The addition of precipitation information to the ISCCP histograms currently used in the cluster analyses would further separate the clouds into raining- versus non-raining clouds, perhaps alleviating some of the uncertainty found in different cloud scenes with similar optical depths and cloud-top heights, but with vastly different component clouds.

We propose that such a histogram may be constructed, and that new cloud regimes found using these histograms will add to the already valuable information found by using two-dimensional histograms based on cloud radiative properties. We turn to this next in Chapter 3.

Chapter 3

Towards Development of a Three-Dimensional Cluster Analysis

3.1 Motivation

As was discussed in the introductory chapter, the role of precipitation has a dominating influence on the energy budget of the tropics. The presence (and magnitude) of precipitation in tropical clouds signals a marked change in the overall stability of the tropical atmosphere, replete with feedbacks on several relevant tropical systems, especially on tropical dynamics (through radical changes in the vertical structure of tropospheric heating through latent heat release) as well as the local radiative budget. The formation of precipitation marks an important milestone in convective development – as such, it may well be supposed that the inclusion of observations of precipitation would have an influence on the results of cluster analysis.

How, exactly, to include this precipitation information is the topic of this chapter. We begin with a discussion of candidate datasets for our precipitation observations, and then discuss how to include these observations in an intelligent manner with our previous observations, followed by an exhaustive analysis of the selection of the appropriate value of clusters in our new, 3D framework.

3.2 The AMSR-E Precipitation Product

We seek a precipitation dataset with similar spatial- and temporal coverage and resolution to complement the MODIS observations used for our 2D analysis. Several such options exist – the bulk of which consist of various infrared techniques, which use a variety of relationships to infer rainrate from differences in brightness temperature as a function of wavelength, and microwave techniques, which observe the emission

properties of suspended hydrometers above the surface. Infrared techniques are limited in that they tend to rely on non-robust relationships between cloud-top properties (which in turn, determine the brightness temperature in the IR) and actual rainfall. Microwave properties are limited when the surface has an unknown microwave emissivity. Over water, microwave retrievals tend to perform better, due to the reflectivity of the surface in microwave frequencies.

For this research, we use Level-2 rainfall retrievals from the AMSR-E instrument on the EOS/PM (Aqua) satellite, as described in Wilheit, et al. (2003). Using a microwave retrieval makes the most sense for our region, which is nearly entirely covered by ocean, and in particular, the microwave retrieval from the AMSR/E instrument has the additional benefit of being located on the same spacecraft as the MODIS instrument we're using for our other observations, thus maximizing data overlap between the two datasets.

AMSR/E rainfall data are processed and gridded (in a non-interpolative manner) with co-located MODIS observations of cloud-top height and cloud optical depth to ensure pixel overlap between the two instruments – pixels containing only MODIS or AMSR/E data, and not both, are discarded. A sample image over the region of interest, containing MODIS observations of cloud-top height and cloud optical depth along with the corresponding AMSR/E rainfall retrieval is presented as Figure 3.1.

3.3 Selection of a Suitable 3D Histogram

Having selected a suitable dataset for rainfall observations, we now seek a method to integrate observations of precipitation along with our previously described observations of cloud-top pressure and cloud optical depth into our cluster analysis.

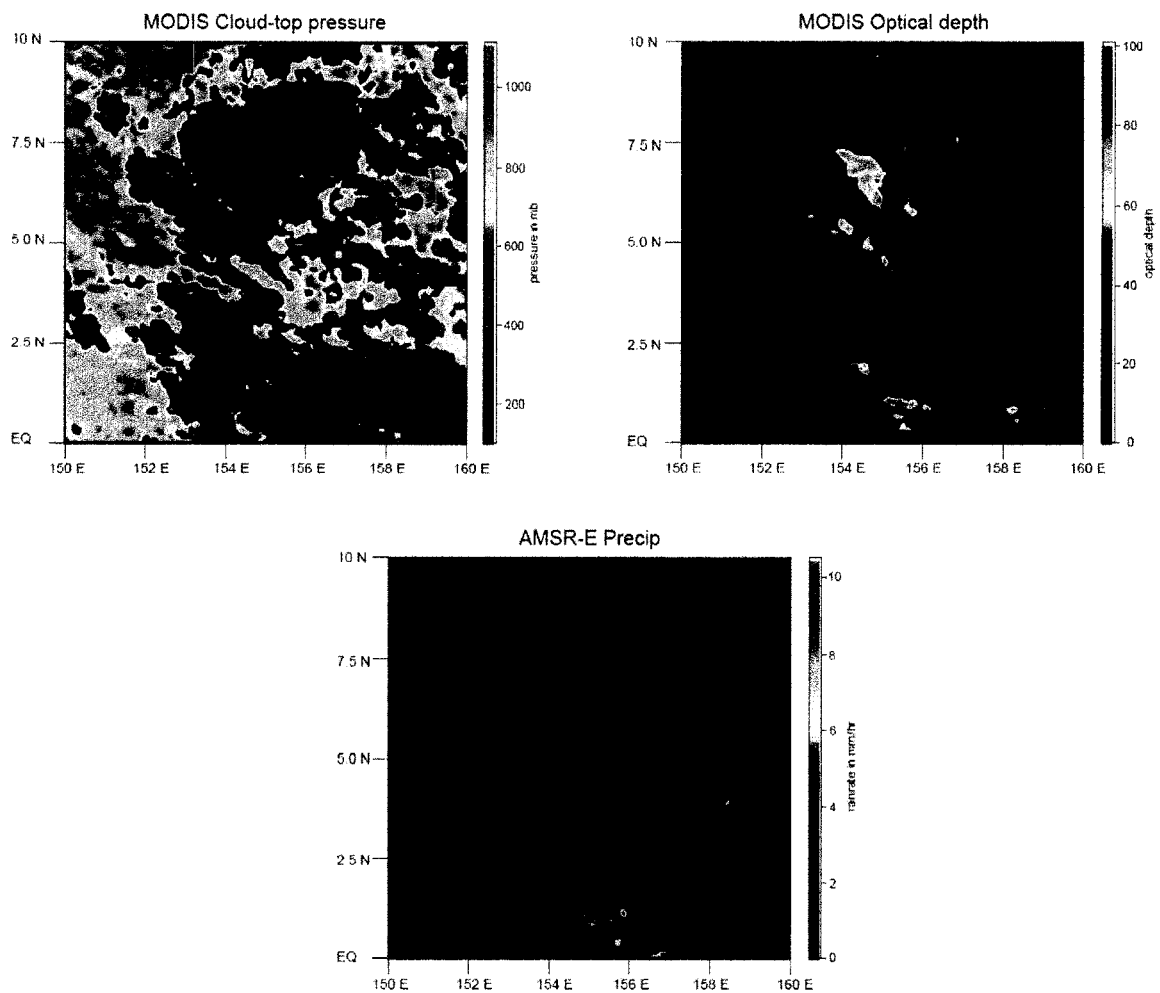


Figure 3.1. 21 June 2002 observation of cloud optical depth and cloud-top height as seen from MODIS, and rainfall as seen from AMSR/E.

Ideally, the combination of the three observations into a cogent 3D histogram will facilitate comparison to the well-known properties of the 2D histogram analysis, but will also demonstrate increased sensitivity to precipitation characteristics of cloud organization, particularly with regards to convective organization.

One way to approach this goal is to re-use the 2D comparison (the ISCCP cloud classification scheme) between cloud-top pressure and optical depth in the framework of varying levels of rainrate. In the 2D histogram representation, cloud optical depth is sorted into increasing values along the x-axis, while cloud top pressure is sorted along the y-axis. Bin sizes for the 2D histograms are, as previously discussed, determined by the ISCCP cloud classification scheme. For the 3D histogram representation, we propose to use the same x- and y-axis bins for optical depth and cloud-top pressure as in the 2D histograms, and sort the precipitation along the z-axis into bins whose sizes are determined by an external analysis of the precipitation – in this case, an additional, one-dimensional clustering of observed rainrates.

Of course, there are several ways one could organize the third dimension of our 3D histograms – what we seek is a simple, intuitive method that relies on the properties of the precipitation itself, and not on externally-derived parameters, most of which rely on assumptions about precipitating microphysics. What is necessary for the clustering histogram is that the bin sizes be distinguishable based on gross properties only – moreover, we desire a mechanism that describes these gross properties in a realistic, and preferably objective, manner.

To this end, the third dimension of the histogram is established as follows: a simple, one-dimensional clustering of AMSR/E rainrates is performed in order to find the

most common rainrate values in the dataset, and then use those rainrate values as the center values for our z-axis bins, with the bin widths determined by the spacing between the one-dimensional clusters. This entails, of course, an exercise in miniature of the entire research project – we must categorize the AMSR/E data in a large vector readable by the cluster algorithm, decide on an appropriate number of clusters to use (which, in turn, will decide the number of bins for our 3D analysis later on) and run the requisite tests to ensure that the cluster results are repeatable and meaningful in a statistical sense.

The data used to determine the rainrate bin value are the same data to be used in the 3D analysis – namely, the AMSR/E Level-2 rainrate product from June 2002-May 2003 over the previously-discussed region of the tropical West Pacific. At the end of the analysis, we will have, in effect, an amalgam of the most commonly-found rainrates in this region, at least as observed by the AMSR/E platform. To our knowledge, this is the first time that tropical rainrate has been categorized using clustering techniques.

The precipitation observations are broken down into a one-dimensional vector of satellite pixels containing rainrate values. (Since the intent is to categorize the properties of the *precipitation*, non-precipitating pixels in the cluster analysis are rejected – when we use the results of the one-dimensional cluster analysis to create bins for our 3D analysis, we will add an additional bin for non-precipitating pixels.) We then apply our clustering algorithm to this vector, choosing k initial centroids from the database, and then iterate through the vector. The principals discussed previously guide our choice of k – each derived cluster value for an appropriately-chosen value of k will be sufficiently different from other cluster values, and will not change significantly when different initial centroids are chosen from the vector.

We find three clusters of rainrate values in our dataset (namely, for this clustering, that $k=3$). They are, in order of increasing mean rainrate, 0.342 mm/hr, 2.88 mm/hr, and 8.04 mm/hr, with relative frequencies of occurrence of 68.04%, 23.07%, and 8.89%, respectively. For the sake of convenience, we label these clusters as ‘low’, ‘medium’, and ‘high’ precipitating clusters, again respectively. As might be expected, the bulk of precipitating pixels found were of medium-to-low rainrates, with a smaller fraction of high rainrates (corresponding to the smaller areal fraction of convective cores, presumably.) Since we cannot ascribe any cloud properties to the clusters based on our simple 1-D clustering, it is difficult to compare these values with other studies of cloud fraction as of yet. If we assume, however, that the lowest rainrate is more likely to be associated with stratiform (as opposed to convective) cloud structure, then the relative frequency of occurrence of our ‘stratiform’ rain compares favorably with analyses of stratiform rain fraction discussed in Schumacher and Houze (2003). There are some important differences – namely, that the RFO values from our analysis are not strictly a frequency in time or of areal coverage, but instead one of *pixel occurrence*. Naturally, however, there should exist some relationship between the three – if a region experiences a majority of stratiform precipitation in both frequency and areal coverage, one would expect a similar majority in the number of pixels reporting stratiform-like precipitation.

For this particular analysis, we found that $k=3$ was the optimal value. An attempt to run the analysis with $k=2$ failed entirely to converge on final cluster values, presumably due to the great variability of rainrates in the dataset. Choosing $k=4$ for the cluster analysis yielded the following cluster rainrate values: 0.184 mm/hr, 1.185 mm/hr, 3.37 mm/hr, and 8.34 mm/hr, with relative frequencies of occurrence of 52.03%, 22.46%,

17.74%, and 7.76%, respectively. However, histograms of rainrate amount separated by cluster for the $k=4$ case demonstrate a significant amount of overlap in rainrate between clusters. This leads to some uncertainty in the significance of each cluster in the $k=4$ case, as overlapping clusters represent individual observations of rainrate that are assigned to different clusters, despite having the same value. Rossow et al. (2005) describes this phenomenon as well, and it is typically used to ensure that the value of k is not too large. While there exists some overlap in the $k=3$ case, it is less than in the $k=4$ case, making it a better solution than the $k=4$ case. Figure 3.2 presents histograms (by cluster) for the $k=3$ and $k=4$ cases, and demonstrates the degree of overlap between the two cases. As one would expect, increasing the number of clusters (setting $k > 4$) yields increased overlap between clusters. Following the guidelines of Rossow et al. (2005) we therefore select the best option, which is the $k=3$ case.

The results of both the $k=3$ and $k=4$ cases give similar enough results to draw some general inferences about the nature of tropical precipitation, at least in the region of interest (and as observed by AMSR/E.) Both cases break down into roughly three precipitation scenarios – one of sub-mm/hr rainrate, with the highest frequency of occurrence, another of light-to-moderate rain of the order of ~ 3 mm/hr, and one of heavier precipitation likely associated with deep convection. The relative occurrence of each precipitation type agrees generally with other observations of tropical rainfall – furthermore. Using the results from our $k=3$ cluster analysis, and recalling that we want

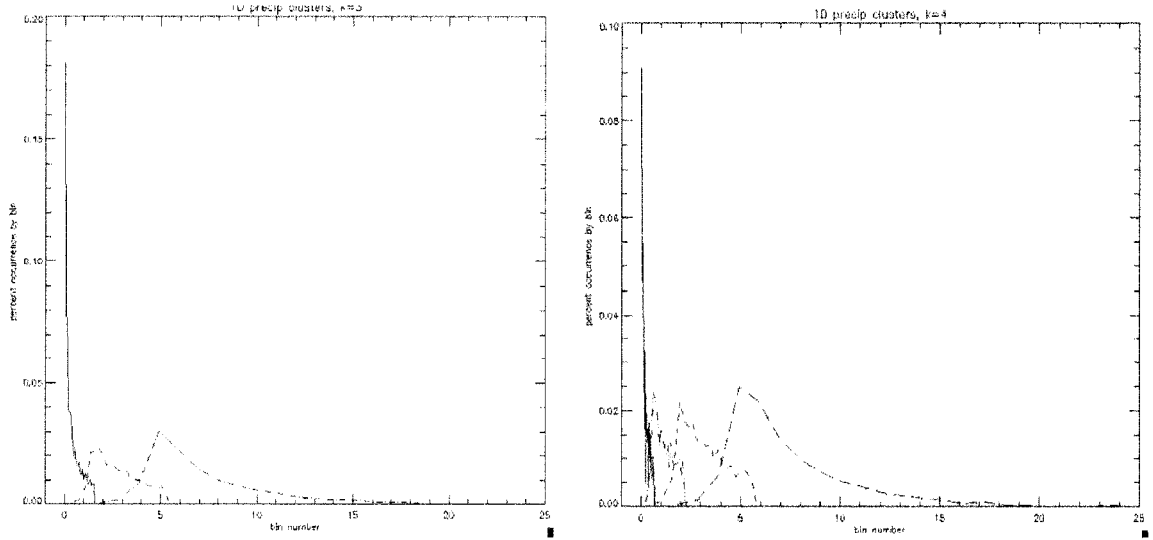


Figure 3.2 – histograms of precipitation amount sorted by cluster. The histogram for the $k=3$ case is shown on the left; $k=4$ on the right. Overlapping regions in the histogram indicate that individual identical rainrate amounts are being assigned to different clusters, signifying less certainty in ‘membership’ to a particular cluster. Overlap in the $k=3$ case is less than in the $k=4$ case, particularly for low rainrate observations.

Table 3.1: Sorting bin values for MODIS-AMSR/E 3D histogram

Cloud Optical Depth	Cloud-top Pressure	Observed Rainrate
Bin 1: 0-1.3	Bin 1: sfc-800 mb	Bin 1: 0-0.06 mm/hr
Bin 2: 1.3-3.6	Bin 2: 800-680 mb	Bin 2: 0.06-1.5 mm/hr
Bin 3: 3.6-9.4	Bin 3: 680-560 mb	Bin 3: 1.5-5.1 mm/hr
Bin 4: 9.4-23.0	Bin 4: 560-440 mb	Bin 4: > 5.1 mm/hr
Bin 5: 23-60	Bin 5: 440-310 mb	
Bin 6: > 60	Bin 6: 310-180 mb	
	Bin 7: < 180 mb	

to allow for non-precipitating cloudy pixels, we propose to use the following definitions to sort rainrate amounts in our 3D histograms: *Bin 1.)* contains pixels with rainrates < 0.06 mm/hr – our non-precipitating cluster. *Bin 2.)* contains pixels with rainrates between 0.06 and 1.5 mm/hr, centered on the ~ 0.3 mm/hr cluster. *Bin 3.)* contains pixels with rainrates between 1.5 and 5.1 mm/hr, roughly centered on the 2.88 mm/hr cluster. Finally, *Bin 4.)* contains pixels with rainrates higher than 5.1 mm/hr. The range of each bin is based on the rainrate value at the overlap points between clusters from the histogram described in Figure 3.n – this effectively ‘splits’ the uncertainty between clusters, assigning lower rainrates into lower bins and vice versa. Our 3D histogram will thus be sorted according to criteria outlined in Table 3.1. An example 3D histogram, constructed using the data from the 21 June 2002 shown in Figure 3.1, is presented as Figure 3.3.

3.4 Selection of value for k in 3D Histogram Cluster Analysis

Arguably, the selection of a histogram with which to sort multiple satellite observations into a coherent picture is the most important step in this kind of cluster analysis. A close second would be the selection of the number of centroids with which to sort the histograms, especially since there exists a certain amount of subjectivity in this selection. Earlier, in Chapter Two, we discussed different approaches to the selection of k , varying from choosing a number leading to results that are ‘visually significant’ to the approach of Rossow et al. (2005). Further experimental attempts to objectively determine k involve more rigorous statistical analysis of resulting clusters for different choices of k than are described by Rossow et al. (2005) – typically, a parameter that describes an ‘optimal’ relationship (defined as λ , which varies depending on the kind of

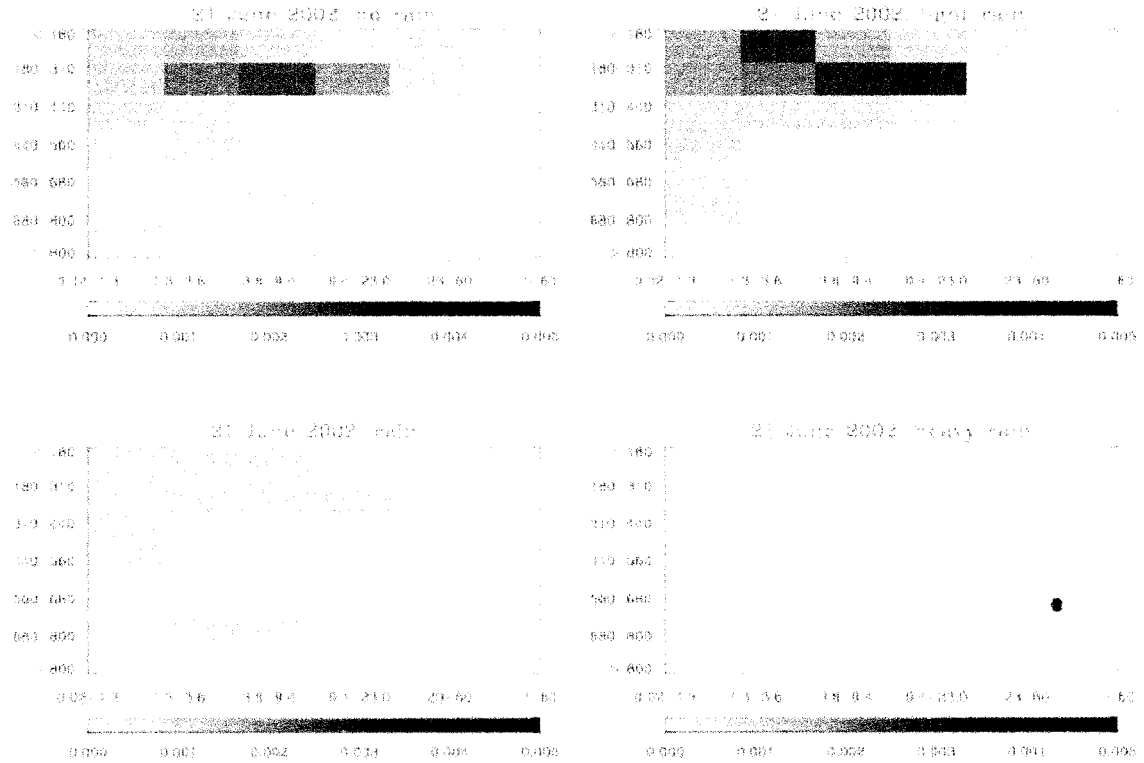


Figure 3.3. 3D histogram of satellite observations from 21 June 2002, (shown in Figure 3.1) Each of the four panels for each cluster depict a 2D cloud-top height/optical depth histogram for a certain rainrate bin. Here, “no rain” represents rainrates < 0.06 mm/hr, “light rain” for rainrates between 0.06 and 1.5 mm/hr, “rain” for rainrates between 1.5 and 5.1 mm/hr, and “heavy rain” for rainrates > 5.1 mm/hr. Brighter colors in a histogram represent higher counts of cloud pixels per bin, and more generally, a higher amount of cloud cover for that histogram.

data being analyzed) is computed for varying values of k , and an iterative approach to minimize λ is used to determine the appropriate value for k . Such methods (known as *lambda cluster analyses*) fall outside of the strict definition of our familiar k-means analysis, and perhaps more properly belong in the field of statistics. We would like to take the *via media* when determining k – i.e. to remove as much subjectivity in the selection as possible, without necessarily having to break new ground in the field of statistics to do so.

With some modifications, we propose that the criteria determined by Rossow et al. (2005) constitute a useful set to determine an appropriate value for k , at least for the style of cluster analysis of this study. The criteria used are:

- 1.) Resulting clusters for an appropriate value of k will have high pattern correlation with themselves if different initial centroids are chosen (which we will refer to as the *stability principle*.)
- 2.) The Euclidian distance between individual observations and their ‘parent’ cluster (defined as the cluster to which the analysis assigns said observation) will be less than the intra-cluster distance between the ‘parent’ cluster and the other clusters. We refer to this as the *confidence principle*.
- 3.) Finally, resulting clusters for an appropriate value of k will have low pattern correlations *in concert with high ‘confidence’ (as defined above)* with other resulting clusters, which we will refer to as the *significance principle*.

The fourth criterion of Rossow et al. (2005) omitted here is that of low correlation of cluster frequency over space and time, which is particularly useful for cluster analyses over large regions (say, the entire tropics.) For large regions, there is sufficient variability in cloud observations to potentially create ‘sub-clusters’ of clusters – that is, nominally identical clusters which exhibit just enough difference in pattern correlation to qualify as separate clusters via the significance principle. However, when the spatial- and temporal characteristics of these ‘sub-clusters’ are analyzed, there exists a significant amount of overlap between the ‘sub-clusters’, and that in fact, the differences are merely an artifact of the observation dataset rather than a real difference in clusters. For our analysis, we have chosen a spatial scale that is suitable to observe cloud-scale phenomena - while the clouds in our defined region do exhibit considerable variability, the variability is nothing like that which would be found in a global-scale analysis. We expect, therefore, that differences in pattern correlation between resulting clusters to be sufficient to satisfy the significance principle for our smaller region.

Another consideration is the utility of using pattern correlation in determining the ‘significance’ of resulting clusters. Rossow et al. (2005) determine that resulting clusters are significant (that is to say, significantly different from other resulting clusters) if the pattern correlation (defined as the Pearson expected value coefficient of linear correlation) between the two clusters is ≤ 0.6 . However, there has been little or no discussion as to *why* exactly pattern correlation is the preferred method of determining cluster significance.

We propose the following thought experiment to explore the utility of pattern correlation: consider a function $f(t)$ varying in time. We trivially define the following additional functions in terms of $f(t)$:

$$\begin{aligned} g(t) &= c_1 f(t) \\ h(t) &= c_2 f(t) \end{aligned} \tag{3.1}$$

Without assigning any physical meaning to the functions, we can perform a linear correlation analysis between $f(t)$ and $g(t)$, and between $f(t)$ and $h(t)$. This is a simple exercise – since the functions $g(t)$ and $h(t)$ are simply linear multiples of $f(t)$, the linear correlation in both cases will be unity. Let us now assign a physical meaning to the functions – say that $f(t)$, $g(t)$, and $h(t)$ represent the different areal cloud fraction of stratocumulus for different regions as a function of time. (For the sake of simplicity, we also assume that the only clouds in the regions are stratocumulus – thus the fractions described by the functions represent also the total cloud fraction for the region.) In this thought experiment, the cloud cover amount for the three regions will differ based on the value of the constants c_1 and c_2 . Let us assign values of $c_2 \approx 1$, and $c_1 \ll c_2$, and then evaluate our hypothetical model in the framework of cloud radiative forcing. For this case, there will be a marked difference in cloud radiative forcing between the cloud fractions represented by $g(t)$ and $f(t)$, and considerably less difference in cloud radiative forcing between the cloud fractions represented by $h(t)$ and $f(t)$ – despite the fact that all three regions are represented by cloud fraction functions that correlate identically with one another. This finding is not surprising – the linear correlation describes only how the functions vary identically with time in this case, and clearly has no bearing on the physical properties (or significance) of the cloud fractions described by the three functions.

With this in mind, let us turn to our cluster analyses. As we described in Chapter 2, it is possible to integrate over the bins of a cluster's histogram to determine the total cloud cover (TCC) of the region – this integration is possible because each bin contains the number of pixels matching a certain range of cloud parameters. Integrating over every bin therefore counts the total number of cloudy pixels in the scene, which can then be compared to the total number of pixels observed in the scene to compute the fraction of total cloud cover. Thus, the *magnitude* of the bin counts in a histogram can convey important *physical* information – in this case, the more counts per histogram, the greater the total cloud cover. This information *must* be included along with the spatial pattern of histogram bin population when considering the total amount of information found in a histogram – but linear correlation only considers the spatial pattern of bin histogram population, and disregards the magnitude information. As in our simple example above, two histograms with superficially similar spatial patterns, but different magnitudes of bin populations (and therefore, different cloud cover amounts) will correlate well with one another. This does *not*, however, necessarily mean that the two histograms are not physically significant from one another, simply because the correlation analysis does not take into account all of the relevant information to determine significance. In other words, using correlation analysis to determine cluster significance rules out the possibility of having two clusters with similar cloud *types*, but radically different amounts of cloud *cover*. If we want to describe the unique properties of each cluster as a cloud regime, we need to include the possibility of regimes that vary primarily by cloud cover amount as opposed to cloud type to be complete.

How then do we assess the significance of cluster results if correlation analysis doesn't tell the whole story? One of the nicer features of k -means cluster analysis is that these kinds of issues are directly addressed – two histograms with similar spatial patterns but different magnitudes of population will have non-trivial Euclidian distances from each other, and will self-identify thusly. Of course, this is of little help when attempting to select an ideal value of k . In an ideal world, variations between observations would be sufficiently large that observations would neatly sort themselves, with little or no ambiguity, into the correct number of clusters. In this ideal case, there would be little difference between the properties of observations within a cluster – much less than the differences from cluster to cluster (the so-called confidence principle.) In the real world, there is more ambiguity in how individual observations are sorted into clusters (due to the nature of the observations themselves), but the fact remains that, for an appropriate value of k , the differences between members of a cluster will be less than the differences between centroid values – choosing a value of k that is too large will result in duplicate clusters that exhibit smaller intra-cluster distances. We propose that in order to assess the significance of cluster results, the pattern correlation analysis used in previous works must be examined in the framework of inter- and intra-cluster distance in order to properly select the value of k , and that fixed values of pattern correlation used as accept/reject criteria may not be the best method to determine significance (in any case, it should at least not be the *first* method used).

Thus, an appropriate value of k for this analysis will result in unique clusters, which will persist when different initial values for the centroids are chosen, and are made up of observations that 'belong' more closely to each cluster than they do to any other

cluster. In order to test these properties, for each value of k selected, we run the following tests:

- 1.) To test the stability principle, the cluster algorithm is performed once, with the initial centroids selected evenly throughout the observations at intervals of k/N , and saving the cluster results. The algorithm is then repeated 5000 times, this time selecting the initial centroids randomly. The cluster results from each random-centroid run are pattern-correlated against the results from the evenly-spaced centroid run. Naturally, the order of the new clusters will differ from the order of the evenly-spaced runs – for this test, we pattern correlate all of the randomly-derived clusters against each of the evenly-derived clusters one at a time, and the cluster with highest pattern correlation thus found is assumed to be the randomly-derived ‘counterpart’ to the evenly-derived clusters. Note that for the purpose of testing stability, a simple pattern correlation is sufficient to determine similarity of clusters. High pattern correlation between the random runs and the evenly-spaced runs are required to confirm the stability principle.
- 2.) To test the confidence principle, the mean intercluster distance for each cluster is computed, and compared to the intracluster distances computed between each cluster. If the largest mean intercluster distance is less than the lowest intracluster distance, we determine that the confidence principle is satisfied.
- 3.) Finally, we test the significance principle by comparing the inter- and intracluster distances computed above with the pattern correlations between the resulting clusters. If the pattern correlation between a selected cluster and

each of the other clusters is less than 0.6, then we determine that the significance principle is satisfied. If the pattern correlation between any cluster and another cluster is greater than 0.6, but the intercluster distances for each cluster are smaller than the intracluster distance between the two clusters, then we also confirm the significance principle. Any other combination fails to confirm significance, and indicates that k is likely too large.

In order to qualify as a candidate value for k , all three tests must be passed. We now discuss the results of these tests for values of k between three and six.

3.4.1 Results for $k=3$

The three clusters found using $k=3$ are presented in Figure 3.3. Each cluster is represented by a four-panel set, with each panel representing a 2D cloud-top pressure/optical depth histogram at one of the four rainrate bins described in section 3.3. Cluster 1 in the $k=3$ case appears to be largely made up of high, thin clouds with along with a population of lower, also thin clouds, with little precipitation. This cluster has a relative frequency of occurrence of 63.5%. Cluster 2 is a classic deep- and organized convection cluster, exhibiting high thick clouds with moderate to heavy rainfall. The relative frequency of occurrence for this cluster 2 is 7.3% - similar to the values found for organized convection in the 2D case. Cluster 3 is also made up of high, thick clouds, although somewhat less high and thick when compared to cluster 2. Rainfall for cluster 3 in this case is largely in the light- to moderate category, although some heavier rainfall exists as well. The relative frequency for cluster 3 is 29.2%.

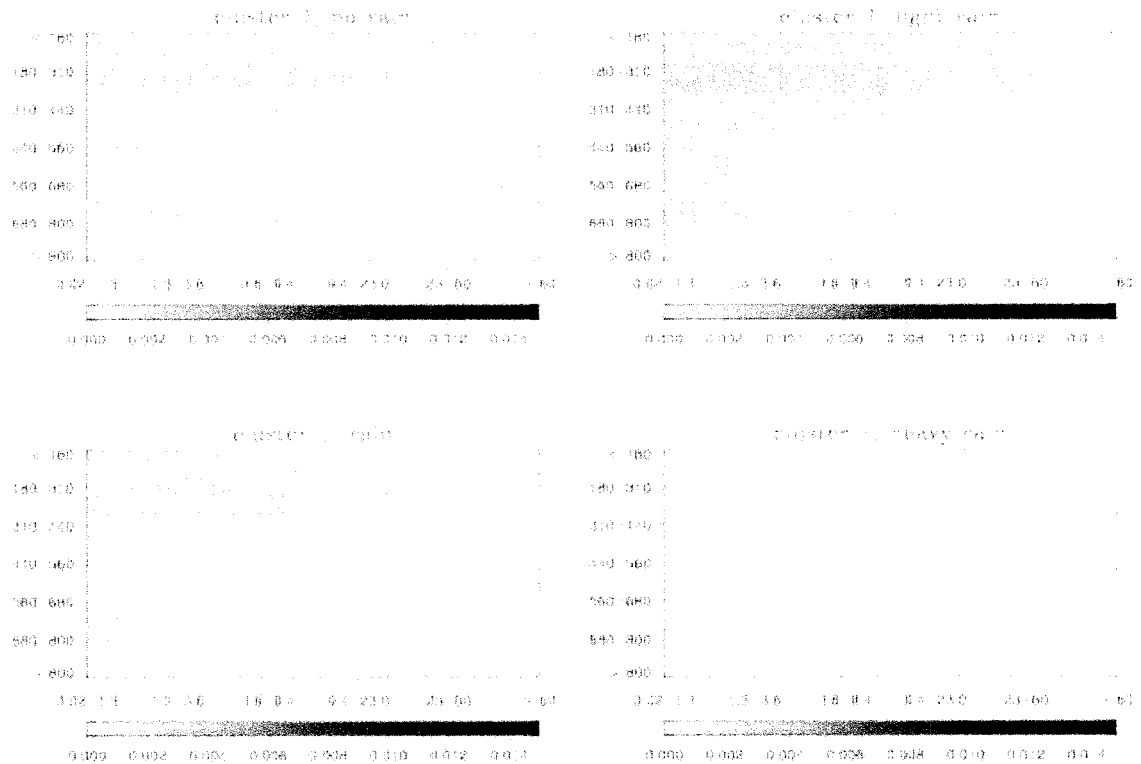


Figure 3.4a. Cluster 1 found using k-means analysis with $k=3$. Each pane of the figure represents a joint CTP-optical depth histogram for a single precipitation bin, in the manner for Figure 3.3.

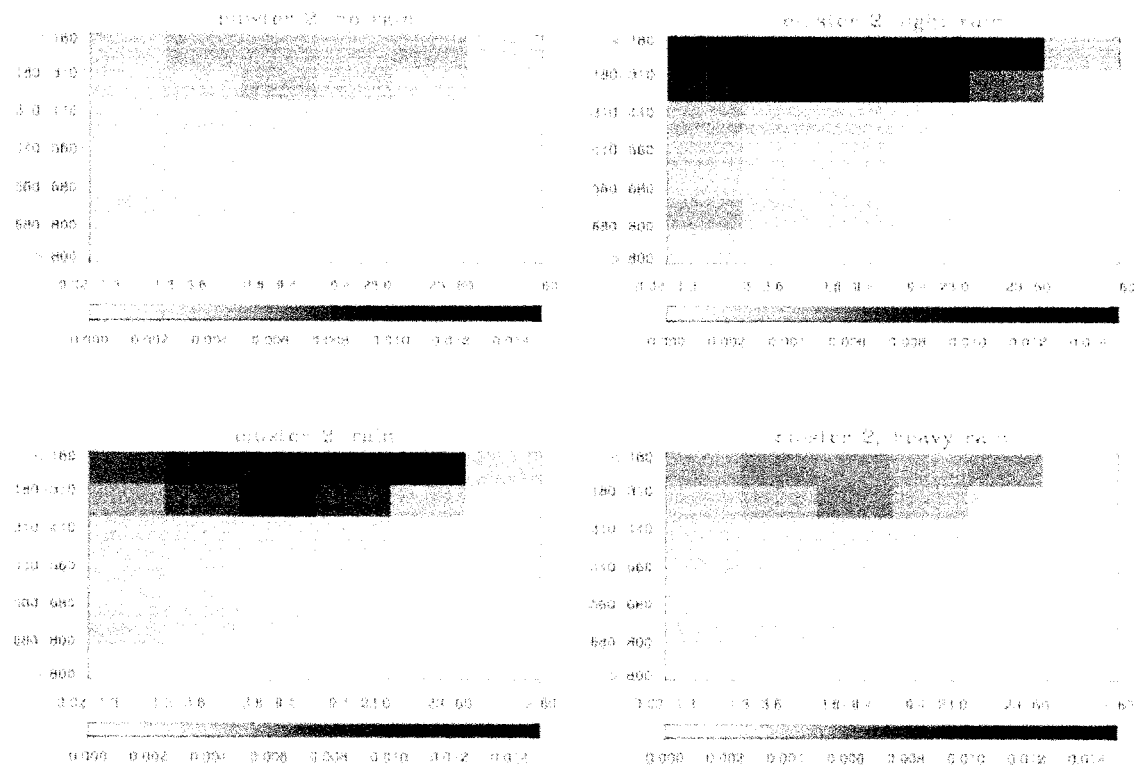


Figure 3.4b. As in Figure 3.4a, but for cluster 2 found using k-means analysis with $k=3$.

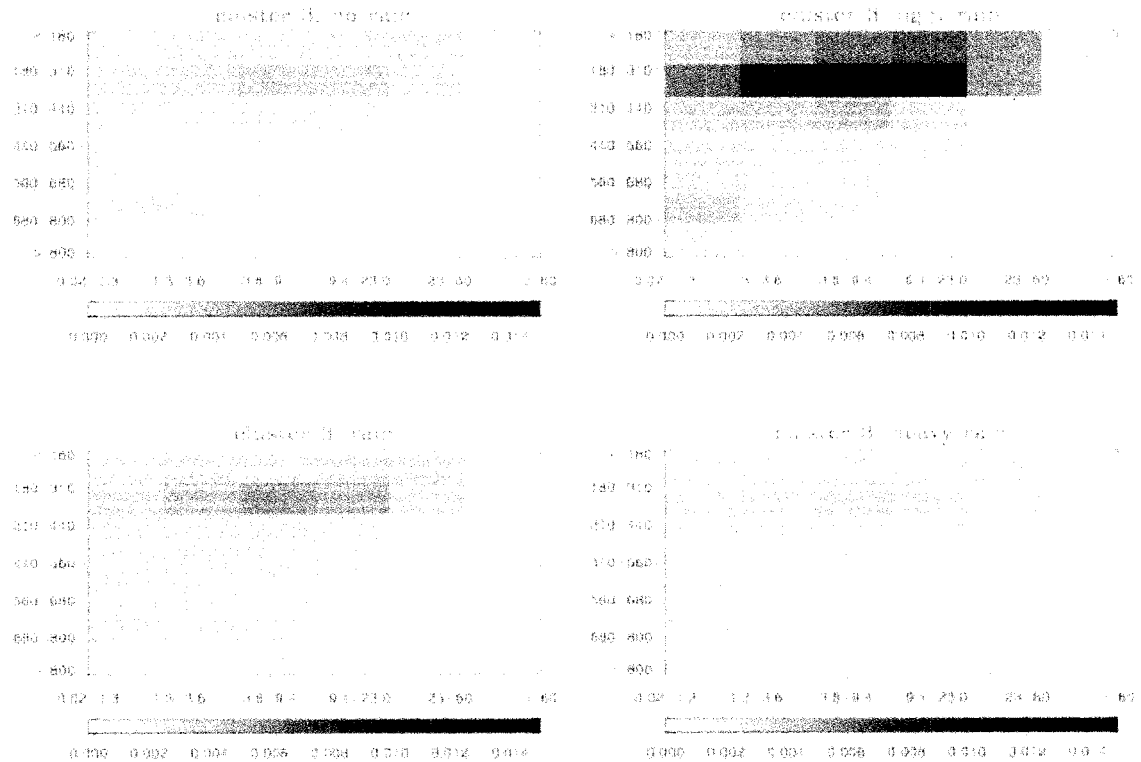


Figure 3.4c. As in Figure 3.4a, but for cluster 3 found using k-means analysis with $k=3$.

After 5000 re-runs of the $k=3$ case, using randomly-chosen initial centroids, we find that the $k=3$ case fulfills the requirements for the stability principle – the average correlation between randomly-chosen and evenly-chosen clusters were, respectively, 0.99, 0.99, and 0.99 for clusters 1, 2, and 3. Extremely high pattern correlations are typical for cluster stability comparisons using low values of k – in the extreme case of $k=1$, the correlation would be unity, exactly, after all.

Intercluster distances (again, defined as the mean Euclidean distance between each member observation and the parent cluster to which each belongs) is computed to be 3.32×10^{-5} (unitless) for cluster 1, 1.26×10^{-4} for cluster 2, and 2.16×10^{-4} for cluster 3. The intracuster distances between the clusters are computed to be 3.9×10^{-3} for cluster 1 and cluster 2, 5.95×10^{-4} between cluster 1 and cluster 3, and 1.96×10^{-3} for cluster 2 and cluster 3. The largest variability (and therefore, ambiguity in membership) is found in cluster 2, but this variability remains less than the distance between cluster 2 and any other cluster, thus satisfying the confidence principle.

Intracuster pattern correlations were found to be 0.72 between clusters 1 and 3, 0.82 between clusters 1 and 2, and 0.829 between clusters 2 and 3. The highest pattern correlation is, as might be expected, between the two convective clusters. However, as we have demonstrated a high level of confidence that individual observations ‘belong’ to their clusters, and with the previous discussion about the utility of pattern correlation in determining significance in mind, we are comfortable with the relatively high pattern correlations found here. Having confirmed stability, confidence, and significance of the clusters found in the $k=3$ case, we submit that $k=3$ would be a valid choice for consideration in further analyses.

3.4.2 Results for $k=4$

The four clusters resulting in a selection of $k=4$ are presented as Figure 3.5. Cluster 1, with a relative frequency of occurrence of 21.9%, appears to be made up of high, moderately thick clouds with light- to moderate rainfall, with properties similar to those of developing convective towers. Cluster 2, with a relative frequency of occurrence of 17.5%, describes higher and thicker clouds than those of cluster 1, with greater incidence of moderate- to heavy rainfall. Cluster 3, with the highest relative frequency of occurrence (53.3%) is a low-precipitation, low total cloud cover regime similar to non-convective clusters found by JT03. Cluster 4, with a relative frequency of occurrence of 7.3% (again, the same as found in the 2D and the $k=3$ case above) describes very high and thick clouds, exhibiting a good deal of precipitation, including moderate- and heavy rainfall.

Again, 5000 re-runs of the analysis, each using randomly chosen initial centroids, are compared to the results above to assess stability. Pattern correlations between the randomly-derived clusters and the clusters shown above were very high – the lowest pattern correlation was for cluster 4, which had a pattern correlation of 0.99. Again, high pattern correlations between the clusters indicate that the analysis is insensitive to the choice of initial centroids; the stability of the analysis for $k=4$ is thusly confirmed.

Intercluster distances for the $k=4$ case are presented in Table 3.2. The maximum intercluster distance is again for cluster 4, at 0.000188 (unitless.) Intracluster distances for the $k=4$ case may be found on Table 3.3. The minimum intracluster distance, between clusters 1 and 3, is 0.000203 – again, since the intercluster distances are smaller than the

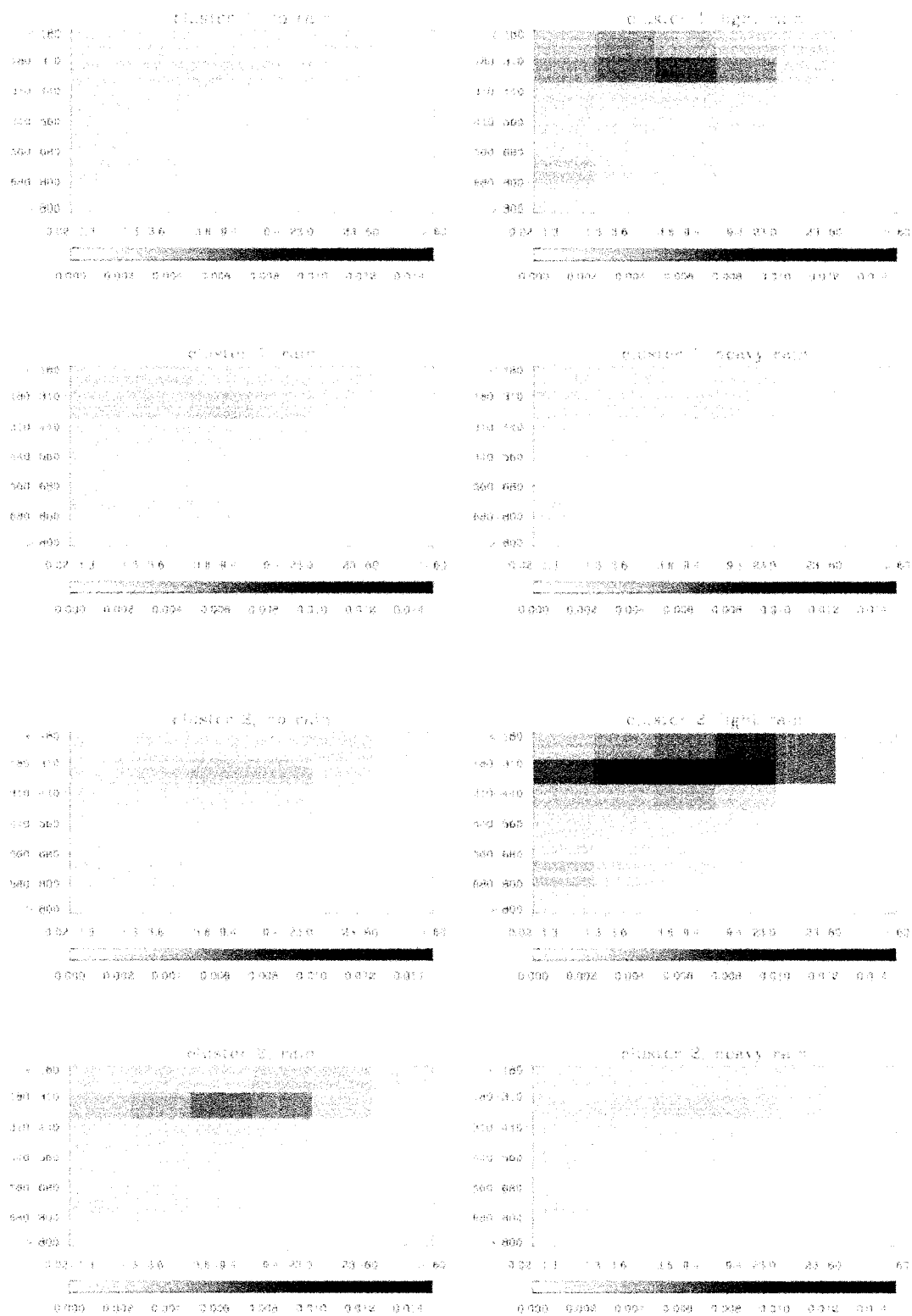


Figure 3.5a. Clusters 1 and 2 for k-means analysis with $k=4$. The upper four panels represent cluster 1, the lower four panels cluster 2

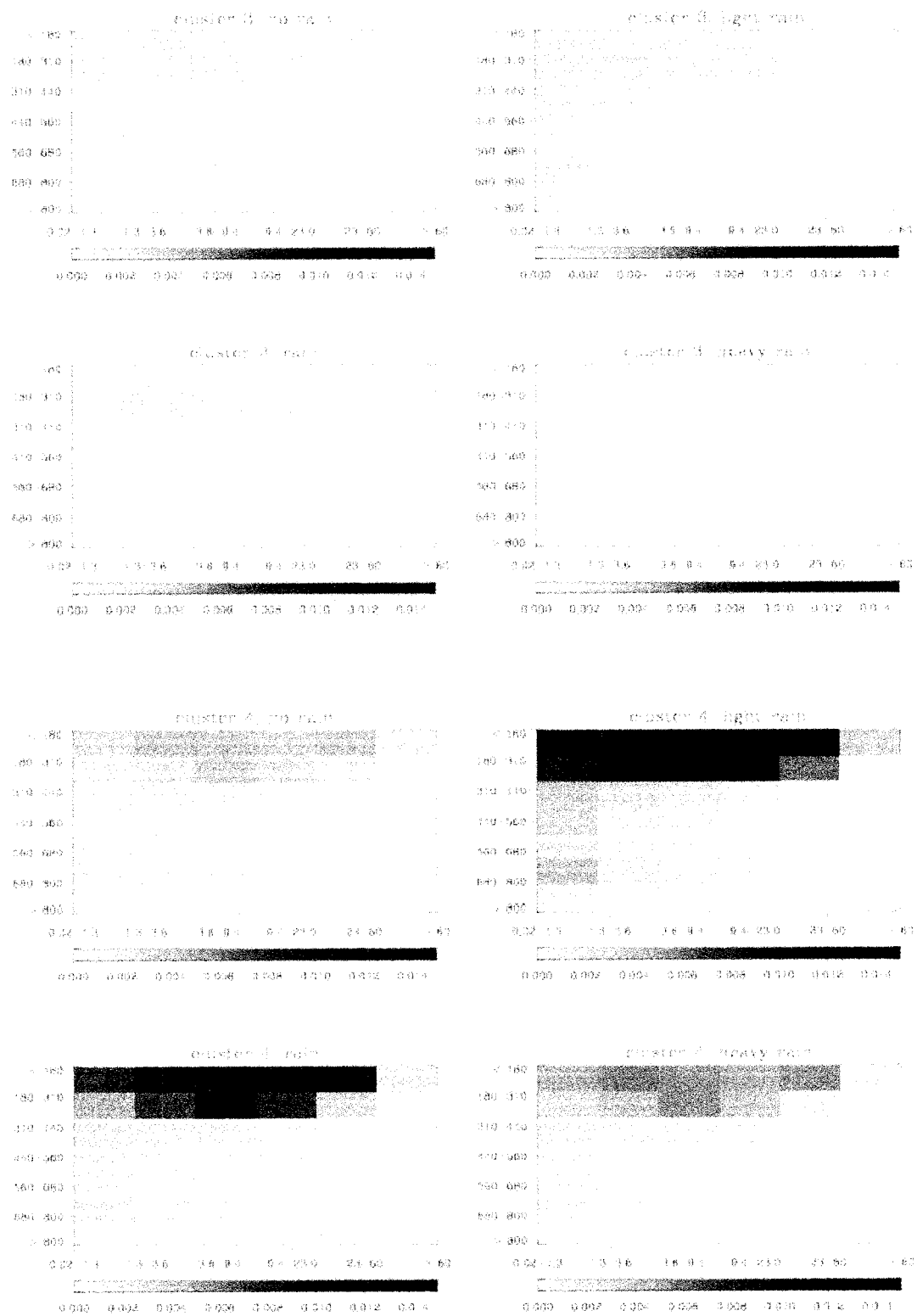


Figure 3.5b. As in Figure 3.5a, but for clusters 3 and 4.

Table 3.2. Intercluster distances for $k=4$

Cluster w/ description	Mean Intercluster Distance (unitless)
Cluster 1	1.18×10^{-5}
Cluster 2	9.71×10^{-5}
Cluster 3	1.89×10^{-4}
Cluster 4	6.54×10^{-5}

Table 3.3. Intra-cluster distances for $k=4$

	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>	<i>Cluster 4</i>
<i>Cluster 1</i>	---	3.34×10^{-4}	2.03×10^{-4}	2.96×10^{-3}
<i>Cluster 2</i>	3.34×10^{-4}	---	9.59×10^{-4}	2.08×10^{-3}
<i>Cluster 3</i>	2.03×10^{-4}	9.59×10^{-4}	---	4.43×10^{-3}
<i>Cluster 4</i>	2.96×10^{-3}	2.08×10^{-3}	4.43×10^{-3}	---

Table 3.4. Coefficients of pattern correlation between clusters in the $k=4$ case.

	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>	<i>Cluster 4</i>
<i>Cluster 1</i>	---	0.928	0.921	0.839
<i>Cluster 2</i>	0.928	---	0.789	0.792
<i>Cluster 3</i>	0.921	0.789	---	0.657
<i>Cluster 4</i>	0.839	0.792	0.657	---

distances between the clusters themselves, we have confidence that individual observations are being assigned correctly to the centroids.

Pattern correlation between the clusters, as one might suspect from Fig. 3.5, are somewhat higher – the pattern correlations between the clusters are shown in Table 3.4. The highest pattern correlations are found between clusters 1 and 2 (0.928) and between clusters 1 and 3 (0.921.) Both cases are of predominantly high clouds, of low- to moderate thickness and moderate- to low rainrates being compared to one another – the key difference is the amount of cloud coverage in each cluster. Both clusters 1 and 2, for example, contain moderately-thick high clouds of similar appearance (at least in the sense of how they are represented in our 3D histogram) – but there is a great deal more cloud cover in cluster 2 than in cluster 1, especially of clouds with low- and moderate rainrates. Similarly, cluster 3 contains markedly *fewer* clouds than cluster 1 – one might envision (and we’ll certainly do so, in greater detail, in just a bit) cluster 3 as a prototypical trade-inversion cumulus regime, with fewer clouds (and certainly less convection), and cluster 1 as a transitional regime (perhaps due to a breakdown of stability in the region) containing more convective clouds, and an increased amount of total cloud cover. From a histogram-centered perspective, the clouds in these three clusters are topologically similar, and pattern correlation is therefore high. The amount of cloud cover, and the distribution of *where* that cloud cover occurs at certain rainrates, is what distinguishes these clusters from one another in the cluster analysis. When considered along with a high confidence in assignment of individual observations to these clusters, we feel comfortable with the significance of the clusters as computed with $k=4$.

3.4.3 Results for $k=5$

Clusters found for the $k=5$ case are shown in Figure 3.6. The five clusters shown in Figure 3.6 depict our familiar deep-convective (cluster 5) and hot-tower (cluster 4) structures, with relative frequencies of occurrence of 7.3% (by now, a familiar number) and 27%, respectively. We also see a low-cloud cover/low precipitation cluster containing thin cirrus and low, likely cumuliform clouds (cluster 3) with relative frequency of occurrence of 42.3%, which is reminiscent of the trade-regime cluster, and two high-cloud clusters of moderate thickness and rainrate (clusters 1 and 2.) These latter clusters have relative frequencies of occurrence of 14.5% and 8.8%, respectively.

After performing 5000 re-runs of the analysis using randomly chosen initial centroids, we start encounter some issues. Pattern correlations between clusters are lower than previous cases, and in fact, several of the cases exhibit different characteristic clusters. Although we omit figures of the clusters themselves in the name of brevity, it is worthwhile to describe the differences between the random and the evenly-spaced runs. Cluster 1 from the random run is roughly similar to cluster 5 from the evenly-spaced run depicted in Figure 3.6; likewise, cluster 3 in the random case approximates cluster 4 of the evenly-spaced case, and cluster 4 of the random case is similar to cluster 1 of the evenly-spaced case. However, clusters 2 and 5 in the random case are markedly different than clusters 1 and 2 in the evenly spaced case. The differences are largely ones of cloud cover, although some topological differences do exist. These differences are sufficiently large enough that pattern correlation between the ‘related’ clusters for the evenly-spaced and the randomly-initiated runs are low: as low as 0.669 for randomly-initiated clusters

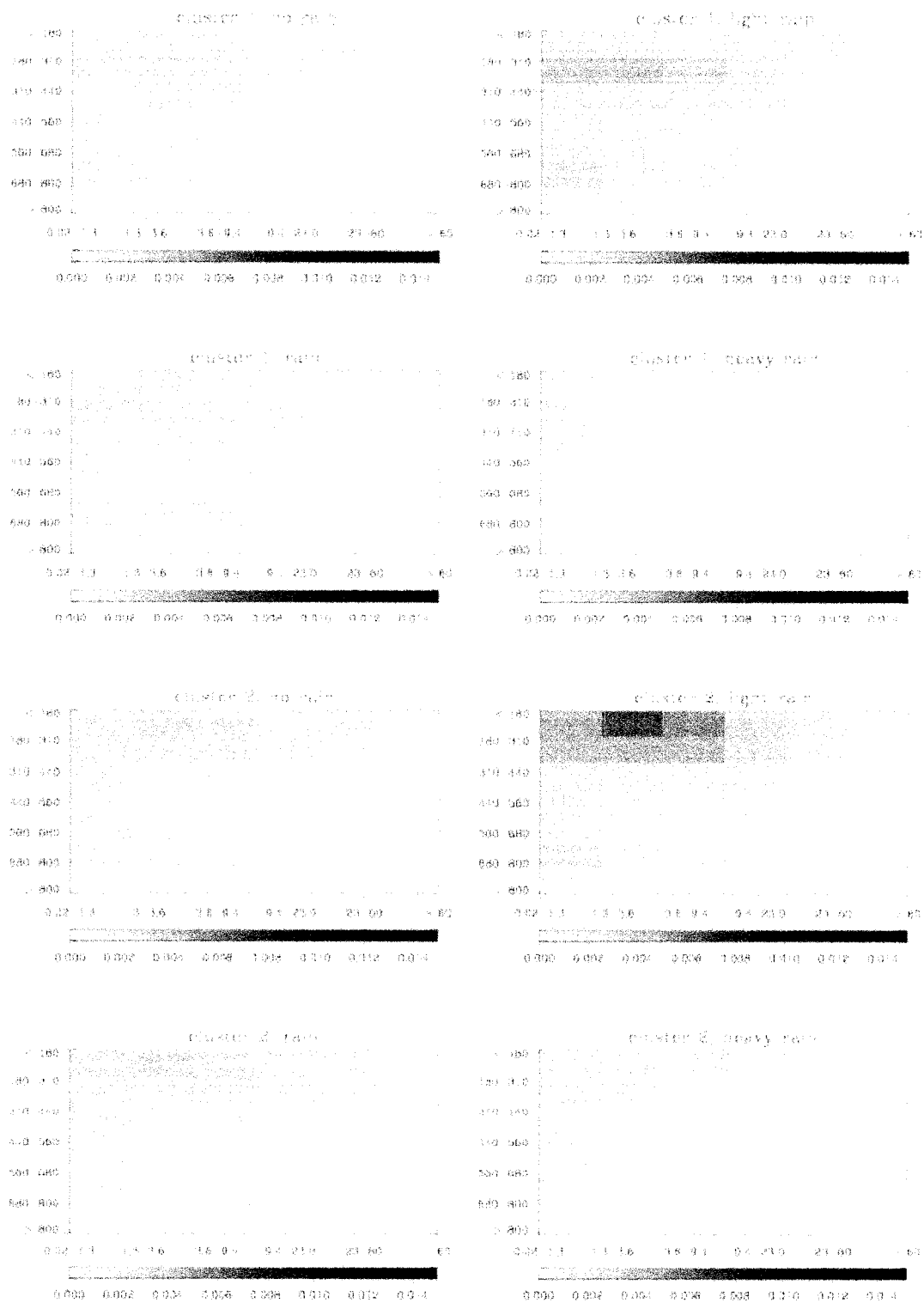


Figure 3.6a. Clusters 1 and 2 for k-means analysis with $k=5$. The upper four panels represent cluster 1, the lower four panels cluster 2



Figure 3.6b. Clusters 3 and 4 for k-means analysis with $k=5$. The upper four panels represent cluster 3, the lower four panels cluster 4

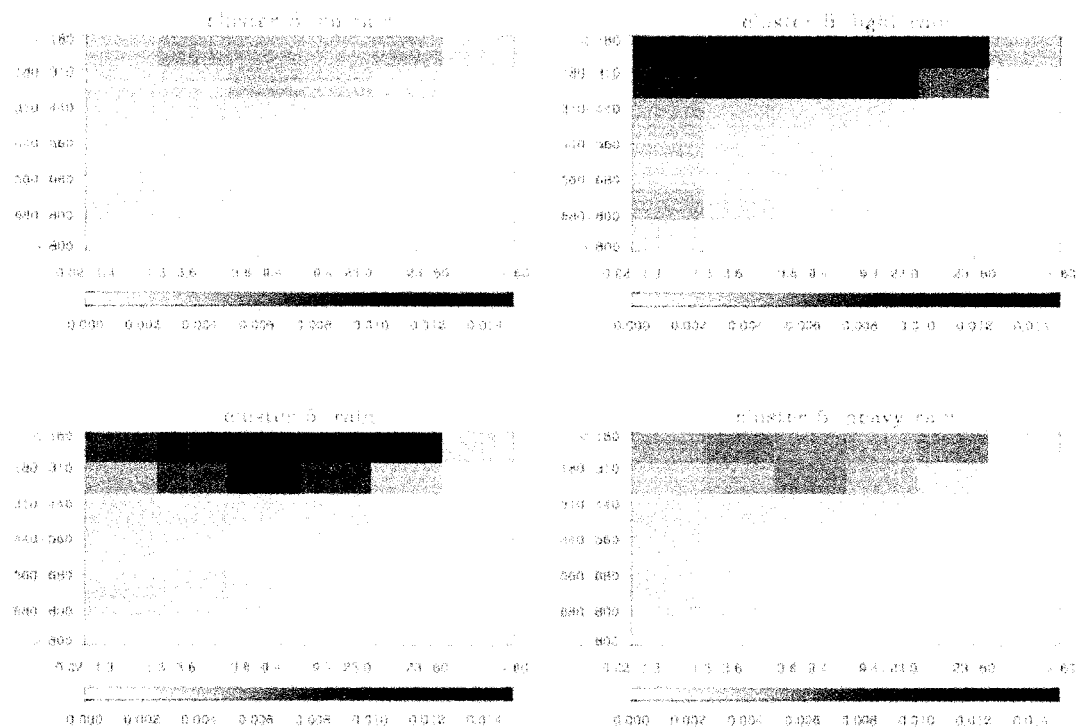


Figure 3.6c. Cluster 5 for k-means analysis with $k=5$.

that are identified as most similar to cluster 1 of the evenly-spaced runs. Table 3.5 contains the coefficients of pattern correlation for the $k=5$ case.

Average intercluster distances for the $k=5$ case are provided in Table 3.6. Intercluster distances for clusters 3-5 are generally small, and are comparable to other intercluster distances computed so far. The average intercluster distance for both clusters 1 and 2, however, are larger, signifying more ambiguity in membership of individual observations to each cluster. Mean intercluster distances for cluster 1 and 2 are, respectively, 0.000115 and 0.000187. The intracluster distance between cluster 1 and 2 is computed to be 0.0000814 – which is extremely close and is, in fact, smaller than the intercluster distances of clusters 1 and 2 (since this finding violates the confidence principle, we omit reporting of the intracluster distances in the interest of brevity.) In other words, there exists more variability within cluster 1 and cluster 2 than there exists difference between them. This variability also explains the results of the stability tests: since intercluster variability for the first two clusters is greater than the intracluster variability between them, it is perhaps not surprising that the results of the stability analysis will depend more on the initial choice of centroids.

Finally, we assess the significance principle for the $k=5$ case. Pattern correlations between the five clusters are not markedly higher than in other cases, with the exception of the pattern correlation between clusters 1 and 3, which is computed to be 0.934. Since we have determined that we have low confidence in the uniqueness of cluster 1, we cannot now dismiss this pattern correlation result, and must concede that the clusters in the $k=5$ case lack significance from one another.

Table 3.5. Average coefficients of pattern correlation between clusters for the $k=5$ case

Cluster 1	0.669
Cluster 2	0.984
Cluster 3	0.976
Cluster 4	0.996
Cluster 5	0.993

Table 3.6. Average intercluster differences within clusters for the $k=5$ case

Cluster 1	1.15×10^{-4}
Cluster 2	1.87×10^{-4}
Cluster 3	9.96×10^{-6}
Cluster 4	1.07×10^{-5}
Cluster 5	4.17×10^{-6}

3.4.4 Results for $k=6$

The six clusters computed for $k=6$ are shown in Figure 3.7. Having determined that $k=3$ and 4 are valid choices for analysis, and that $k=5$ is not a valid choice, we expect that increasing values of k above 5 should also fail. This is, in fact the case – specifically, we were unable to assess stability for the $k=6$ case due to great variability in the results when randomly-selected initial centroids were used in place of evenly-selected centroids. This is not entirely unexpected, since raising the value of k for a fixed number of observations splits the observations into more clusters composed of fewer members (in fact, when $k=N$, each observation becomes its own cluster.) This of course has the ultimate effect of increasing the mean intercluster distance. Minor differences between clusters composed of only a few observations are also not terribly stable, and the minor differences imparted by using different initial centroids become increasingly influential in the final outcome. For the $k=6$ case, the clusters found for each random case often bore little resemblance to the clusters depicted in Figure 3.7, and as such, had extremely poor pattern correlations. Since we have essentially no confidence in our ability to produce six unique clusters for the $k=6$ case, we therefore forego the remaining tests, and truncate our search for increasing values of k .

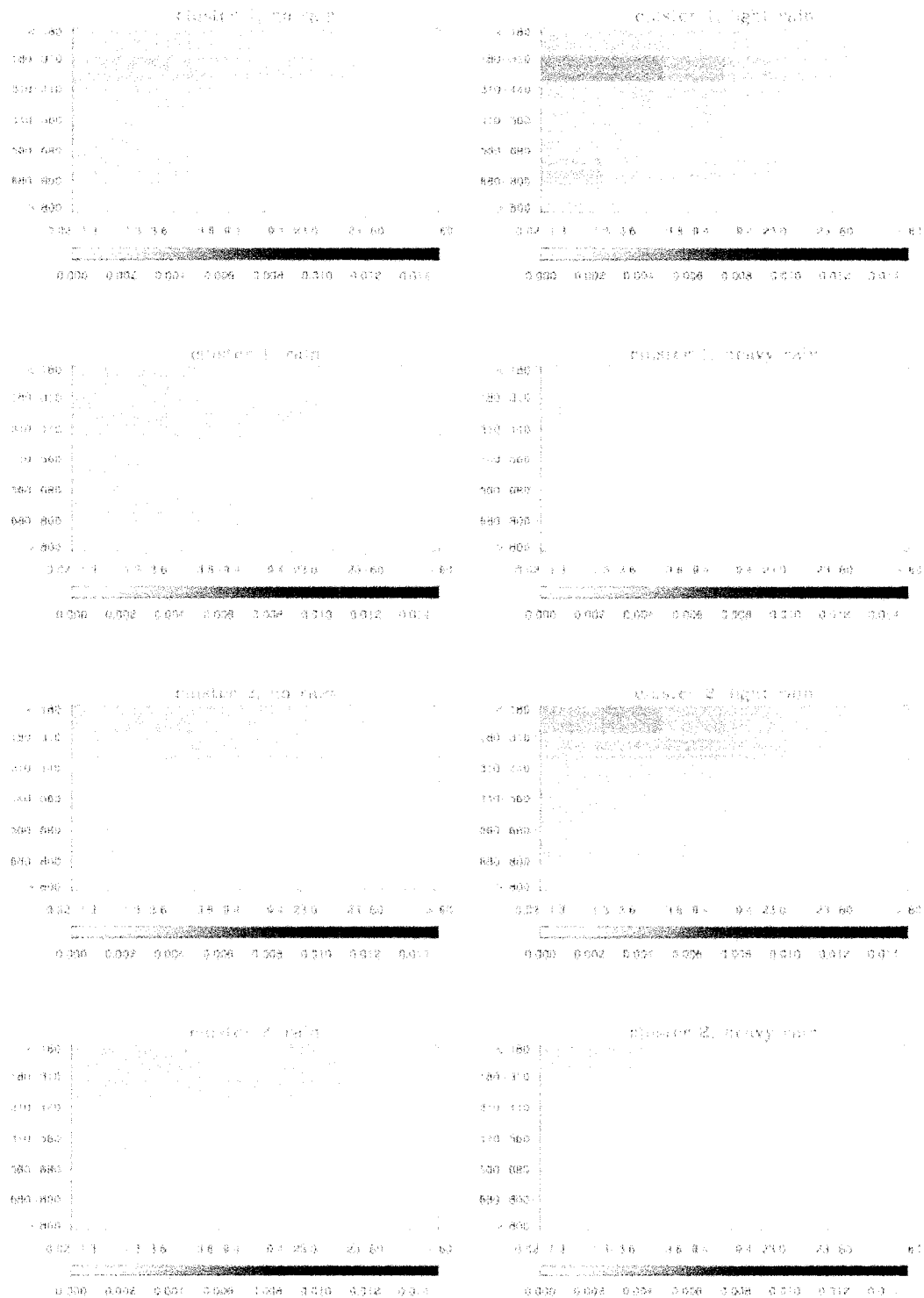


Figure 3.7a. Clusters 1 and 2 for k-means analysis with $k=6$. The upper four panels represent cluster 1, the lower four panels cluster 2

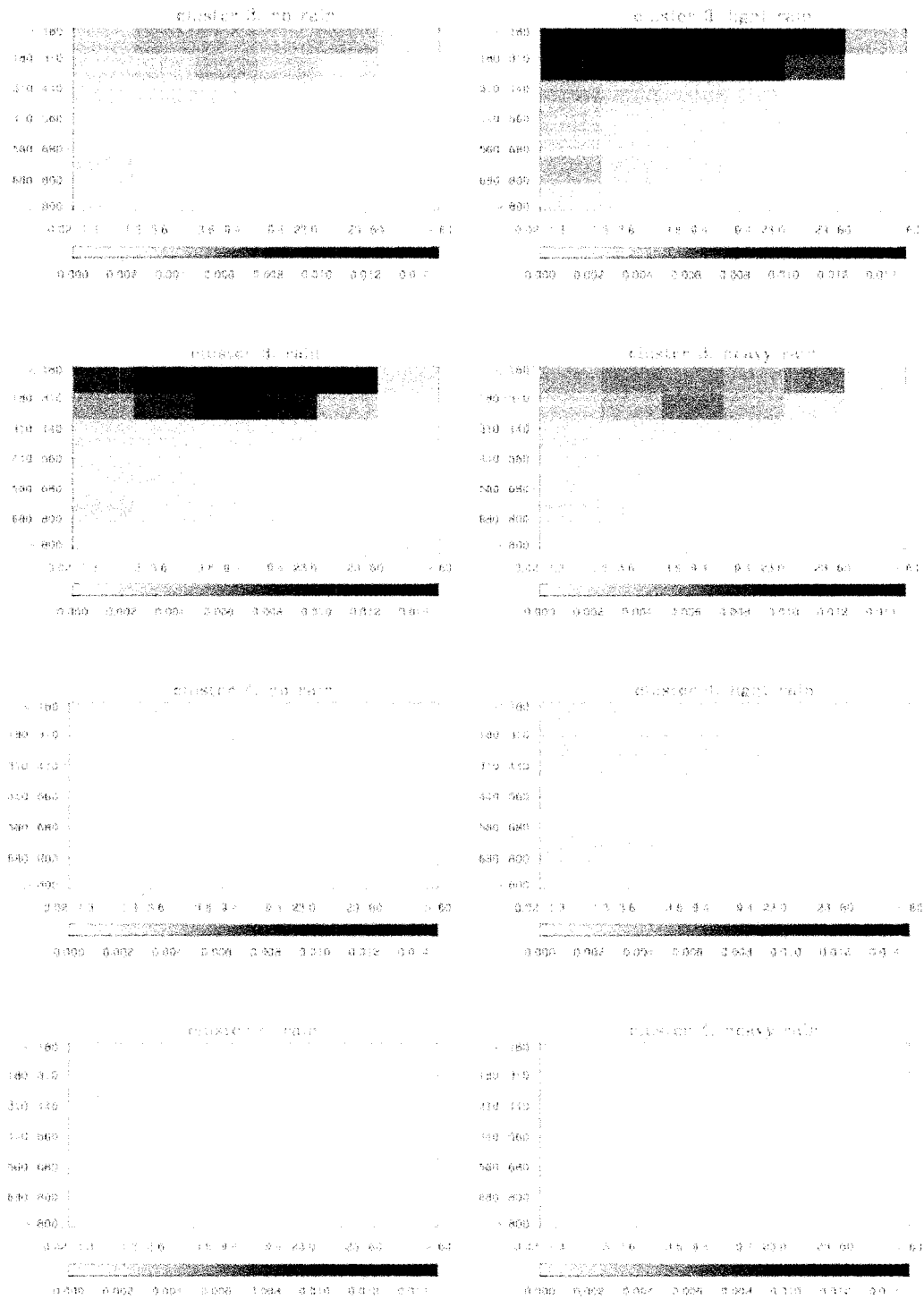


Figure 3.7b. Clusters 3 and 4 for k-means analysis with $k=6$. The upper four panels represent cluster 3, the lower four panels cluster 4

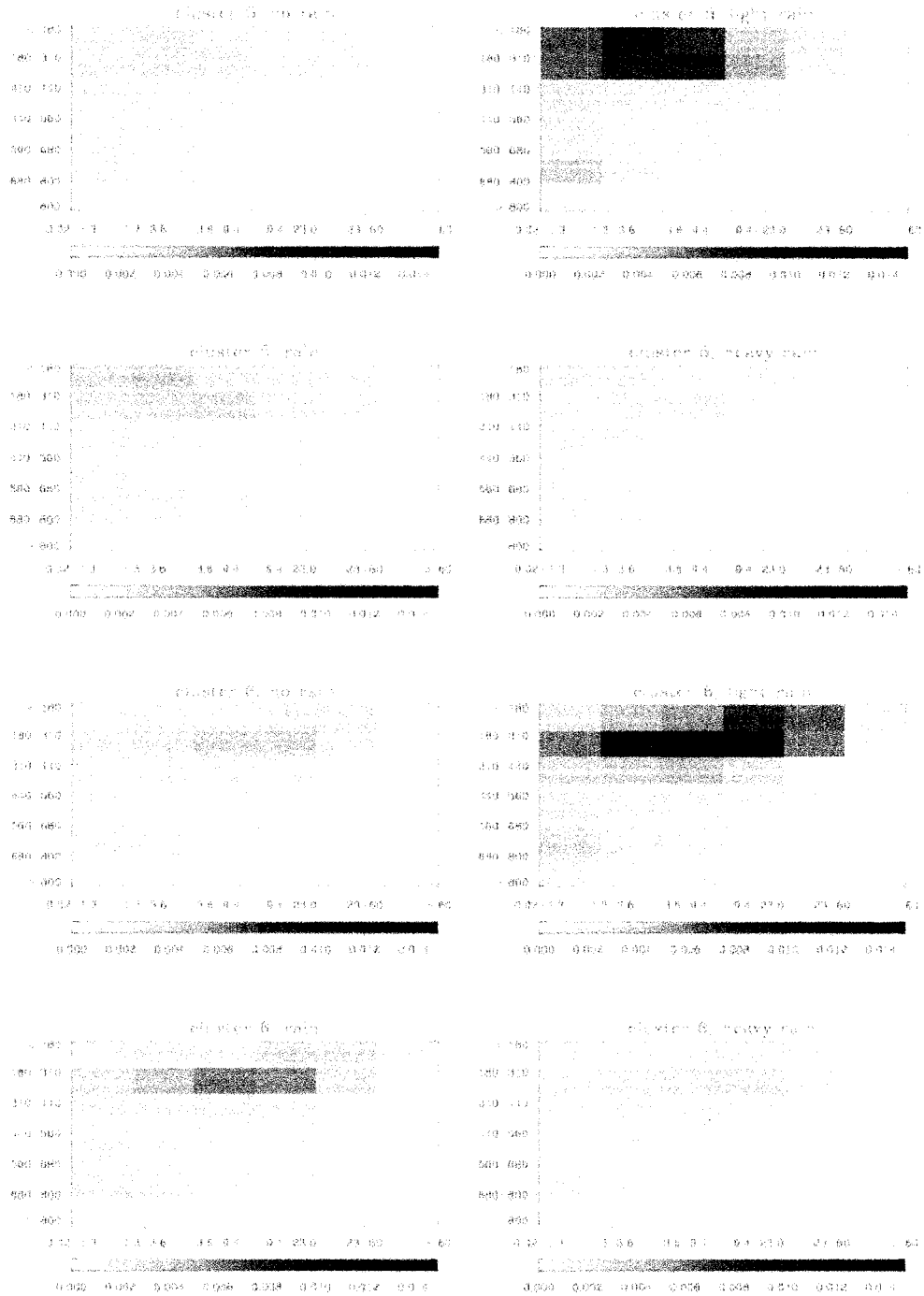


Figure 3.7c. Clusters 5 and 6 for k-means analysis with $k=6$. The upper four panels represent cluster 5, the lower four panels cluster 6.

3.4.5 Discussion on selection of k

We are left with two viable candidates for k – either $k=3$ or $k=4$ produce unique and significant clusters for our particular dataset. As is noted in JT03, choosing the largest valid option for k has the benefit of choosing the most number of ‘distinctly different’ clusters, and therefore provide the most granularity when discussing the properties of the domain as a function of cluster-derived properties. We therefore select $k=4$ as the best choice for continued analysis for the remainder of this work. Coincidentally, the 2D analysis also used $k=4$, but for very different reasons, although it should be noted that the general properties of the clusters found in the 3D analysis do bear some resemblance to those of the 2D analysis. We shall argue that the *differences* between the 2D and 3D analyses are of some interest, however.

Finally, it is worth mentioning that our selection of $k=4$ for this analysis should not be construed as a suggestion that the region of interest in the tropical West Pacific contains four, and only four regimes – had we used different variables, we certainly could have found other appropriate values of k . As stated in Chapter Two, it is, rather, our intent to ‘index’ the atmosphere by using important cloud properties of a region – not to classify the atmosphere *in toto* as a function of our regimes. We are mindful, therefore, that our results (which certainly are appropriate for our particular combination of data and region) cannot be extrapolated to other regions without re-performing the steps described in this chapter.

3.5 Discussion

If we interpret the clusters found in our 3D analysis in the same manner as JT03 interprets the results of their cluster analysis, then we conclude that the clusters shown in

Figure 3.4 represent the four dominant cloud regimes found in the region of interest as determined through our 3D combination of cloud-top pressure, cloud optical depth, and rainfall amount. Comparing the results of Figure 3.4 to those of Figure 2.5 (our 2D MODIS-derived clusters) we see several key similarities, along with some important differences. Both the 2D results (Figure 2.5) and the 3D results (Figure 3.4) contain what appear to be at least two convectively-active regimes, and at least one convectively suppressed regime, with identical convective regimes consisting of a ‘hot-tower’ like regime, and a likely organized deep-convective regime in both analyses. Furthermore, the identical convective regimes have similar relative frequencies of occurrence, especially when comparing the frequencies of occurrence of the organized deep convective clusters. The general similarities between the two analyses are comforting, at least in the sense that we have not gone too far astray from our original results with our new analysis technique.

Notable *differences* between the 2D- and 3D analyses are perhaps more interesting. One of the motivations of attempting this kind of analysis with a 3D dataset is a fair amount of ambiguity in the nature of moderately thick cirrus, perhaps above an underlying layer of congestus cloud, in the manner depicted in Figure 2.6. Indeed, the results of Figure 3.4 bear this out – the nature of the non-convective regimes found in our 3D analysis is notably different than those of the 2D analysis. The most immediately recognizable difference between the 2D non-convective and the 3D non-convective clusters is the presence of a weakly-convective regime consisting of moderately-thick congestus and cirrus clouds in our 3D analysis – this is certainly a departure from the 2D analysis, although hints of some convection do appear in the 2D-derived high-cloud

regime. Our 3D analysis, at first glance, appears to contain a convectively suppressed regime consisting of low height- and thickness clouds (perhaps cumulus or scattered stratocumulus) with a smattering of thin cirrus present as well. Overall cloud cover in this suppressed regime is low, suggesting the possibility for a large amount of shortwave surface heating. The next three clusters appear to vary by convective amount and strength – a marginally convective regime (that we just discussed) consisting of congestus and cirrus, with low- to moderate rainrates, followed by an isolated convective, and finally the ubiquitous deep-convective regime. These results echo other observations of tropical cloud construction more strongly than does the 2D analysis, and could be arranged to support several different theories regarding the possible progressive nature of tropical convection (e.g. Stephens et al. (2003)). In order to investigate these possibilities, however, we will need to first derive the average properties of each of our regimes, assess the progression of each regime as a function of time, and compare other observations of tropical variability with our results. We present this work in forthcoming chapters.

Chapter 4

Properties of the Tropical West Pacific from 3-Dimensional Cluster Analysis

4.1 Introduction

The central hypothesis of this research is that the current state of the tropical environment can be identified by a continuing progression of a few unique, self-similar cloud states, differentiable by a small set of cloud properties, and that through cluster analysis centered on these properties, we may more accurately and objectively investigate the evolution of the properties of the tropical atmosphere. Of special consideration are those properties that describe the radiative, hydrological, and thermodynamic states of the atmosphere – and perhaps more importantly, the interaction between these properties, and what we can determine about the progressive nature of the tropical system from those interactions.

Before developing these concepts further, one formal detail must be ironed out – we must make the connection between a cloud *cluster* and a cloud *regime*. What has been described so far have been cloud *clusters*, which is to say, the results of a cluster analysis describing the several statistically unique patterns in a dataset consisting of cloud observations. These are naturally-occurring patterns in nature and are unique in their occurrence (in other words, the patterns don't overlap when they occur.) This provides a convenient definition of a cloud *regime*: a naturally occurring and ubiquitous cloud population that manifests itself over a period of time in a real dataset. If we use the results of our cloud cluster analysis to identify sets of naturally occurring and ubiquitous cloud populations, then we have the mechanism to describe cloud *regimes* in the tropical

West Pacific. The key distinction is how each is applied – for our purposes, we will refer to a cloud *cluster* as specific to the particulars of cluster analysis, and will use cloud *regime* to refer to the occurrence of one of our cloud clusters in a real dataset.

We begin with a detailed description of each cloud regime based solely on the cloud properties of each regime (as determined from the MODIS and AMSR/E observations used to describe the regime), and then follow up with an examination of the progression in time of each regime. We then compare the results of our cloud regime analysis with external datasets (specifically, CERES OLR and flux data, and NCEP reanalysis products) to assess the germane atmospheric properties of each cloud regime, and the evolution with time of these properties. The results of this comparison are of particular use in examining contemporary theories of tropical evolution, which we will discuss in Chapter 5.

4.2 Description of 3D regimes

The individual cloud clusters have heretofore been referenced by their cluster numbers, e.g. Cluster 1, etc. These numbers are, of course, arbitrarily assigned – in moving forward to describe the properties of each regime (based on the occurrence of each cluster) we will want to assign a name for each regime that is descriptive of the overall properties of the regime. As we describe the individual properties of each cloud regime, as we do here, it is convenient to include the name we assign to each regime. We now present the four regimes of our 3D analysis, sorted roughly by increasing amount of convection in each regime.

4.2.1 Cluster 3 – Trade Cumulus and Cirrus regime

The first regime contains the fewest convective clouds and the least cloud cover of the four regimes, and is also the most commonly found cloud regime in the dataset. The 3D histogram representing this regime is provided as Figure 4.1. The relative frequency of occurrence (RFO) for this regime is 53.3% - just over half of the cloud scenes in our analysis fit into this regime. As discussed in Chapter 2, it is possible to integrate over the histogram bins to determine the total cloud cover (TCC) in the regime – for this regime, total cloud cover is a scant 4.799 percent. By way of comparison, TCCs for convectively suppressed regimes described in JT03 were upwards of 40%. (This difference is also reflected in the scaling of the figures.) A likely explanation for this discrepancy is found in the disparate nature of the datasets used: JT03 used ISCCP observations of 5km resolution, averaged over 3 hours – the presence of any cloud in the 5km box over the three-hour averaging period would register the entire pixel as cloudy. For our study, the 1km-resolution MODIS pixels are observed relatively instantaneously – for an equivalent 5km area in our study, only those pixels that actually contain cloud *at the time of overpass* are counted as cloudy pixels. For the extreme case of a single, 1km² cloud existing in a 25 km² area (the size of the 5km pixel) the MODIS observation would return a single, 1km² cloudy pixel, and 24 clear pixels, whereas the ISCCP observation would return a 25km² cloudy pixel, with the single cloudy area ‘smeared’ over the entire domain. The difference in perceived cloudiness is roughly an order of magnitude, which corresponds with the difference in TCC seen in our study. A second explanation is found in the disparate nature of the regimes themselves – in our analysis, we found two regimes

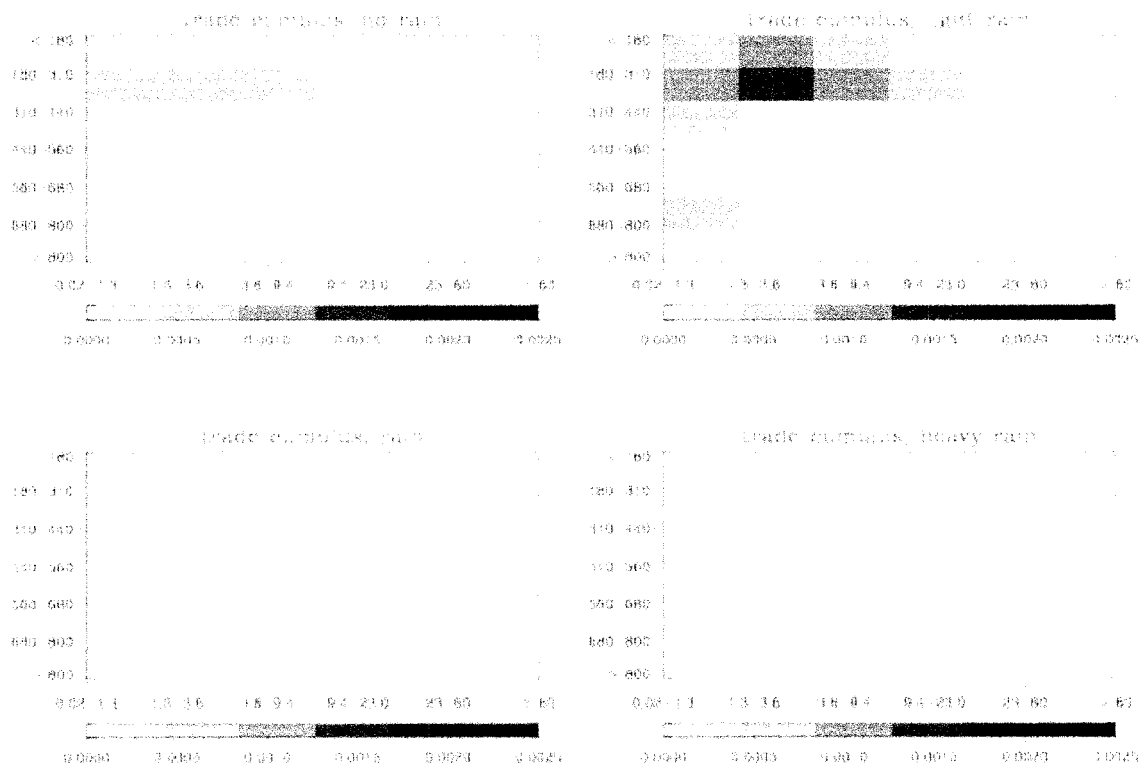


Figure 4.1. 3D histogram of the trade cumulus and cirrus regime. Each panel consists of a 2D histogram of cloud top height along the y -axis, and cloud optical depth along the x -axis, with a single panel representing the 2D histogram of the cloud population within a certain rainrate bin (described in Chapter 3.) The regime consists primarily of thin, upper-level clouds and low-topped (CTP > 800mb) thin clouds consistent with fair-weather cumulus clouds.

consisting of generally shallow cumuliform clouds with attendant populations of thin cirrus (clusters 3 and 1), but with marked difference in precipitation (and in total cloud cover.) It is possible that the suppressed low-cloud regime of JT03 (which did not use precipitation information to determine cloud cluster) contained sporadic convective clouds (which typically have higher cloud cover amounts, as we will see), artificially inflating the total cloud cover of the JT03 regime. A third explanation is simply that there were fewer clouds over the domain during the period of observation for this study than there were during the period used in JT03.

As shown in Figure 4.1, the trade cumulus and cirrus regime consists of relatively high (cloud tops between 310 and 180mb) and thin (optical depth < 3.6) clouds, with a secondary population of very thin (optical depth < 1.3) and low (cloud tops < 800 mb) clouds. The bulk of observed clouds contained rainrates less than 1.5mm/hr, although a scant few pixels had higher rainrates. The lack of significant convection in the regime (as would be expressed by high-thickness, high-topped clouds with significant amounts of moderate- to heavy rainfall) indicate the general stability of the regime – this agrees with the conventional understanding of the tropics as existing largely in a state of subsidence, populated with boundary-layer clouds generated by eddies. As such, we would expect of this regime generally high OLRs representative of the ocean surface, with attendant high surface downward shortwave flux, leading to warmer SSTs. The paucity of total cloud cover also suggests a relatively lower surface humidity, pointing to calmer surface winds. In section 4.4, we further test these hypotheses with datasets external to the cluster analysis.

4.2.2 Cluster 1 – ‘Enhanced’ Cumulus regime

This cloud regime is perhaps best viewed as an extension of sorts of the trade cumulus regime – the enhanced cumulus regime contains many clouds of a similar nature to the trade cumulus regime. Clouds of the regime are typically thicker, with optical depths typically less than 9.4 (with a small fraction in the range between 23 and 9.4.) Cloud tops for the enhanced cumulus regime fall into two categories; first, higher level clouds with cloud top heights between 180- and 310mb, divided between convective clouds (as observed through at least moderate rainrates) as well as a population of thicker cirriform clouds (with low or no rainfall) 1; and a second population of low-topped thin clouds similar to the trade cumulus found in the trade cumulus regime. As with the previous regime, the bulk of the cloud population exists at rainrates below 1.5 mm/hr, although there exists a significant fraction of upper-level, moderately thick clouds with rainrates between 1.5 and 5.1 mm/hr, as well as a tiny fraction of cloud pixels with heavier rainfall (rainrates exceeding 5.1 mm/hr). Total cloud cover for the regime is found to be 29.33%, and the RFO for the regime is found to be 21.9%, making it the second-most common regime. The histogram for this regime is presented as Figure 4.2.

When compared to the more convective regimes (which we will describe shortly), total cloud cover for the enhanced cumulus regime is relatively low – approximately seventy percent of the region remains cloud free, and a fair amount of the cloudy parts of the region contain low clouds that are more representative of a convectively-suppressed atmosphere. However, it is the presence of convection that marks this regime as unique from both the convectively suppressed regions and the other convective regions. As a prototypical ‘transitional’ regime, it is expected to have unique properties compared to

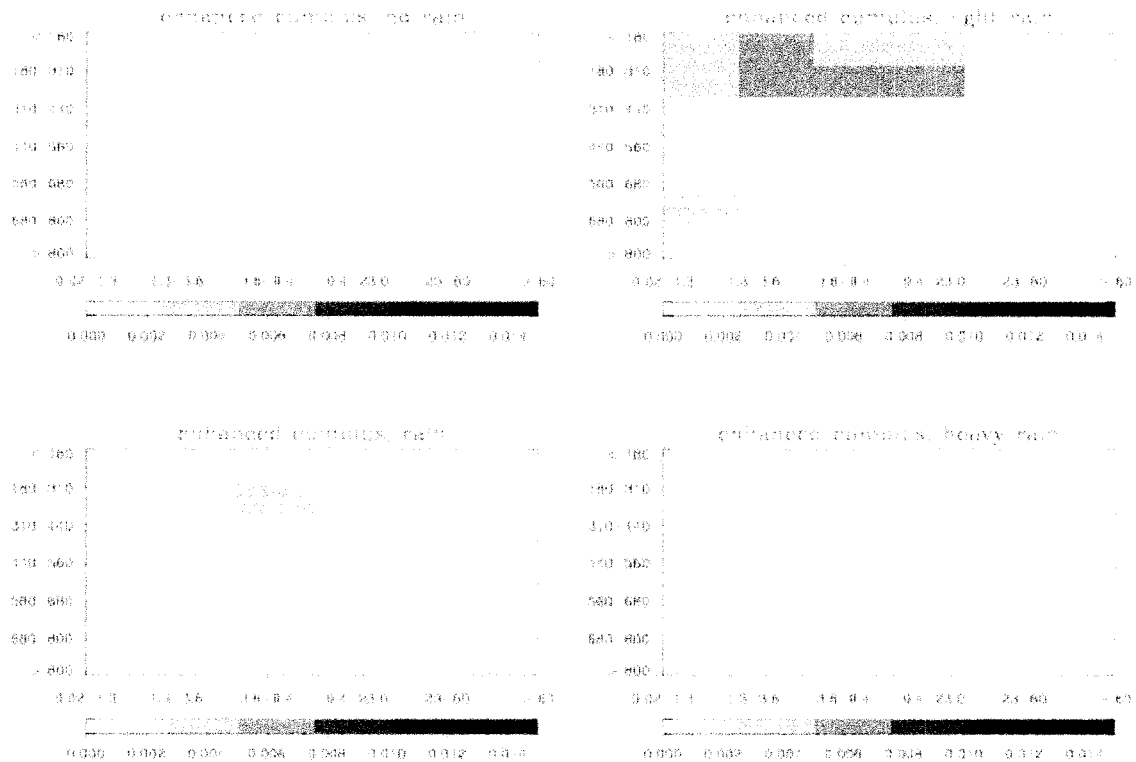


Figure 4.2. As in Figure 4.1, but for the enhanced cumulus regime. Note that the scaling factor for individual bin population has been increased to account for the greater total cloud cover of this regime compared to the regime of Figure 4.1.

the convectively suppressed regime. Although total cloud cover is still generally low, the presence of deeper cloud suggests greater surface convergence of moisture, as well as differences in the surface wind profile and mid-tropospheric moisture profiles. OLR for the regime should be slightly lower, with appropriately lower shortwave surface fluxes as well – we will return to test these hypotheses again in section 4.4.

4.2.3 Cluster 2 – Isolated Convection

Figure 4.3 depicts the histogram for the isolated convective cluster. The third-most common regime of the four (RFO = 17.52%), the isolated convective regime consists mainly of moderately thick, high-topped clouds with moderate- to heavy rainfall. Cloud-top pressures for the isolated convection regime are generally between 180-310mb, with a number of pixels exceeding the 180mb upper limit, representing deep convection. Cloud optical depths in the regime are between 3.6 and 23, with a small fraction exceeding optical depths of 23. The majority of the rainfall in the regime is again in the light rain category, but with a sizable minority in the 1.5-5.1mm/hr category, with roughly twice as much rainfall in this category as the enhanced cumulus regime. Another indicator of the convective nature of this regime is the total cloud cover, which is 59.17%, suggesting greater coverage of convective anvils which commonly occur with convection. Furthermore, there exist in this regime fewer low-topped, cumuliform cloud such as those found in the prior two regimes.

The two convective regimes found in this study share many common features with other convective regimes as found by other studies using cluster analysis (JT03, Rossow et al. 2005) which is perhaps not surprising – the unique nature of tropical convection makes the convective modes of the tropics some of the more easily-

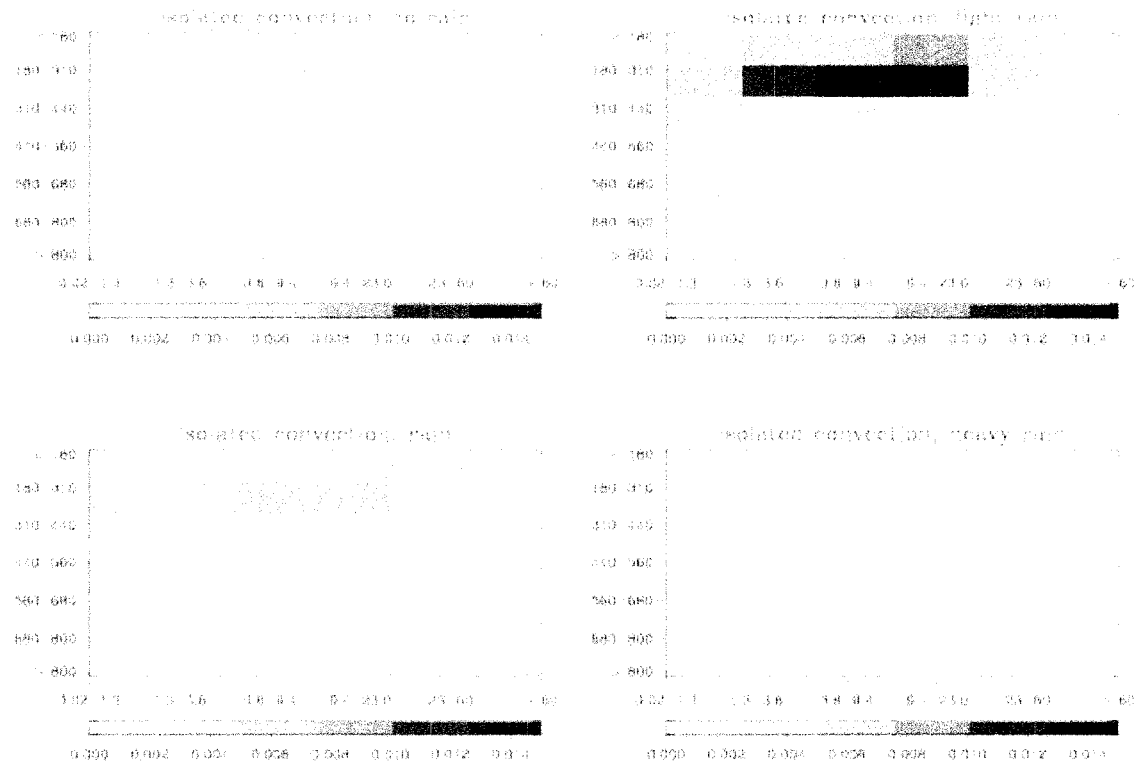


Figure 4.3. As in Figure 4.2, but for the isolated convective regime.

recognizable features of the system. A few key differences, however, offer some insight into the utility of including precipitation in the technique – for example, there is a markedly decreased presence of moderately thick cirrus in the isolated convective regime (compensated by an increased presence of the same in the enhanced cumulus regime). The shift of non-precipitating thicker cirrus (or perhaps, of non-precipitating multiple layers of cirrus and underlying cumuliform cloud) to a different regime addresses one of the more prominent ambiguity issues discussed in Chapter 2. (Whether or not this is the primary benefit of the inclusion of rainfall into the algorithm is an issue we discuss at the close of the chapter – there are other benefits of including rainfall to consider.)

We would expect even lower OLR values compared to the previous regimes and reduced downward shortwave surface fluxes in concert with this regime, along with higher total column water vapor amounts and a beginning of a decrease in SSTs due to shading from convective anvils and increased transport of energy from the ocean surface due to increased winds. The transport of water vapor from the near the ocean’s surface into the upper atmosphere would become increasingly important in this regime, as would the concurrent change in the vertical heating profile due to latent heat release in the atmosphere aloft due to the convection. (We speculate further on this issue in Chapter 6.)

4.2.4 Cluster 4 – Organized Deep Convection

The final regime described by our analysis is one of predominantly deep convection, with near-total cloud coverage and the most rainfall of all the regimes. This regime is also the least-present in the dataset, with a relative frequency of occurrence of only 7.3% (which happens to be the same frequency of occurrence found for the deep-

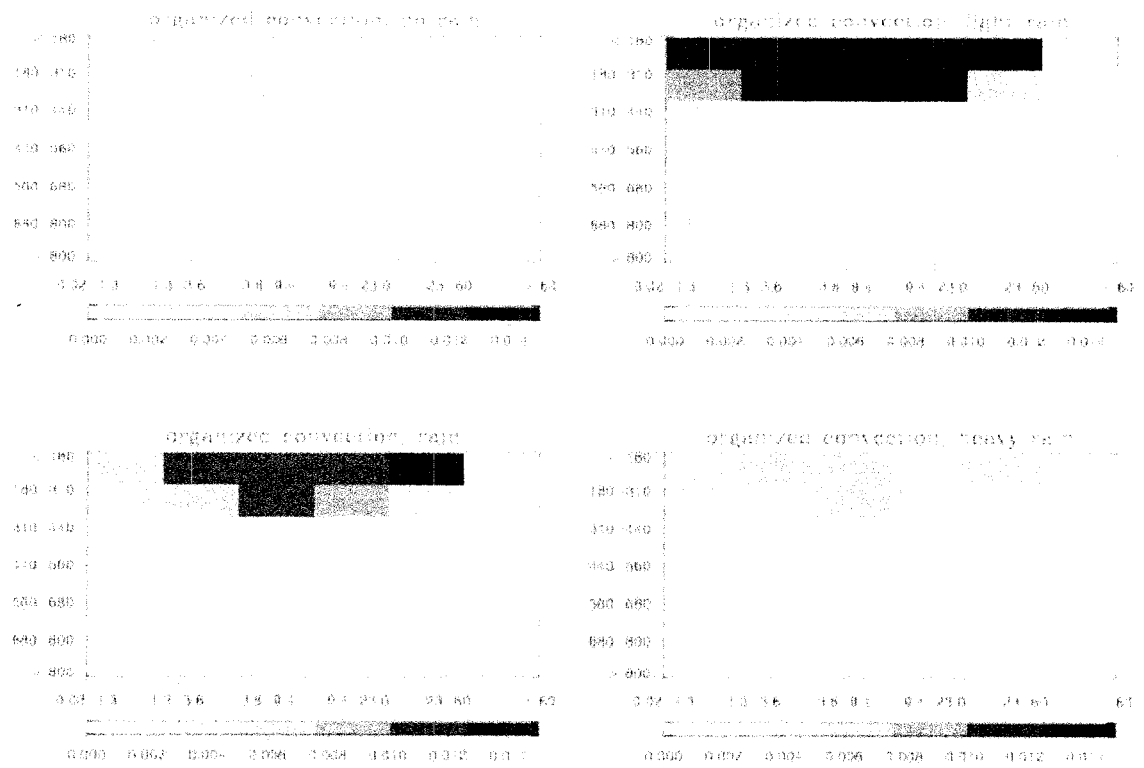


Figure 4.4. As in Figure 4.3, but for the organized deep convective regime.

convective regime in the 2D analysis) and is the most unchanged regime when compared to the original 2D analysis. The histogram representing the deep convective regime is presented as Figure 4.4. Total cloud cover for this regime was found to be 96.1%.

High, thick clouds dominate the regime - the bulk of the pixels have cloud tops above the 310mb level, and exhibit optical depths in excess of 23, although the bulk of pixels have optical depths between 1.3 (in the case of anvil cirrus) through 60 (in the case of convective cores.) The bulk of the raining pixels still inhabit the light rain category, with rainrates under 1.5 mm/hr, but this is concurrent with the highest concentration of rainfall between 1.5 and 5.1 mm/hr of all the regimes. Furthermore, the vast majority of all of the ‘heavy’ rainfall (characterized by rainfall in excess of 5.1 mm/hr) occurred in this regime – more pixels with these rainrates occur in this regime than in the other three regimes, combined.

The nature of organized convection in the tropical West Pacific ocean is a storied topic of the atmospheric sciences. Nakazawa (1988) drew on this history to develop his theory of tropical ‘super clusters’, the properties of which exhibit many of the same properties found in this analysis. The convection associated with the MJO also has the same general properties as those found by this study, and organized clusters of deep convection plays a central role in several dynamic theories of tropical convective interaction with the environment (Moncrieff (1992)).

4.3 Temporal characteristics of 3D regimes

One interesting feature of cluster analysis is the ability to go back to the original dataset of histograms and assess which histogram belongs to which regime, and thereby evaluate the evolution in time of each regime. In principle, one could do this from the

saved centroid-histogram distance information as the analysis is originally performed as well - however, there exists a great deal more ambiguity in cluster assignment in the initial stages of the analysis, making it a more prudent effort to perform after the clusters are already comfortably defined.

Having described the basic cloud properties of the regimes in section 4.2, it is useful to examine the temporal properties of each regime. As previously described, during the clustering procedure, each histogram is assigned to the ‘best’ cluster, and this assignment information describing which observations ‘belongs’ to which cluster is stored locally. For a dataset of M daily observations, we therefore have M assignments between one of k clusters. For each cluster, we can construct a vector of length M containing an arbitrary, non-zero value for days when the observed histogram was assigned to that particular cluster, and a value of zero for days when the observed histogram was assigned to a different cluster. We can then create a $k \times M$ array combining the vectors for all values of k - plotting the array for our results yields Figure 4.5, where the non-zero values (representing the presence of a particular cluster at a certain time) are plotted in dark blue. In the case of Figure 4.5, the arrangement of the array is ordered by convective activity, starting with the trade cumulus vector at the bottom, and ending with the organized convection vector at the top. Several general trends are immediately noticeable in the figure – a general sawtooth-shaped pattern (Figure 4.6) with five roughly-defined peaks from 1 July through 26 November 2002 suggest a cyclic pattern of convective development from trade cumulus through

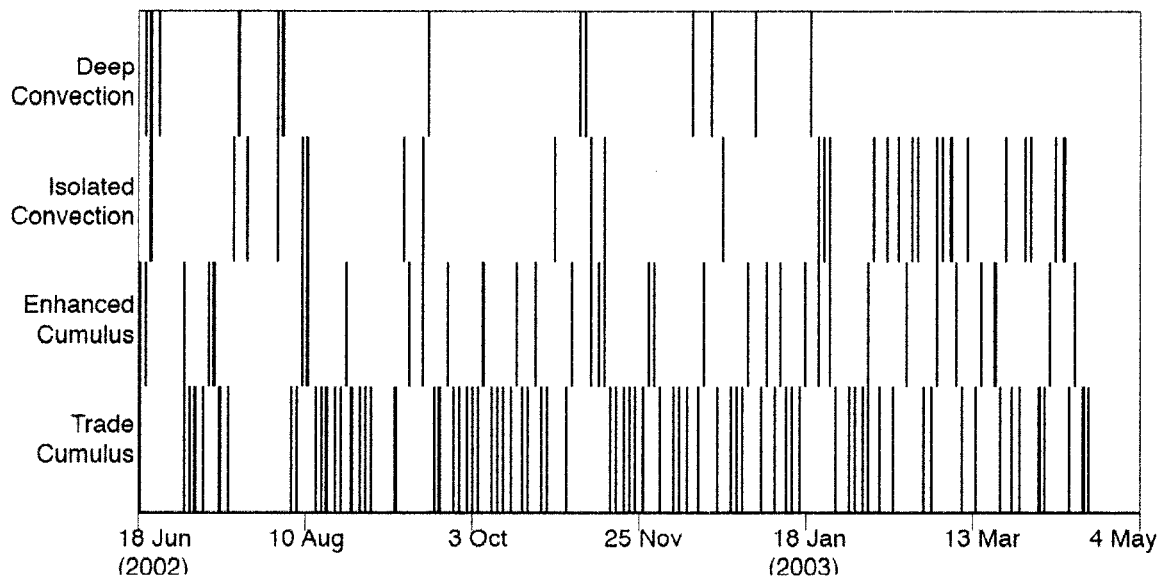


Figure 4.5. Evolution with time of each cloud regime, with dates for years 2002 and 2003 located along the x -axis, and individual regimes, in increasing order of convective activity, presented on the y -axis.

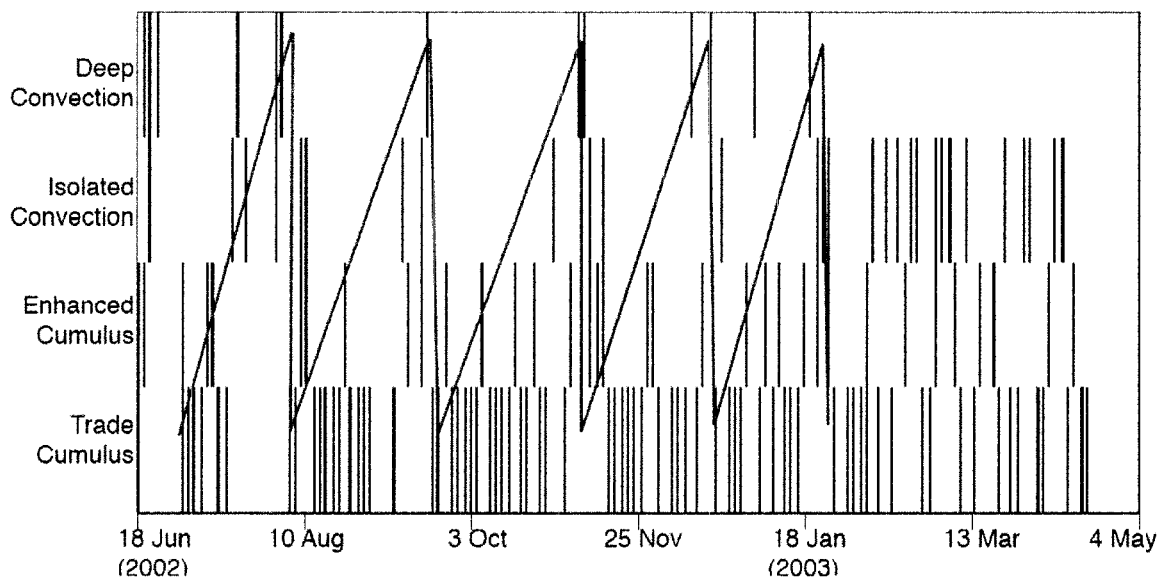


Figure 4.6. As in Figure 4.5, with cartoon overlay of generic sawtooth pattern with a period of around 40 days.

organized deep convective regimes. The period of these cycles is roughly 40 days, with the most clearly-defined cycles occurring in late fall through winter. After 18 January 2003, the pattern oscillates at a much higher frequency, with no occurrence of the deep-convective regime as the year progresses through May 2003.

The pattern described in Figure 4.5 and 4.6 is quite suggestive of the MJO, and indeed the pattern coincides with low OLR anomalies (not shown at present, we return to the topic shortly) generally associated with an active MJO. After the ‘MJO’ period ceases, the convective cycles are generally more random, although higher-frequency sawtooth cycles (terminating with the isolated convection regime) are seen from 30 January through 6 March 2003.

The presence of these possibly cyclic- and progressive patterns in cloud regime evolution is clearly worth further investigation. It is difficult to do so, however, with only the information provided in the datasets internal to the cluster observations. We move forward, therefore, by using our regime information, and specifically, the knowledge of the occurrence in time of each regime to compile a regime-sorted database of external observations from satellite and reanalysis data. Through the use of these regime-sorted compilation datasets, we can investigate more of the mechanisms that may govern the cyclic behavior observed.

4.4 External observations of 3D regimes

The various hypotheses addressing the organization of the tropical cloud cycle are legion, but most of the major theories involve some combination of a select set of atmospheric parameters combined in differing physical frameworks. One such example is the ‘humidistat’ hypothesis of Stephens et al. (2003) which posits a three-phase cycle of

the following: a *destabilization* phase marked by warming SSTs in concert with upper-atmosphere cooling that results in a moistened, unstable atmosphere; followed by a *convective* phase, initiated by an unspecified convective trigger (thought to be a consequence of a superimposed MJO dynamics profile which exerts sufficient control to initiate convection) marked by organized convection which results in strengthened surface winds, increased surface evaporation with decreased SSTs and a moistening and warming of the upper troposphere; and finally, a *restoring* phase which occurs after the MJO-like convective trigger has moved out of the region, leaving in its wake a large amount of residual upper-level cloud which (through LW heating effects aloft and SW cooling effects at the surface) restores the atmospheric stability, and eventually dissipates, returning the system to the beginning. In this theory, as with other theories regarding convective control of the tropics, the key interactions are those between surface parameters (SST, evaporation/humidity/surface winds, and surface radiative fluxes) and free- and upper-tropospheric parameters (LW heating, humidity, precipitation (with attendant latent heating) and vertical temperature profiles) as governed by the cloud processes within the system. Since our goal is to examine potential mechanisms that govern the tropical cycle of cloudiness, our choice of external datasets to use in concert with our cloud regime information is dictated by these relevant variables.

Thus far, we have used a combination of CERES instrument data to observe radiative properties of the atmosphere, and NCEP reanalysis products to observe other properties. In coming chapters, we will undergo examination of the region using regime analysis including CloudSat observations of precipitation (Haynes et al., in preparation.) This research is addressed in Chapters 5 and 6.

4.4.1 Radiative characteristics

With respect to the physical models we intend to investigate with our regime analysis, there are three key parameters that are reliable measurements of the radiative state of the earth-atmosphere system that are also easy to co-locate with the MODIS and AMSR/E observations used to create our cloud regimes. Pending further discussion, the three parameters that we use to infer atmospheric radiative properties are – OLR, surface downwelling longwave flux, and surface downwelling shortwave flux. Naturally, the latter two properties are more difficult to retrieve for cloud scenes exhibiting a high amount of cloud cover (namely, our organized deep convection regime) - in the case of highly cloudy scenes, surface flux properties are generally computed using radiative models (e.g. Li et al. (1993), Li et al. (2002)).

Our motivation for choosing these three parameters is based primarily on our objective of comparing the radiative properties of our cloud regimes with the hypothesized radiative properties of theorized cloud systems (e.g. the cycle of destabilization, convection, and restoration described in Stephens et al. (2003)). OLR, while not specifically tied to the humidistat hypothesis, is a useful parameter to observe, especially when comparing cloud regime tendency to phases of tropical oscillations such as the MJO which are often indexed by OLR anomalies. Surface downwelling fluxes of shortwave and longwave radiation are used as proxies for the magnitude of surface heating by insolation (as well as likely cooling of the upper atmosphere) and of upper-atmosphere heating due to cloud greenhouse effects. It is our hope that by obtaining domain- and regime-averaged properties for these parameters, we should gain insight into

(at least the radiative) aspects of several current hypotheses regarding convective regulation in the TWP.

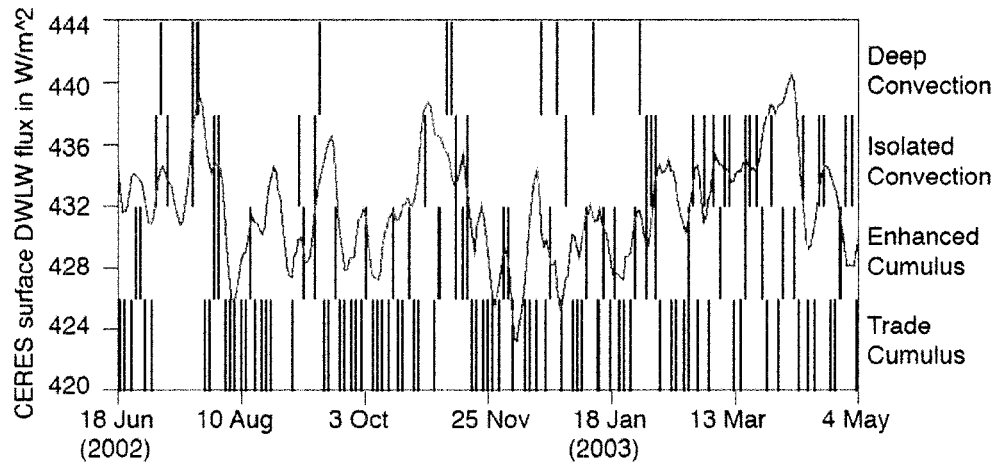
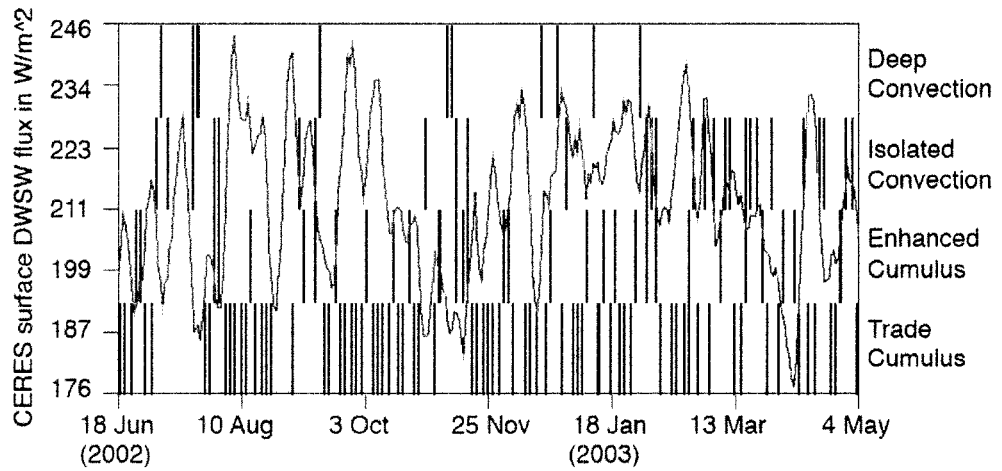
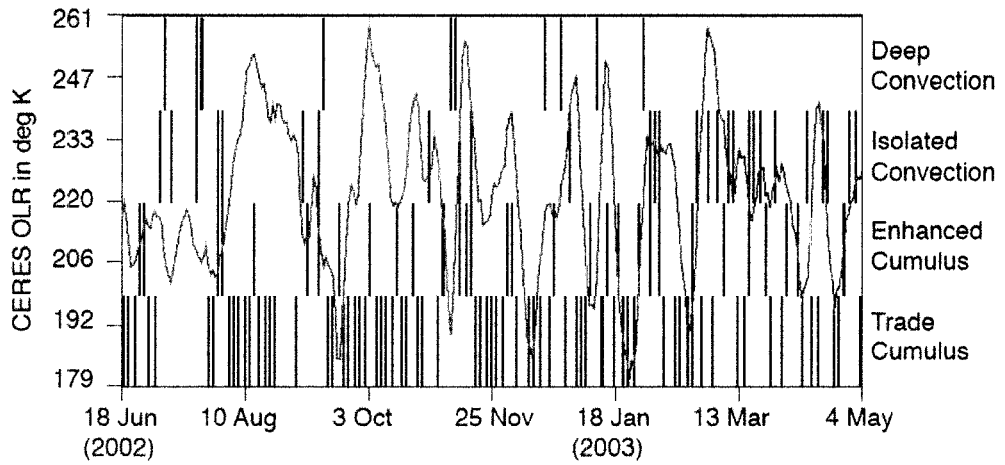
A convenient source of OLR and surface flux information comes from the CERES instrument aboard the Aqua satellite, which also carries the MODIS and AMSR/E instruments used to create our cloud regime analysis. This greatly simplifies the process of co-locating our radiative properties with our regime-based cloud scenes. Two CERES instruments (Wielicki et al., (1996)) aboard Aqua employ three-channel scanning broadband radiometers to observe a variety of scientific parameters, including ERBE-like fluxes and observations of TOA/surface fluxes. For this comparison, we use the Single Scanner Footprint (SSF) CERES product, using observations from both the FM3 and FM4 instruments aboard Aqua. OLR and surface downwelling SW and LW fluxes are both available from the product, with the downwelling surface flux parameters computed using a radiative transfer model described in Li et al. (2002). CERES observations are collected for the same region described in Chapter 3 (150° - 160° E, 0° - 10° N) and averaged over the domain to compute a daily average of these properties. Figures 4.7, 4.8, and 4.9 depict overlays of daily, domain-averaged values of OLR, downwelling SW flux at the surface, and downwelling LW flux at the surface (smoothed with a running 5-day filter for clarity) superimposed on the regime trend plot (as previously described and introduced as Figure 4.5.)

Figure 4.7 depicts the relationship between CERES OLR and the regime trend. Perhaps the most noticeable trend in the OLR/regime relationship is that during quiescent periods of convection, OLR values are higher, with lower, colder OLR values (more representative of deep convection) occurring in concert with the convective regimes.

This is especially true during the first two-thirds of the time period, which is marked by periodic cycles of convection with a frequency of approximately 40 days. Mean OLR values, by regime are presented (along with the flux data) in Table 4.1.

Figure 4.8 depicts the relationship between downwelling SW flux at the surface as observed from CERES and the regime trend. A generally similar pattern as in Figure 4.7 is found – the lower cloud cover amount of the non-convective cluster fosters greater insolation, and greater downwelling shortwave flux. It is expected that this increased insolation would lead to increased surface heating, and warmer SSTs during non-convective periods. (We investigate this in the next section.) As convection in the domain becomes more predominant, insolation decreases as cloud cover increases. Mean values for downwelling shortwave flux at the surface for each regime are shown in Table 4.1.

Figure 4.9 presents the relationship between CERES-observed downwelling longwave flux at the surface versus the regime trend. As the domain becomes increasingly convective, surface downwelling LW flux increases as well, which agrees with our expectations that increased amounts of thicker cloud and increased water vapor (especially with increasing clouds amounts located higher in the atmosphere) would lead to increased trapping of longwave radiation, thereby warming the upper troposphere. Combined with the decreased insolation, we would expect a general decrease in SST and a cooling of the ocean surface combined with this upper-tropospheric warming. Again, we investigate this further in the next section.



Figures 4.7, 4.8, and 4.9 (from top to bottom) Cluster trend versus CERES observed OLR, downwelling SW flux at the surface, and downwelling LW flux at the surface, respectively. CERES observations shown in predominantly orange, regime blocks in dark blue.

Table 4.1. CERES observed radiative properties by regime type

	Trade Cumulus	Enhanced Cumulus	Isolated Convection	Organized Convection
OLR	257.9 W/m ²	242.1 W/m ²	225.5 W/m ²	192.6 W/m ²
Downwelling surface SW flux	211.6 W/m ²	207.5 W/m ²	204.1 W/m ²	195.9 W/m ²
Downwelling surface LW flux	418.0 W/m ²	433.1 W/m ²	433.0 W/m ²	435.3 W/m ²

The results shown here are promising – the radiative characteristics of the regimes as derived from our 3D cluster analysis generally agree with our expectations of what the radiative properties for regimes consisting of the kinds of clouds that we see in our regimes. How these radiative parameters interact with the other atmospheric parameters that are thought to govern the tropical cloudy cycle will determine how closely the regimes derived from our analysis agree with contemporary theory.

4.4.2 Thermodynamic characteristics

Assessing the thermodynamic properties of the atmosphere is simultaneously the most important and the most difficult task to accomplish over the tropical oceans. Comprehensive data collection sites in the tropical West Pacific such as the ARM facilities located on Nauru and Manus islands and in the Australian coastal city of Darwin are typically well-equipped, but provide only point-specific information. Using point-specific datasets in concert with cluster analysis requires a resampling of the cluster domain, as is done in Jakob et al. (2005.) Estimating the total thermodynamic domain properties of cloud regimes requires a data source that covers the entire domain, which at this time, does not exist for the tropical West Pacific.

We may make some progress, however, by using reanalysis data. The NCEP reanalysis product, described in Kalnay et al. (1996) uses forecast model analysis updated every 12 model hours with observations from the region. The forecast model ingests the few observations available, along with relevant satellite retrievals, and forecasts the domain properties using a standard forecast model approach. After a 12-hour forecast run, the model solution is compared to the observations of the domain at the end of the period, and the model is modified such that the model predictions match the observations. The cycle is then repeated, continually modifying the model solution to agree with the observations, while filling in the details for regions where observations are not possible. The result is a high temporally- and spatially-covered dataset that is accurate for analytical use, within the limits of the model used to generate the dataset.

We can use the domain-averaged properties from reanalysis products such as the NCEP reanalysis project to assess some of the more difficultly retrieved thermodynamic parameters for our domain. Among these parameters that we investigate are: 850mb zonal wind, SST, total column water vapor, vertical velocity profiles, and atmospheric stability, represented by the computed lifted index in the reanalysis product. For each variable, we obtain a daily domain-averaged value over our region of interest from the 12-hourly averaged reanalysis product, giving us a half-daily domain average to compare with our regime properties. We discuss each variable separately.

4.4.2.a 850mb Zonal Wind

A depiction of the mean 850mb zonal wind plotted against the regime trend (in the manner of Figures 4.7-4.9) is provided as Figure 4.10. Zonal wind at this level is often associated with the onset of deep convection (often called ‘westerly wind bursts’,

referenced in Johnson and Lin (1997)) and is suspected to be a mechanism of low-level convergence and enhanced surface evaporation. As such, the low-level zonal wind function as a convective triggering mechanism, and is suspected to be driven by the large-scale dynamics in the tropics.

Although the reanalysis results of Figure 4.10 are not as reliable as actual sounding observations, we may detect some of the important trends using the reanalysis data. One general trend we see is a stronger pattern of easterlies over the winter months, with departures to westerlies (or at least a significant weakening of the prevailing easterlies) in advance of deep convection. The summer months on either end of the figure experience more unsettled wind conditions. With regard to the 850mb wind properties by regime, 850 mb winds are milder and easterly (mean value of -0.149 m/sec) in the trade cumulus regime, with a shift towards stronger westerlies in the enhanced cumulus and isolated convection regimes (mean values of 0.913 and 1.242 m/sec, respectively). Winds are still westerly, but much weaker in the deep convective regime, with a mean value of 0.241 m/sec.

4.4.2.b SST

The domain-averaged SST anomaly for the period is plotted against the regime trend in Figure 4.11. In nearly every mechanistic theory of tropical convective regulation, increases in SST during cloud-free periods, due to increased insolation, are predicted in advance of convective triggering (e.g. Stephens et al. (2003)). In our analysis, we observed increased insolation (in the form of heightened surface downwelling SW flux) during non-convective periods; as such, we would likewise expect increasing SSTs during these periods. This is confirmed, as is shown in Figure 4.11 –

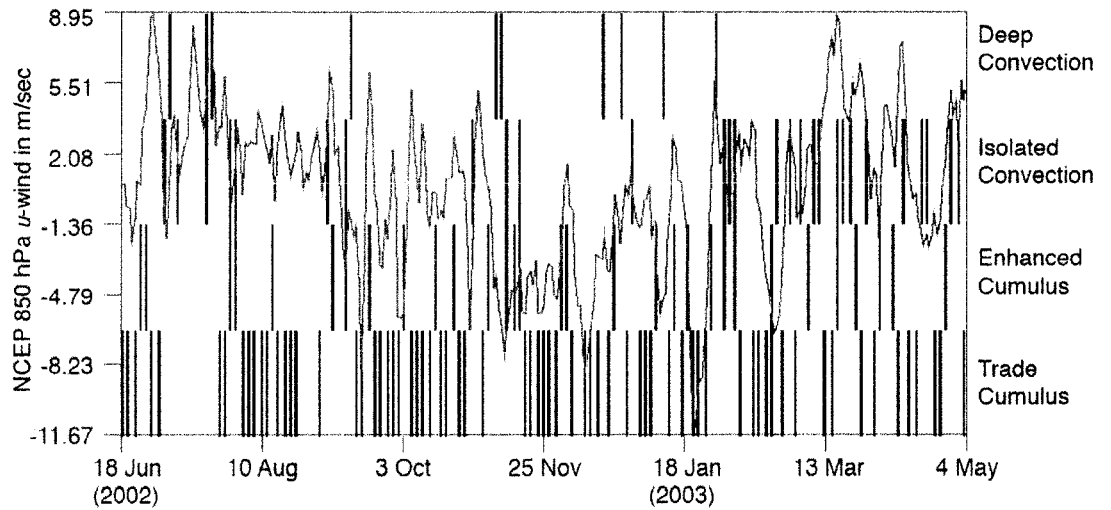


Figure 4.10. Trend of 850mb zonal wind velocity from NCEP reanalysis versus regime trend

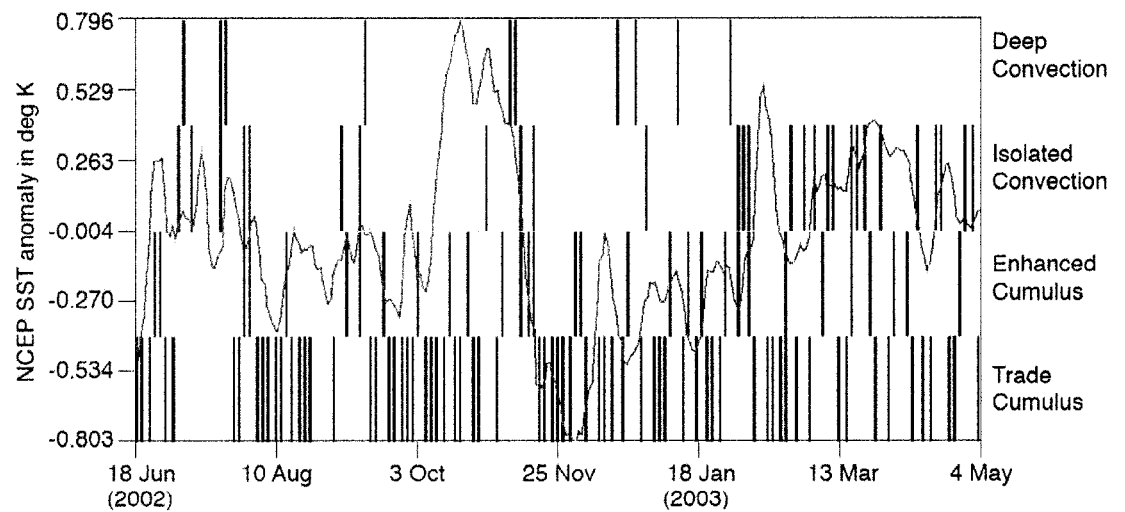


Figure 4.11. As in Figure 4.11, but for NCEP SST versus regime trend

SSTs continue to increase during the trade cumulus and enhanced cumulus regimes (on the order of ~ 0.7 K), then experience a marked decrease in temperature during the convective regimes.

4.4.2.c TCWV

The amount of column water vapor is very much tied to the amount of convection in the atmosphere – the ability of the atmosphere to transport latent heat, especially into the middle troposphere, where it can have a marked feedback effect on convectively-forced waves in the tropics, is tied to the amount of moisture evaporated. Figure 4.12 presents the total column water vapor (TCWV) for the region compared to the regime trend. As is seen in Figure 4.12, increases in convective activity are commensurate with increases in total column water vapor. Mean TCWV values, by regime, are as follows: for the trade cumulus regime, 4.97 cm, for the enhanced cumulus regime, 5.16 cm, for the isolated convection regime, 5.37 cm, and for the deep convective regime, 5.81 cm.

Of equal importance to the amount of water each regime possesses is the altitude at which said water exists. Regime-averaged vertical profiles of water vapor (in terms of the specific humidity q , and expressed as a difference between the trade cumulus regime and each of the enhanced cumulus, isolated convection, and deep convective regimes) are presented as Figure 4.13. As is seen in Figure 4.13, surface values of water vapor increase with convective activity for all four regimes – an increase in mid- and upper-tropospheric water vapor is also exhibited by the more convective regimes (especially in the deep convective regime).

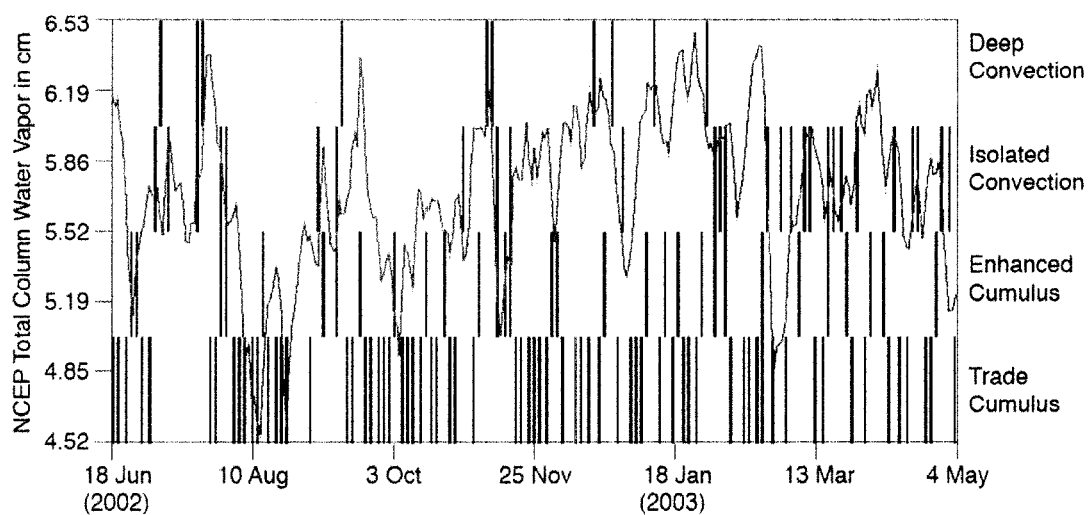


Figure 4.12. As in Figure 4.11, but for NCEP TCWV

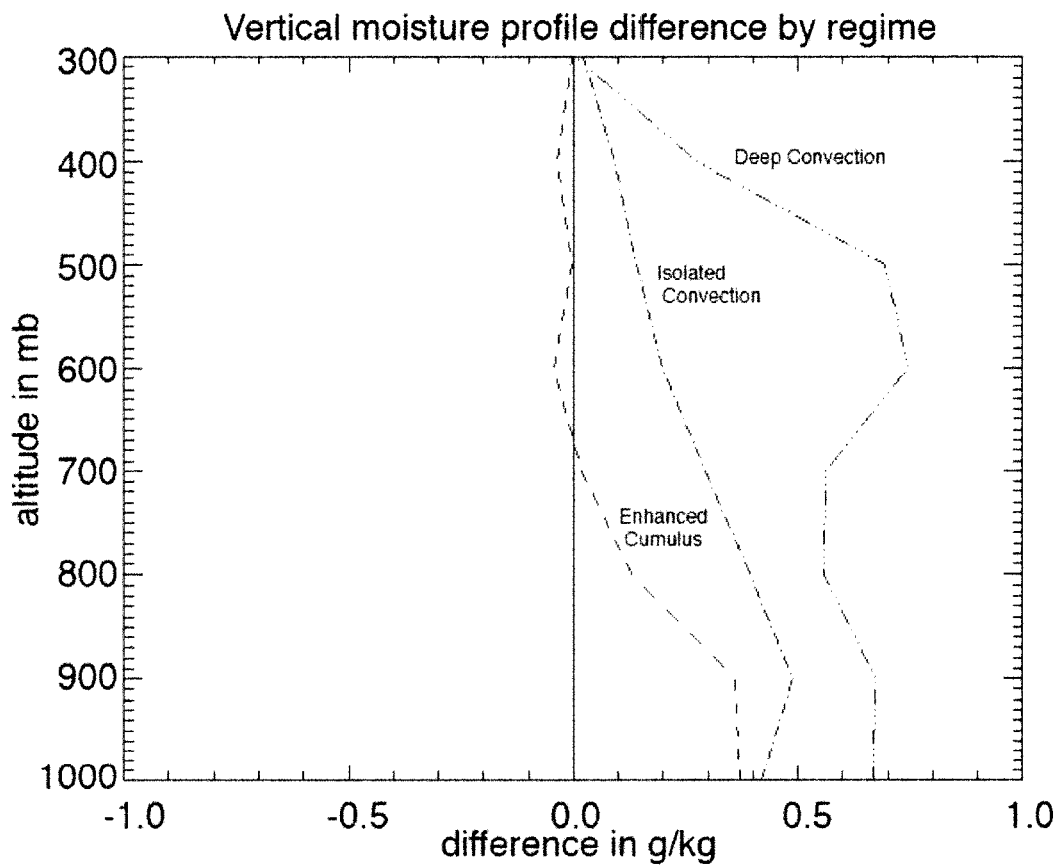


Figure 4.13. Regime-averaged vertical profiles of specific humidity for the enhanced cumulus, isolated convection, and deep convective regimes, compared to the trade cumulus regime.

This would lead to an increase in mid- and upper-tropospheric warming due to the retaining of LW energy (agreeing with the LW flux results described previously), as well as a marked increase in mid-tropospheric latent heating – both of which will contribute to modification of tropospheric stability, as we will discuss in later chapters.

4.4.2.d Lifted Index

Assessing the vertical stability of the atmosphere in the absence of regular sounding data is best left to model analysis such as found in the NCEP reanalysis product. One of the more useful products is the domain-computed lifted index, defined as the difference between the 500mb environmental temperature and the temperature of a surface parcel adiabatically lifted to the 500mb level. In an unstable atmosphere, the temperature of the environment would be cooler than the temperature of a lifted parcel, and the difference between the environmental and parcel temperature (defined as the lifted index) would be negative. Conversely, in a stable environment, the environment is warmer than a lifted parcel, and the lifted index would be positive. The value of the lifted index, and to a lesser degree, the magnitude, are therefore useful indicators of atmospheric stability. A plot of NCEP-derived lifted index versus regime trend is presented as Figure 4.14.

The results shown in Figure 4.14 are particularly interesting – we see marked instability primarily in the trade cumulus cluster, but less so in the enhanced cumulus

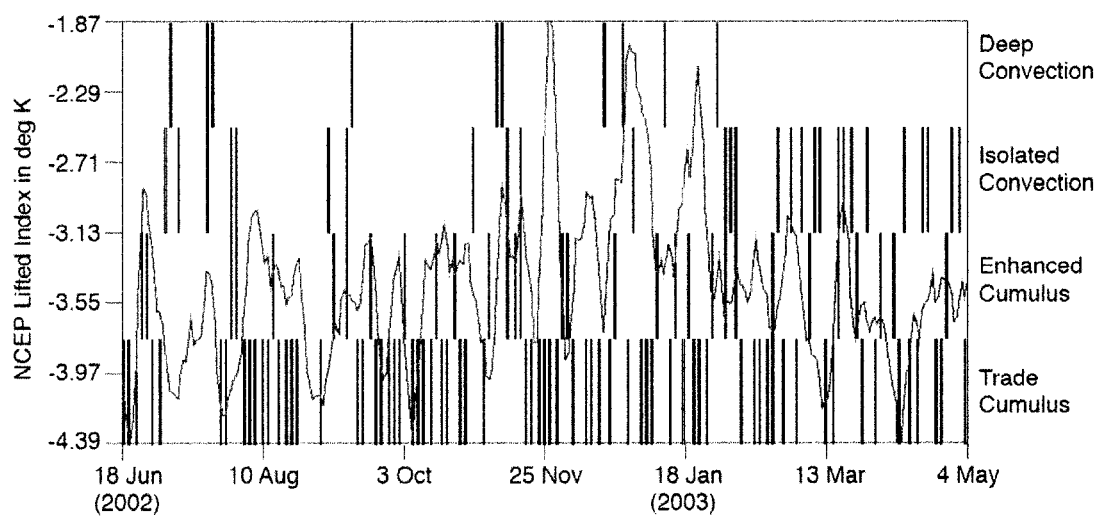


Figure 4.14. As in Figure 4.12, but for NCEP-derived lifted index.

cluster. Convective regimes tend to work towards neutral atmospheric stability, agreeing with the general precepts of radiative-convective equilibrium (e.g. Manabe and Strickler, 1964.) Destabilization of the tropical tropopause appears to be most prominent during the convectively-suppressed regimes, with restoring of atmospheric stability (presumably due to heating aloft due to latent heat and LW effects, concurrent with surface cooling due to anvil shading) occurring primarily during the deep convective regime.

4.4.2.e Vertical Velocity

The vertical motion of the tropical atmosphere is typically described as being in a state of general descent, save for marked regions of enhanced convection, which provides mass flux to the upper atmosphere, as well as the circulation driving the mass descent over the bulk of the tropical atmosphere. As such, diagnosing vertical motion is of interest when examining the tendencies toward convective activity of a particular region in the tropics. Unfortunately, computing vertical motion, along with the convergence and divergence associated with these motions, is notoriously difficult. Mapes and Houze (1995) were successful in using radar observations of tropical clouds, along with a model of diabatic divergence, to compute vertical motion in the atmosphere – they then used this model to compute atmospheric heating profiles as a function of ascent in tropical cloud systems. This seminal study provided a great advancement in our understanding of the effects of clouds on tropical dynamic systems.

Computation of vertical motion in *non*-cloudy regions is much more difficult – especially as the bulk of the tropics exists in the regions of large-scale descent at any given time. Here, reanalysis products are of limited value – the dynamics governing the patterns of atmospheric ascent and descent are poorly understood, and, are therefore

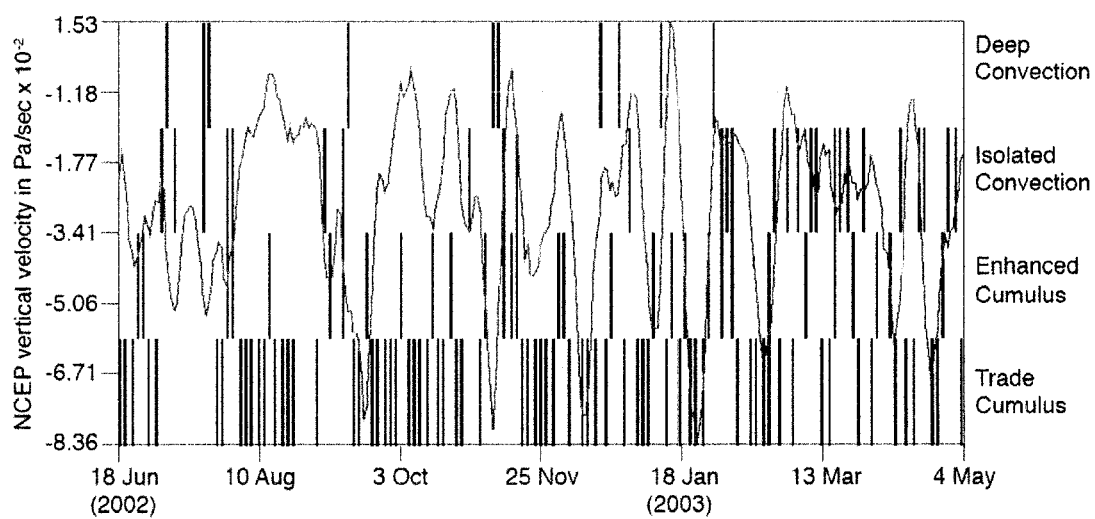


Figure 4.15. As in Figure 4.12, but for NCEP-derived vertical velocity.

represented in forecast models by relatively crude parameterizations. Furthermore, the dearth of observations of atmospheric ascent and descent with which to compare the modeled patterns of ascent and descent make any attempt to improve the model's representation of vertical motion problematic. As such, there exists no real dataset that definitively defines the mean properties this important piece of the atmospheric puzzle.

At best, what may be determined from our rudimentary ability to diagnose tropical vertical motion are trends – we cannot hope to find physical verity in the magnitudes (or indeed, even the signs) of vertical motion as diagnosed through reanalysis. Figure 4.15 presents the NCEP-diagnosed vertical motion at 500mb versus regime trend.

The most obvious result seen in Figure 4.15 is that vertical motion (here expressed in Pascals/second, with negative values indicating upward motion) very rarely attains positive values, even for the convectively-suppressed regime. Since the vertical velocities of Figure 4.15 are domain-averaged over a relatively large domain, we would otherwise surmise that the areally-small amounts of upward vertical motion due to boundary-layer eddies in the trade cumulus regime would be averaged out by the more dominant large-scale descent almost certainly present in the regime; yet the overall results indicate net *upward* motion over the regime. Given the growing instability in this regime, and especially given that the presence of this particular regime is proportional to the presence of observed trade cumulus in genuinely convectively suppressed atmospheres, we conclude either that the discrepancy must lie in the model's representation of vertical motion, or in some mismatch in location between observed convection and the manner in which it is represented in the reanalysis. We are not alone

in asserting this – several other studies (primarily Jakob et al. (2005) discuss in more detail the great difficulties faced in computing vertical motion in a model framework with little or no observations to confirm the model results.

This result has relevance to several recent climate diagnostic studies (Bony and Duvel (1994), Allan and Soden (2007), Back and Bretherton (2006)), which use reanalysis-derived vertical motion as a metric to diagnose several properties of cloud-climate interaction. The results of this study, which seem to indicate that our expectations of convective vertical motion are not necessarily well-represented in reanalysis products suggest that such studies using reanalysis as a diagnostic tool should be careful to relate their use of this parameter with other observations that are more readily tied to convection – Back and Bretherton (2006), for example, combine TRMM rainrate observations with reanalysis vertical motion in their assessment.

While the overall *trend* of vertical motion by regime, does fit with our basic expectations that convective regimes should have increased upward vertical motion than their convectively-suppressed siblings, we are left with the inescapable conclusion that, at least for this analysis, there appears to be a nontrivial discrepancy between reanalysis-derived vertical motion and our expectations based on convective activity sorted through cluster analysis. The mean values of vertical velocity, by regime, are found to be: -0.028 Pa/sec for the trade cumulus regime, -0.034 Pa/sec for the enhanced cumulus regime, -0.066 Pa/sec for the isolated convection regime, and -0.067 Pa/sec for the deep convective regime. We would expect better results from comparison between this particular parameter with our regime analysis as more sophisticated techniques for retrieving large-scale atmospheric motion become available.

4.5 Summary of Results and Discussion

What we have discussed in this chapter is the properties of each of our tropical cloud regimes – to a greater extent, we have derived the occurrence in time of each cloud regime, along with important properties of each regime that will allow us to compare each regime (with their important properties) to the hypothesized progression of the tropical hydrologic cycle as seen through several contemporary theories. As we recall from Figure 4.5, there occurs a general trend of progression from trade cumulus through enhanced cumulus, to isolated convection and then deep convective cloud trend.

Our most commonplace regime is the trade cumulus regime, consisting of a low-total cloud cover admixture of high, thin cirrus mixed with low-topped, low optical depth cloud, which exhibits high OLR values with high insolation, low LW trapping, warming SSTs, increased zonal wind, destabilization aloft, and likely descending motion (although the latter was not definitively proscribed.) The next-most prevalent regime is the enhanced cumulus regime, consisting of greater coverage of upper-level cirrus and slightly thicker low cloud, and exhibiting an increase in upward motion, maximum SSTs, increased instability, and an increase in water vapor uptake. The isolated convection regime consisted of moderate cloud cover of thick, high convective clouds with moderate rainfall, marked by an increase in LW trapping and a decrease in OLR and SW heating at the surface, along with a marked increase in precipitation, column water vapor (along with mid-tropospheric moisture), and vertical motion. The final, and least prevalent cloud regime consisted of near-total cloud cover of deep, very high convective clouds with moderate- to heavy rainfall, a maximum of LW heating concurrent with a minimum

of OLR and SW heating at the surface, with a maximum of both column- and mid-tropospheric water vapor, and marked stabilization throughout the regime.

Regime-averaged precipitation may be trivially computed by summing over the raining pixels in the histograms – performing this analysis yields regime-averaged rainrates for the trade cumulus, enhanced cumulus, isolated convection, and organized convection regimes, respectively, of 0.52 mm/hr, 0.61 mm/hr, 0.67 mm/hr, and 0.91 mm/hr. Maximum rainrates for the regimes top out at around 21 mm/day, which is a reasonable value compared with other AMSR-E studies of rainrate.

Several parallels may be drawn between the regimes found in this study and hypothesized cloud populations in contemporary theories. A notable example is the similarity between the cycle of convective growth in our regimes and the natural cycle of the MJO as seen during the latter part of 2002. In order to further explore the physical properties behind these parallels, we will spend the next few chapters applying cluster analysis to simulated cloud populations and examining the results.

Chapter 5

Cluster Analysis of Modeled Results and Comparison to Observations

5.1 Introduction

To this point, we have worked towards establishing a mechanism to sort tropical clouds into cloud regimes via cluster analysis, and with these results, have computed the environmental properties of these cloud regimes. We have also determined the evolution with time of each cloud regime, and found there to exist cycles of convective growth and decay, with characteristic timescales common to the region. Our application of cluster analysis to cloud properties has proven to be a useful analysis tool for investigating the properties of the environment based on sorting atmospheric observations – but we are not limited to applying our analysis tools to only observations.

Recently, the ongoing evolution of climate modeling has seen the replacement of conventional convective parameterization with embedded cloud-resolving models (Grabowski and Smolarkiewicz (1999), Khairoutdinov and Randall (2001), Khairoutdinov et al. (2005)). The improvement by using this technique allows for much greater fidelity between model computations of cloud-scale physics and the processes of the real atmosphere. As such, the newest generation of cloud-resolving models represent a wealth of potential information with which to compare to atmospheric observations – it is expected that application of our cluster analysis technique, which yielded good results on observations, will allow us to perform a quantitative assessment of the properties of these large-scale cloud resolving models. In this chapter, we perform such an analysis and discuss the results of the comparison.

5.2 RAMS simulation of Tropical West Pacific convection

The large-scale cloud resolving model used in this study is the Regional Atmospheric Modeling System (RAMS) developed at Colorado State University (Cotton et al. (2003), Saleeby and Cotton (2004)). RAMS is a non-hydrostatic numerical model that incorporates sophisticated microphysics, radiation, surface and turbulence schemes. As described in the above papers, the surface processes were parameterized using the Land Ecosystem-Atmosphere Feedback-2 (LEAF-2) scheme (Walko et al. 2000), turbulence is represented by the Smagorinsky (1963) deformation-K closure scheme with stability modifications by Lilly (1962) and Hill (1974); and the two-stream radiation scheme of Harrington (1997) was called every 5 minutes.

The model includes a detailed bulk cloud microphysical scheme that assumes a gamma-shaped particle size distribution for three species of liquid water and five species of ice (Walko et al. (1995), Meyers et al. (1997)). The representation of the size distribution in such a manner is consistent with observational studies that have shown particle size distributions in tropical convective clouds closely conform to a modified gamma distribution (Heymsfield et al. (2002)). A description of recent updates to the microphysics is available in Saleeby and Cotton (2004). The cloud droplet spectrum is decomposed into two modes, one for droplets 1 to 40 microns in diameter, and the second for droplets 40 to 80 microns in diameter. Collection is simulated using stochastic collection equation solutions, facilitated by look-up tables, rather than by continuous accretion approximations. The philosophy of bin representation of collection is extended to calculations of drop sedimentation. Concentrations of aerosol species serve as prognostic variables in RAMS to determine cloud droplet concentrations.

For the 2D simulations in this section, we employed a model domain that spanned 9600 km in the zonal direction, from which we extract ten 960km subdomains approximating in the simulated domain the area used for our observational analyses. The horizontal grid spacing of this model is 2.4 km, and the vertical grid consists of 38 levels, which stretch gradually from a depth of 20 meters in the boundary layer to over one kilometer in the upper troposphere, resulting in a model top located at approximately 26km AGL. The resolution of this model approximates the resolution available to satellite sensors, and as such, is a good fit for our cluster analysis. RAMS is initialized with a sounding taken during the TOGA COARE experiment on 0Z 5 December 1992, with an initial zero mean wind and a uniform and fixed sea surface temperature (SST) of 300K. The solar zenith angle is kept constant such that the daily insolation is maintained at 447 W m^{-2} , representative of the mean annual insolation at the equator. As the initialization sounding is weakly convective, no initial forcing is used to initiate convection. The model undergoes an 11-day spin-up period to allow convection to develop and fill the model domain. These two initialization parameters will have some affect on the regime statistics – we discuss these effects presently.

After spin-up, the simulation is run for 55 days, sampled twice daily for each of our ten subdomains to yield a total of 1100 cloud fields for analysis. Another consideration is the affect of the constant zenith-angle approximation on the time evolution of the cloud fields – effectively, the ‘sun is always on’ in this simulation. With a constant solar forcing in the model, crucial parameters such as surface insolation, and the ensuing SST and stability profiles in the simulated environment will largely dictate

the onset of convection in the model. This is an important difference between the model and the environment, and as such, we can expect some differences in the statistical representation of the frequency of occurrence for convection in the model versus in our observational study. As such, we make no attempt to explain the mechanisms behind convective organization and their associated frequencies of occurrence from these modeled results – we leave that for a future model run incorporating more realistic solar forcing – and instead look to see if the generic *structure* of our modeled cloud regimes correspond to the structure of the regimes found using our satellite analysis.

In the RAMS simulation, convection is initiated through random perturbations to the initial potential temperature field, based on the 5 December 1992 TOGA COARE sounding. Although this particular sounding is representative of a relatively quiescent period of convection in the campaign, the initial perturbations to the domain used to trigger convection are of sufficient magnitude to trigger a wide spectrum of convection, and during the 55-day run of the model a great variety of convective systems develop as the model is allowed to react freely to the perturbations induced by cloud formation and dissipation in the model.

To perform our cluster analysis, we use the cloud hydrometeor mixing ratios from the model to determine the uppermost pressure level containing cloud water/ice - from this we derive the familiar cloud-top pressure field. Given that we have complete information in the model regarding cloud-top heights *viz* multiple cloud tops, it is possible to perform the cluster analysis using information from *each* cloud-top height. Future implementations of this algorithm may utilize this information, but for now, we choose to use only the highest cloud-top level to better match our previous work.

Cloud optical depth and surface rainrate are separate fields in the model output – combined with the cloud-top pressure information described previously, we can populate 3D histograms using the same parameters described in Chapter 3 for the cluster analysis.

5.3 Cluster analysis of simulated Tropical West Pacific convection

As previously mentioned, we use the twice-daily histograms from the model simulation for our cluster analysis. The histogram structure used (specifically, the values delineating each bin in the histogram), is identical to those of our previous studies using MODIS and AMSR-E observations to facilitate intercomparison. With respect to how these histograms are populated, however, there are some key differences between our modeled and observed histograms. First, the values from the model are necessarily exact (at least from the point of view of the model), whereas the values from the observations are taken from retrievals, with associated error bars attributed to the retrieved values. This introduces the possibility of a retrieval-versus-computational bias, which is often compensated for by using the model to simulate radiances which are then fed to the retrieval algorithms. This process eliminates any bias inherent to the retrieval algorithm, but is somewhat time-consuming. What's more, the robustness of a histogram-based algorithm such as ours preempts such measures, as the values used to populate an individual histogram bin can vary over a fair percentage before being assigned to a different bin. Those values that for whatever reason *do* get assigned to a new bin due to bias will most certainly be assigned to a neighboring bin, the Euclidian distance to which is the smallest possible in our clustering algorithm. As such, any biasing errors introduced by using the model-computed values rather than simulated retrieval values

will be small enough to not influence the cluster algorithm at all, and may safely be ignored.

A final complication is computing cloud cover and regime-averaged rainrate - as we compute cloud cover by counting the pixels assigned to histogram bins, and since each profile containing multiple cloud layers contributes an increased pixels to the histogram, we would be over-estimating cloud cover. Similarly for rainrate, we would be double-counting raining profiles, resulting in an overestimation of rainrate. As such, we use only the uppermost cloud-top pressure for each profile in computing total cloud cover, which restores the correct statistics. Similarly, we only integrate over the rainrates using the highest cloud-top pressure to compute regime-averaged rainrate. As these two properties are the only ones computed directly from the histogram, these corrections are sufficient for our purposes.

The 110 histograms from the simulation, computed as described, are then processed using the same k -means cluster algorithm developed in Chapter 3. The resulting centroids are then subjected to the same stability, confidence, and significance tests used to evaluate the observational centroids described in Chapter 3.

To facilitate comparison with the earlier results, we start with a value of $k=4$ and evaluate the results, discussing other values of k in brief. As with our observational analysis, a $k=4$ solution of 3D cluster analysis of the model results yields four regimes varying by convective activity level, with some caveats. Figures 5.1-5.4 depict the four regimes found using the model data. As before, we have two convectively-active regimes and two convectively-suppressed regimes. The general properties of the four

model regimes are comparable to those of the observed regimes, with differences that we will discuss in detail presently.

Cluster 1 in our model represents the model trade cumulus regime, and has a relative frequency of occurrence in the model of 64.55%, and is composed primarily of high thin clouds with a sizable amount of non-precipitating low cloud of moderate optical depth. Cluster 2, with a relative frequency of occurrence of 17.27%, represents the ‘enhanced cumulus’ regime, with fewer high, thin clouds and more clouds of moderate optical depth in light-to-moderate rainfall regimes. Cluster 3, representing a regime of isolated convection, has a relative frequency of occurrence of 11.82% and consists of predominantly high, thick clouds of moderate-to-heavy rainfall, with few low clouds in evidence. Cluster 4, the least frequent regime with a relative frequency of occurrence of 6.36%, contains an abundance of high, thick cloud with a second component of low-topped, optically thick clouds of light rainfall. Compared to the observations, both convective regimes contain a much larger proportion of high, thick cloud, representing one of the important differences in the modeled environment compared to that of the observations used in Chapters 3 and 4.

Before delving into individual regime properties and comparison to observations, it would be prudent to demonstrate that the four regimes thus found obey the three principles set out in Chapter 3 – namely, that the four regimes are *stable* to changes in initial centroid selection, that we have *confidence* that the individual histograms belong to each centroid, and that the four regimes are *significant* with respect to each other. As the reader recalls, we test stability of regimes by performing the cluster analysis 5000 times, selecting randomly the initial centroids each time, and then computing the mean pattern

correlation of each regime's 5000 incarnations. Stability is confirmed if this pattern correlation is high, inferring that the same regimes result regardless of the initial choice of centroids.

A final issue to address is the choice of k for this run. As stated in Chapter 3, the goal of a successful k -means analysis is to find the greatest value of k for which significant cluster results can be found. As the reader recalls, for our observations studies, we found valid results for both $k=3$ and $k=4$, and naturally choosing the latter value to fulfill the stated goal. Should we find $k=4$ to be a valid result, we would naturally choose to forgo the tests on the $k=3$ case; since, although the results of this case would likely also be valid, we would still choose the $k=4$ case as it would provide a greater delineation between cloud regimes compared to the $k=3$ by simple virtue of a greater number of regimes for analysis.

This leaves us with the cases for $k > 4$. A test was attempted on the $k=5$ case, but as in the $k=6$ case for the observational clusters, the resulting centroids could not be reliably identified for use in the stability analyses due to the increased sensitivity to initial centroid selection inherent to higher-valued k analyses. For the sake of brevity, we will therefore neglect the unstable $k > 4$ solutions and continue with our analyses using the $k=4$ solution first.

Stability pattern correlation for the trade cumulus, enhanced cumulus, singular convective, and organized convective regimes are, respectively, 0.98, 0.94, 0.89, and 0.87. These values, as the reader will note, are lower than those in the observations, which is to be expected since we are working with fewer histograms compared to the observational dataset. The lowest correlation, at 0.87 belongs unsurprisingly, to the

regime with the fewest observations (the organized convection regime.) As we will see, this also leads to greater uncertainty in the inter-cluster confidence parameter. As we do not have a set rejection threshold, and since our correlations are relatively high compared to the rejected values for $k=5$ and higher in the observational results discussed in Chapter 3, let us provisionally accept the stability of the regimes, pending the outcome of the other two tests.

Our confidence test again measures the mean intercluster distance between a centroid and its member observations, and compares this distance to the intracluster distances between the centroids themselves. We assert *confidence* in the cluster analysis if the mean intercluster distance is lower than all of the intracluster distances, which would signify that the member observations of a centroid are more similar to their centroid than all other centroids. Tables 5.1 and 5.2 respectively present the computed intercluster and intracluster distances confidence test for the $k=4$ RAMS cluster analysis.

As might be expected, the two most similar clusters are the convectively-active clusters, with an intra-cluster distance the same order of magnitude as the intercluster distances for each cluster (albeit larger than the two intercluster distances.) Other clusters have larger intracluster distances than their mean intercluster distances, especially in the case of the enhanced cumulus cluster, Cluster 2. Overall, both inter- and intracluster distances are larger for the simulated case than for the observational case. This is possibly due to the relatively increased instability in the clustering results compared to the observational case, but also might point to a relative homogeneity of cloud structure in the real atmosphere compared to a modeled atmosphere. It is noted that the model simulations are essentially a compacted version of the atmosphere, driven constantly by a

Table 5.1. Intercluster distances for $k=4$

<i>Cluster (description)</i>	<i>Mean intercluster distance</i>
Cluster 1: trade cumulus	0.00483
Cluster 2: enhanced cumulus	0.00987
Cluster 3: singular convection	0.0127
Cluster 4: organized convection	0.0326

Table 5.2. Intra-cluster distances for $k=4$

<i>Clusters</i>	<i>Distance</i>	<i>Clusters</i>	<i>Distance</i>	<i>Clusters</i>	<i>Distance</i>
Cluster 4 – Cluster 1	0.2133	Cluster 3 – Cluster 1	0.2648	Cluster 2 – Cluster 1	0.2507
Cluster 4 – Cluster 2	0.4641	Cluster 3 – Cluster 2	0.5156		
Cluster 4 – Cluster 3	0.0514				

fixed-angle solar forcing with a fixed SST – the results are a much greater representation of ‘transitional’ cloud phases in a temporal sense in the model output than is observed from satellite analysis. This condition would explain the relative increase in both inter- and intracluster distances in the model-derived clusters compared to the observational clusters.

With the exception of the convective cases, the intra-cluster distances are comfortably larger than the intercluster distances; with this limitation in mind we express a reasonable confidence in our cluster analysis and stability thus far. The exception to this is the nature of the convective regimes – although we have confidence in their *existence* as separate regimes compared each other, we would perhaps be in error to

express a good deal of confidence in their computed *structure* without utilizing more observations to bring the statistics more in line.

Finally, we assess the significance of the regimes. As the reader will recall, pattern correlation combined with an assessment of total cloud cover is used to assess regime significance, with low correlation (and differing cloud cover amounts in special cases of higher correlation) between the regimes as confirming the regimes to be significant compared to one another. Table 5.3 presents the results of the significance tests performed here.

The increased heterogeneity of the modeled regimes, perhaps responsible for the larger inter- and intracluster distances as found in the confidence test, also appears to have an influence the results of the pattern correlations presented in Table 5.3, as pattern correlations are quite low. The most similar regimes are the trade cumulus and singular convective regimes, with a pattern correlation of approximately 0.48, followed by the two convective regimes, whose correlation with one another was computed to be around 0.42. This brings into sharp focus the utility of using precipitation amount to separate cloud regimes, as in the framework of the 2D ISCCP histogram, the cirrus and cumulus clouds populating the trade cumulus regime can bear a similarity to the anvil-and-outflow structure seen in convective regimes.

Overall pattern correlations between the regimes are elsewhere quite low, and we therefore find significance in the computed regimes. Combined with our confidence and stability tests, we find the computed regimes of the $k=4$ RAMS simulation to be valid. One reservation must be expressed in these results, however – the stability and confidence analyses of the convective regimes are sufficient to mark these regimes as

Table 5.3. Coefficients of pattern correlation between clusters in the $k=4$ case

<i>Clusters</i>	<i>Correlation</i>	<i>Clusters</i>	<i>Correlation</i>	<i>Clusters</i>	<i>Correlation</i>
Clusters 1-4	0.045858	Clusters 1-3	0.477614	Clusters 1-2	0.093396
Clusters 2-4	0.318291	Clusters 2-3	0.011276		
Clusters 3-4	0.420417				

generally stable and sufficiently self-similar for simple identification analyses, but regime-averaged properties computed from these regimes must take into account that there is likely some swapping of observations between the two convective regimes, which would be entirely a result of differing initial conditions.

5.4 Properties of individual clusters and comparison to observations

As in Chapter 4, we present each regime separately, beginning with the convectively-suppressed regimes and moving to the convectively active regimes. Also as in Chapter 4, we use the term ‘regime’ as opposed to ‘cluster’ or ‘centroid’ to discuss the results. Table 5.4 summarizes the results, and may be found in Section 5.5.

5.4.1 Cluster 1 – Trade Cumulus and Cirrus regime

Figure 5.1 presents the trade cumulus regime computed from RAMS. As stated previously, this is the most frequent regime in the simulation, with a relative frequency of occurrence (RFO) of 64.5% and a total cloud cover (TCC) of 40.9%. The RFO for the model compares favorably with that of the observed trade cumulus regime, which was 53.3% (and was also the most frequent regime of the observed regimes). The total cloud cover for the observed trade cumulus regime, however, was much smaller - 4.8%, which is nearly an order of magnitude less than the cloud cover for the simulation. Of particular

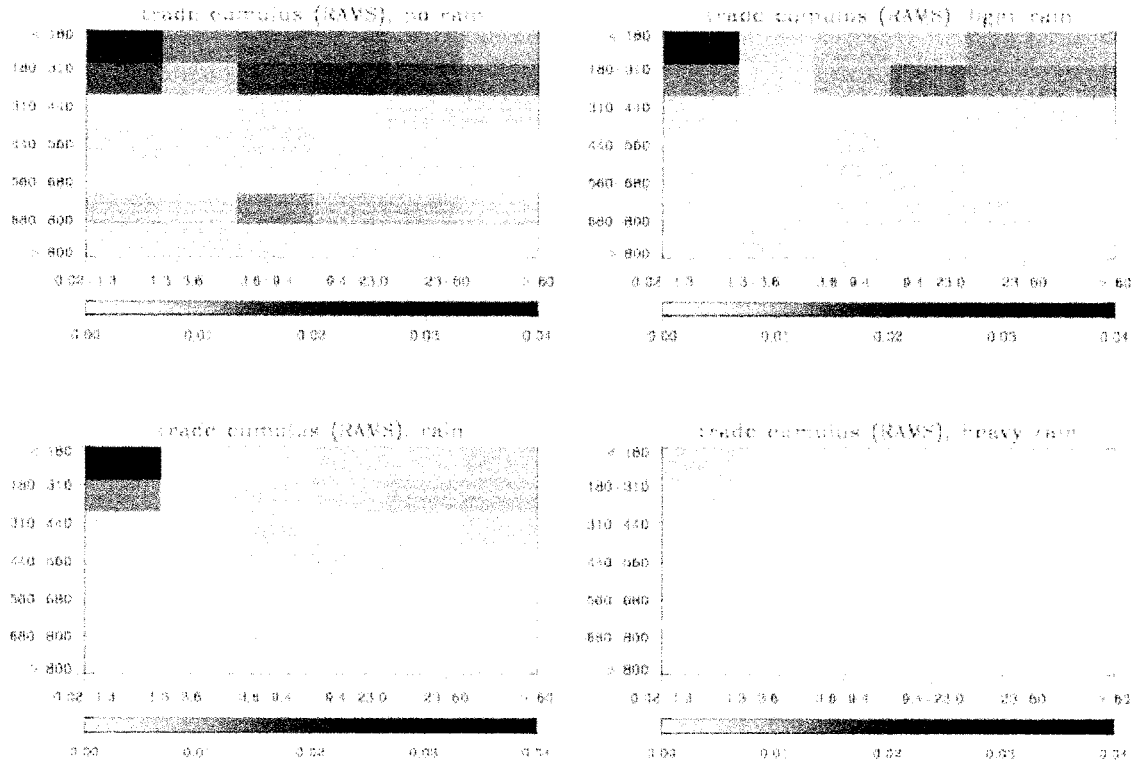


Figure 5.1. 3D histogram of the trade cumulus and cirrus regime for the RAMS simulation. Each panel consists of a 2D histogram of cloud-top height along the y-axis and cloud optical depth along the x-axis, with each panel representing a single rainrate bin.

importance is the difference in cirrus cloud amount between the model cumulus regime and the observed cumulus regime – as we can see, the model regime contains quite a bit more cirrus than the observations, and this increase in cirrus amount is the likely cause of the discrepancy between the model and observations.

As with the observed trade cumulus and cirrus regime, the simulated trade cumulus and cirrus regime consists largely of high, thin clouds combined with an equal proportion of high, thicker clouds and low-topped clouds of moderate thickness. The observed trade cumulus regime contained the high, thin clouds and a signal of high thicker clouds as well, but with a relatively smaller proportion of low clouds. This is the first (and ultimately the primary) difference we see between the simulated and observed cloud populations – our simulated cloud population is a complete population, whereas it appears that passive-only satellite observations may be missing that fraction of lower cloud residing underneath optically thick higher cloud layers. This insight warrants continued investigation into the existence and amount of low-level cloud in the tropical atmosphere.

5.4.2 Cluster 2 – Enhanced Cumulus regime

Figure 5.2 presents the histogram of the enhanced cumulus regime for the RAMS simulation. As with the observed enhanced cumulus regime, the simulated enhanced cumulus regime is marked by an increase in both the frequency and the thickness of the low-level cloud field, with a concomitant increase in rainfall (although the bulk of the profiles remains in the lowest two rainrate bins, with rainrates < 1.5 mm/hr.)

What is notable about the simulated regime compared to its observed counterpart is the relative dearth of high, thin cloud in the simulated regime – the observed enhanced

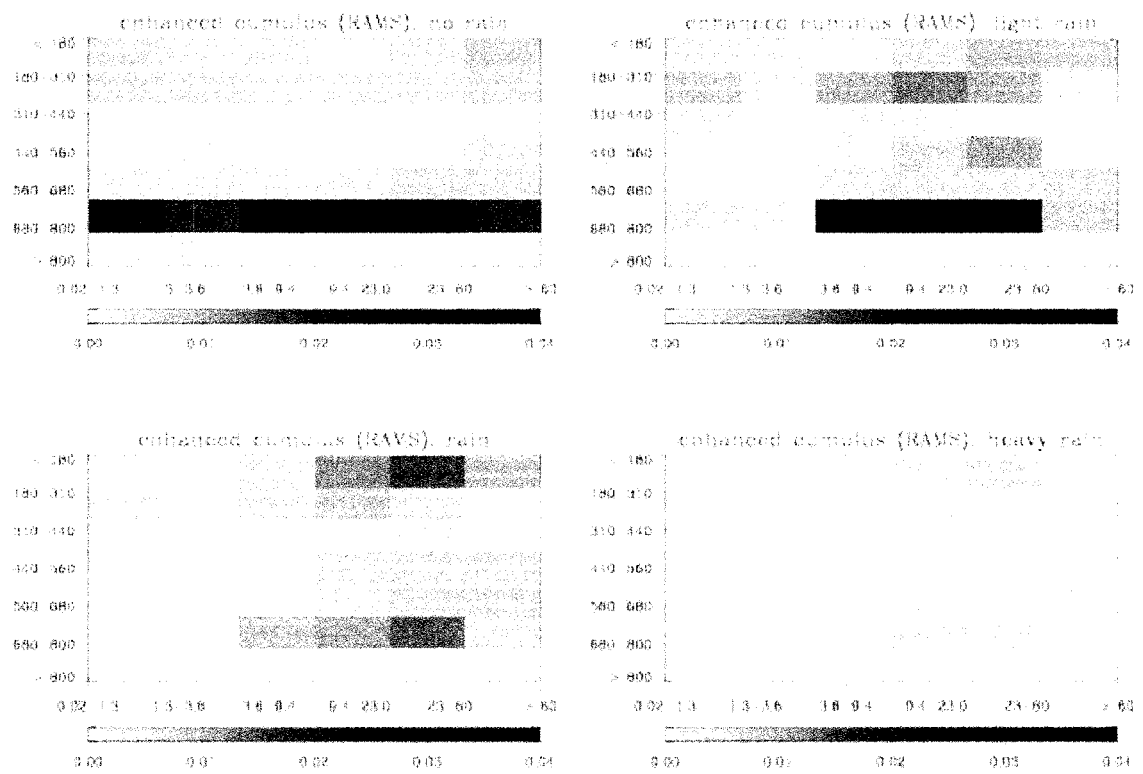


Figure 5.2. As in Figure 5.1, but for the enhanced cumulus regime.

cumulus regime remained primarily one of upper-level cloudiness with a moderately stronger low-level cumulus signature. The simulated regime nearly eliminates the upper cloud signature compared to its low-level cloud signature – this is again perhaps a result of more completely utilizing multiple cloud layers in the model output compared to observations.

In a qualitative sense, these simulated results are more in line with what one would expect in regards to convective evolution of our regimes - as upper-level cirrus amount begins to tail off, increased insolation would contribute to the surface warming and an increase in strength of the boundary layer, resulting in thicker, more substantial cumuliform clouds, such as those represented in our simulated regime, and agreeing with our observational results shown in Chapter 4.

The relative frequency of occurrence for this regime is 17.3%, with a total cloud cover of 46.0%, compared to 21.9% and 29.3% for the observed regime - again, the model is somewhat over-representing total cloud cover, and likely for the same reasons discussed previously in the trade cumulus section.

5.4.3 Cluster 3 – Isolated Convection regime

Figure 5.3 presents the first of the convectively-active regimes, the so-called ‘singular convection’ regime. Structurally, the simulated and observed regimes (reference Figure 4.3 for comparison) are somewhat similar – the simulated regime tends to produce a stronger, more uniform signal of high, thick clouds of moderate-to- heavy rainfall, while the observed regime tends to be spread over a slightly larger range of cloud-top height and optical depth.

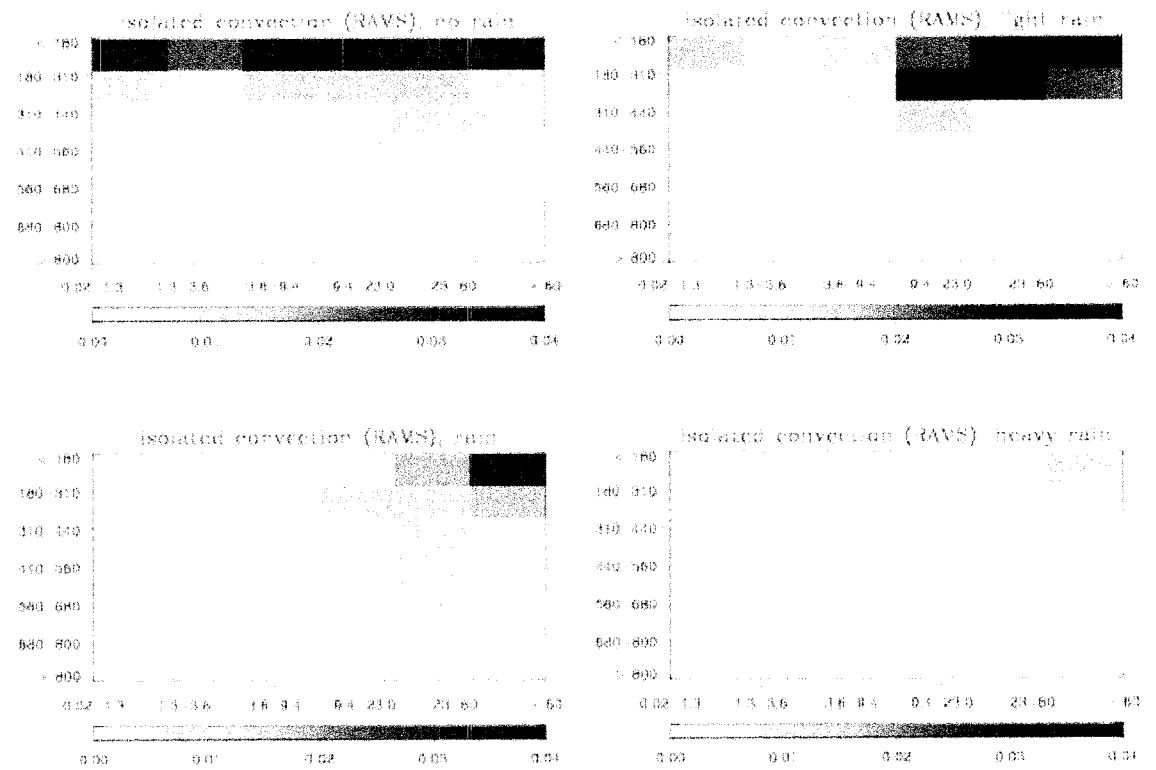


Figure 5.3. As in Figure 5.2, but for the isolated convection regime.

In both regimes, the bulk of the rain falls in the 0.06-1.5 mm/hr rainrate bin, with a reasonably strong signal in the 1.5-5.1 mm/hr bin, and a weak signal in the > 5.1 mm/hr bin, representing the eponymous areally-sparse convective cores of these systems. The bulk of the profiles observed are either high, thick clouds associated with towering convection, or high, thinning clouds likely associated with the cirrus anvils produced by the aforementioned convection – few (if any) profiles represent a significant low-level cloud structure such as that seen in the stratiform regions of more organized convective regimes.

The relative frequency of occurrence for this regime is 11.8%, with a total cloud cover of 67.3%, compared to 17.5% and 59.2%, respectively, for the observed regime. As we move into more convective regimes, the larger areal extent of convection (as represented by anvil outflow in the model) appears to be compensating for the footprint differences between the model and satellite resolving capability, and to that extent, cloud cover amounts computed by the observations and by the model are in better agreement.

5.4.4 Cluster 4 – Organized Convection regime

The fourth, and final, regime for consideration is the organized convection regime, whose histogram is presented as Figure 5.4. The least frequent regime, with a relative frequency of occurrence of 6.4% (versus 7.3% for the observations) the organized convection regime is nevertheless responsible the greatest rainfall and cloud cover of the four regimes. Total cloud cover for the simulated regime is computed to be 92.4%, compared to 96.1% for the observed regime.

An intriguing result seen in Figure 5.4 reveals some details in the nature of the differences between organized convective clusters and isolated convective cells (at least

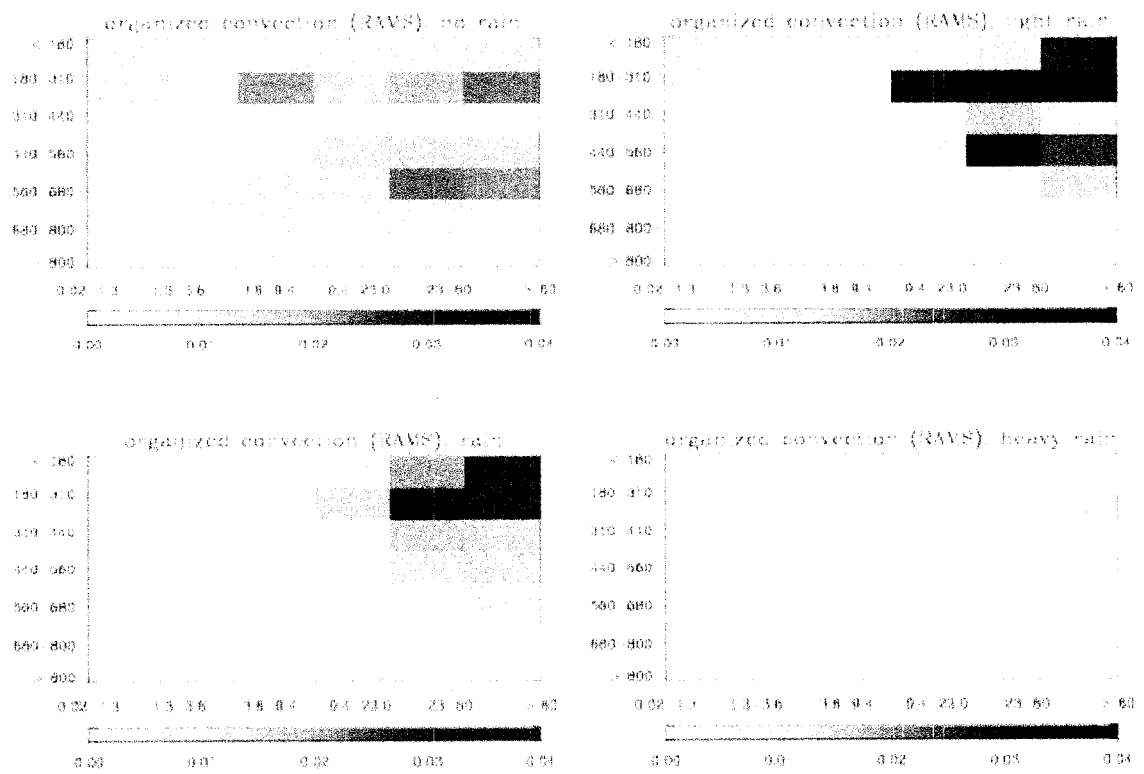


Figure 5.4. As in Figure 5.3, but for the organized convection regime.

in the framework of simulated convection.) Overall, cloud-top heights are lower for the organized convective regime than for the singular convective regime, with a much stronger signal of low, thick cloud of moderate-to-heavy rainfall present in the organized convective regime versus the singular convective regime. A larger fraction of the raining profiles in the organized convection regime reside in the moderate-to-heavy rainrate bin compared to the singular convective regime. What appears to be a smaller fraction of heavily raining profiles in the organized convective regime is compensated by an increase of 25% in cloud cover, yielding a much higher number of heavily raining profiles as well. Compared to the singular convection regime, the organized convection regime consists of slightly lower-topped, thicker clouds of much greater areal extent, and perhaps most importantly, containing a significant fraction of relatively low, thick, raining cloud suggestive of a stratiform component not present in the singular convective regime.

The structural differences between singular convective cells and organized convection in the model are compelling as the profile of atmospheric heating due to latent heat release by convection differs between differing convective modes – our ability to distinguish between these modes through utilization of multiple cloud top information in the model suggests that in the inclusion of active sensors capable of detecting these lower cloud layers, which allows for the distinction between towering and stratiform convection will be of great utility for future applications of cluster analysis. We will describe such a potential use for cluster analysis in Chapter 6.

5.5 Discussion

Table 5.5 summarizes the relevant results of this model study compared with the observations presented in Chapter 4. In this table, we compare the relative frequencies of

Table 5.4. Summary table comparing observed regimes and modeled regimes.

<i>Cluster</i>	<i>RFO</i>	<i>TCC</i>	<i>Rainrate in mm/hr</i>
Trade Cumulus (observed)	53.3%	4.8%	0.52
Trade Cumulus (model)	64.5%	40.9%	0.40
Enhanced Cumulus (observed)	21.9%	29.3%	0.61
Enhanced Cumulus (model)	17.3%	46.0%	0.73
Singular Convection (observed)	17.5%	59.2%	0.67
Singular Convection (model)	11.8%	67.3%	1.08
Organized Convection (observed)	7.3%	96.1%	0.91
Organized Convection (model)	6.4%	92.4%	1.73

occurrence, the total cloud cover, and the rainrate for each regime, both model and observed.

Overall, the structural properties of the simulated regimes and the observed regimes agree well, with a few important differences. Most prominently, the enhanced cumulus regime in our model results shows an increase in low cloud amount with a decrease in upper-level cloud amount relative to the observed enhanced cumulus regime. Similarly an increase in low-topped, thick raining cloud is seen in the model organized convection regime compared to its observed counterpart. Overall, while the generic structure of each regime is quite similar between the observed regimes and the model regimes, the model regimes do contain a larger amount of low-topped cloud populations. This is a key difference between passively-sensed cloud regimes and model cloud

regimes – the question remains as to which result better represents the real atmosphere. Although only preliminary results are available, observations from CloudSat tend to support the model’s view of the atmosphere. Further quantification of the amount of low cloud using active sensors will further delineate these differences.

Regime-averaged rainfall properties for the four regimes are computed in the same manner as for the observed regimes described in Chapter 4. Regime-averaged rainfall rates for the trade cumulus, enhanced cumulus, singular convection and organized convection regimes are, respectively, 0.40 mm/hr, 0.73 mm/hr, 1.08 mm/hr, and 1.73 mm/hr. As mentioned previously, these values are computed by integrating over all raining profiles identified by the uppermost cloud-top height to prevent double-counting of raining profiles, which would result in erroneously high values of regime-averaged rainfall. With the exception of the trade cumulus regime, these values are somewhat higher than the values from our AMSR-E analysis presented in Chapter 4, although not unrealistically so – maximum rainrates for the enhanced cumulus regime are less than 41mm/day. Here we recall that the AMSR-E footprint is larger than the effective footprint of our model – it is possible that the ‘smearing’ of rainrates in the observations over a larger footprint area could account for these differences.

Finally, we discuss the utility of active versus passive-only observations in the formulation of histograms for use in cluster analysis. At the outset of this study, limited or no active sensor data were available for analysis to compare with MODIS observations; with the launch of CloudSat, that situation has now changed. With the ability of the CloudSat CPR to penetrate upper cloud layers to see underlying regions of cloudiness, and especially with addition of precipitation retrievals using CloudSat radar

attenuation (Haynes et al., forthcoming), it is possible to combine MODIS observations with CloudSat observations to generate a new 3D histogram structure, using cloud-top pressures derived both from MODIS and from CloudSat, and per-layer precipitation and LWC values from CloudSat, that captures more fully the true cloud morphology that best reflects the state of the environment in which the cloud resides. Continued development of these retrievals, especially the CloudSat-derived precipitation data, is an area of ongoing research – it is our hope that the groundwork laid by this study will prove useful for these developments. Certainly, it remains important to use a blend of sensors to maximize information content when populating histograms – a CloudSat-only cluster analysis is described by Zhang et al. (2006) in which it is found that using only cloud radar data to count clouds results in an underestimation of total cloud cover by around 30%. In the Zhang et al. work, it is supposed that thin clouds below the minimum detectable signal of the radar as well as low-level clouds lost in the surface return remain uncounted, yielding skewed cloud statistics. Combining the greater sensitivity to high, thin clouds provided by MODIS (or optionally, from the CALIPSO mission (Winker et al. 2006)) would help to reduce this underestimation, while retaining the value of having multiple cloud-tops as observed by the CloudSat radar.

The results of this chapter have provided us with valuable insight into the utility of cluster analysis as applied to large-scale cloud resolving models. As these models are an increasingly important component of analytical tools and retrieval techniques, it is of equal importance to quantify the comparisons between these models and observations – good agreement between cloud regimes as found through model analysis and observations gives us greater confidence in model results, and permits the use of these

powerful tools to represent as-yet irretrievable properties of the atmosphere. In the next chapter of this dissertation, we explore how application of cluster analysis to model results might facilitate a new retrieval technique, and various applications of this technique that might be relevant to analysis of the tropical hydrologic cycle as a whole will be discussed in the final chapter of this work.

Chapter 6

Towards a CloudSat Latent Heat Product Using Cluster Analysis

6.1 Introduction

Thus far, we have concerned ourselves with two primary tasks – first, describing a cluster technique that allows for the identification of unique environmental regime properties based on analysis of cloud properties, and second, analyzing the results of this analysis in the framework of both observations and model analysis. Having described how to use cluster analysis as an analytical tool, we now perform some preliminary work to assess the feasibility of using this tool in active areas of research.

6.2 Identifying Latent Heating Properties through Cluster Analysis

Latent heating through condensation of water provides the energetic source for tropical circulation. Several studies (Nakazawa et al. 1988, Emanuel et al. 1994) have additionally emphasized the key role that tropical latent heating plays in regulation on larger-scale, extratropical circulations. As we gain understanding of the importance of the role latent heating plays in the atmosphere, both in the tropics and in the larger sense, so too have we begun to put more effort into developing techniques to measure vertical profiles of latent heat release in cloud systems.

The TRMM mission (Simpson, et al. 1996) is a spaceborne precipitation radar combined with a passive microwave designed to measure tropical rainfall. Recent work (Tao et al. 2006) describes retrieval algorithms based on TRMM observations that retrieve profiles of latent heating from radar-observed rainfall. Several methods may be employed to achieve these retrievals, including spectral methods whereby lookup tables of heating profiles for different precipitation classifications are generated using cloud-

resolving models (CRMs). Spectral methods such as these rely upon classification schemes to partition rain into stratiform, convective, and anvil components. As we have seen in previous chapters, cluster analysis offers another method for classifying cloud properties that may prove useful for retrievals of latent heating.

The CloudSat mission (Stephens et al. 2003) is uniquely suited to study latent heating, due to the additional resolution and sensitivity of the cloud-profiling radar (CPR) over the TRMM precipitation radar. As such, developing a similar retrieval of latent heating, using cluster analysis as a component for discriminating between different cloud structures, seems a natural evolution. As a first gesture towards developing such retrievals, we turn again to model studies, utilizing a 3D RAMS run, using the domain and initialization properties of the 2D run described in Chapter 5. In addition to standard model outputs, we also collect profiles of latent heating from this RAMS run (latent heat profiles were not saved from the 2D runs used in Chapter 5.) We take 750 vertical profiles of latent heating from the 3D model field, compute the cloud-top pressure, cloud optical depth (by layer) and surface precipitation for each profile, and compute a corresponding histogram using these three values. The histograms are then sorted using the results of the 2D analysis – this simulates how satellite observations of cloud properties (represented by the 750 model profiles) would be sorted using a canonical regime analysis computed periodically using archived satellite observations (represented by the 2D regime analysis results.)

Although we do not explicitly utilize radar reflectivity in these cluster analyses, we will ultimately want to develop radar reflectivity-latent heat profile relationships. To

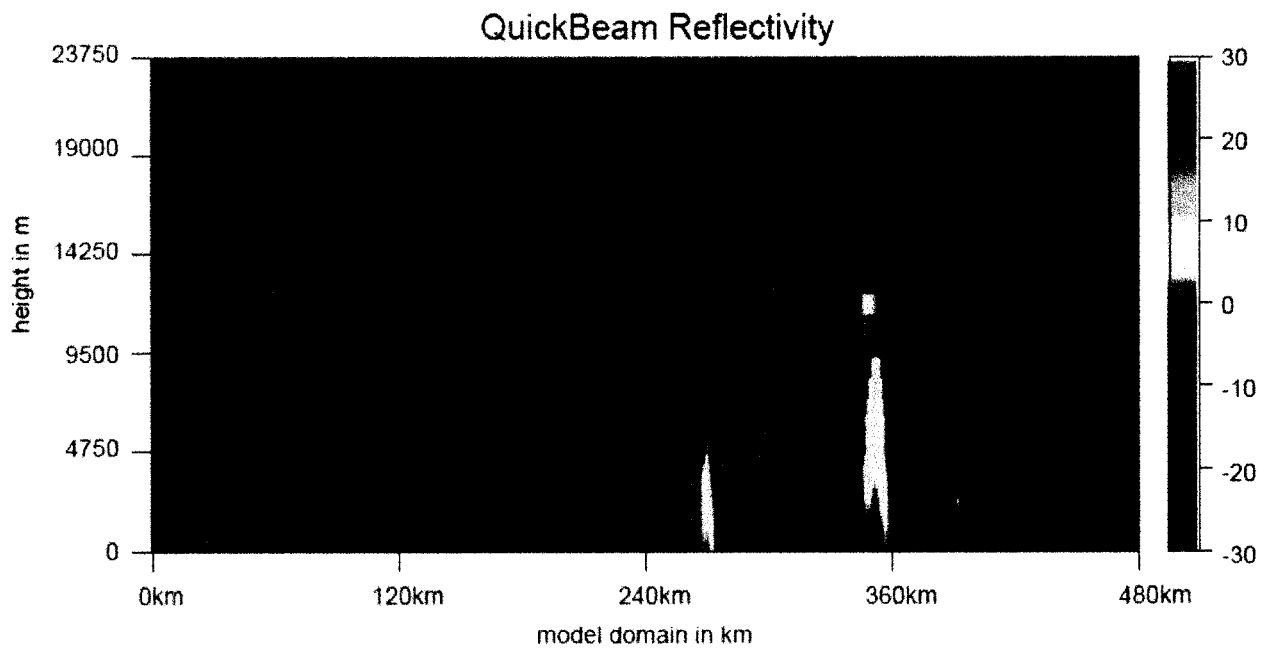


Figure 6.1. Simulated reflectivity profile of an organized convective system present in the 3D RAMS simulation used for our latent heating feasibility study. CloudSat reflectivity is simulated using the QuickBeam simulator package described by Haynes et al. 2007.

that end, we also compute from the 3D model profiles simulated CloudSat observations using the QuickBeam radar simulator package capable of simulating top-down radar reflectivities at a number of microwave frequencies, including the W-band CPR on CloudSat (Haynes et al. (2007)). QuickBeam ingests model profiles of pressure, temperature, relative humidity and hydrometeor mixing ratios, and produces profiles of top-down radar reflectivity. The reflectivity profiles account for attenuation of the simulated radar beam, including those from gaseous absorption by the atmosphere and from the hydrometeors themselves. An example of a simulated radar profile produced by QuickBeam is provided as Figure 6.1. To generate this figure, one of the 750 individual longitude-height profiles was passed to the radar simulator program. As mentioned, for this feasibility study we are primarily concerned with gathering statistics for cloud properties as a whole in the region, rather than comparing individual model profiles to CloudSat observed profiles. Longitudinal-height slices such as these are available at each time step and at each grid point, and are used for our statistical analysis, and will eventually be used in development of the latent heat retrieval.

As a comparison, we have included a representation of model-computed latent heating as Figure 6.2. The structure of latent heating as seen in Figure 6.2 is, exhibits a broad area of heating just above the melting level in the model, along with a cooling signature at and just underneath the melting layer. The structure of the heating is located in the primary updraft/raining region of the convection, corresponding to the highest areas of reflectivity in the simulated radar profiles. These results may be compared to analytical studies of the structure of vertical heating by different convective modes proposed by Mapes et al. (2004), Tulich et al. (2006), and Wheeler et al. (2000).

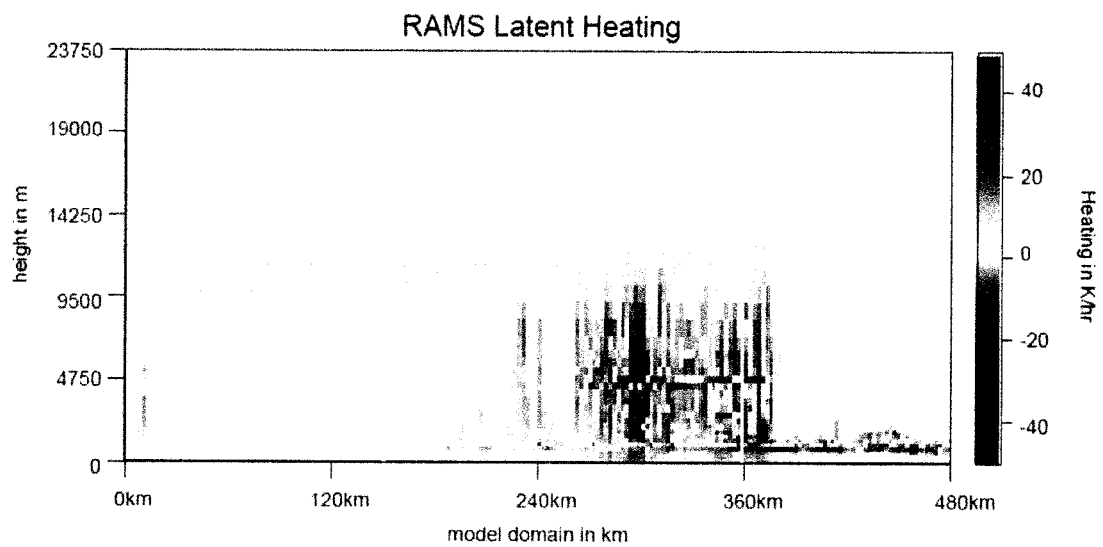


Figure 6.2. Simulated CloudSat reflectivity profile and computed latent heating of a convective element identified as isolated convection using 2D cluster analysis results.

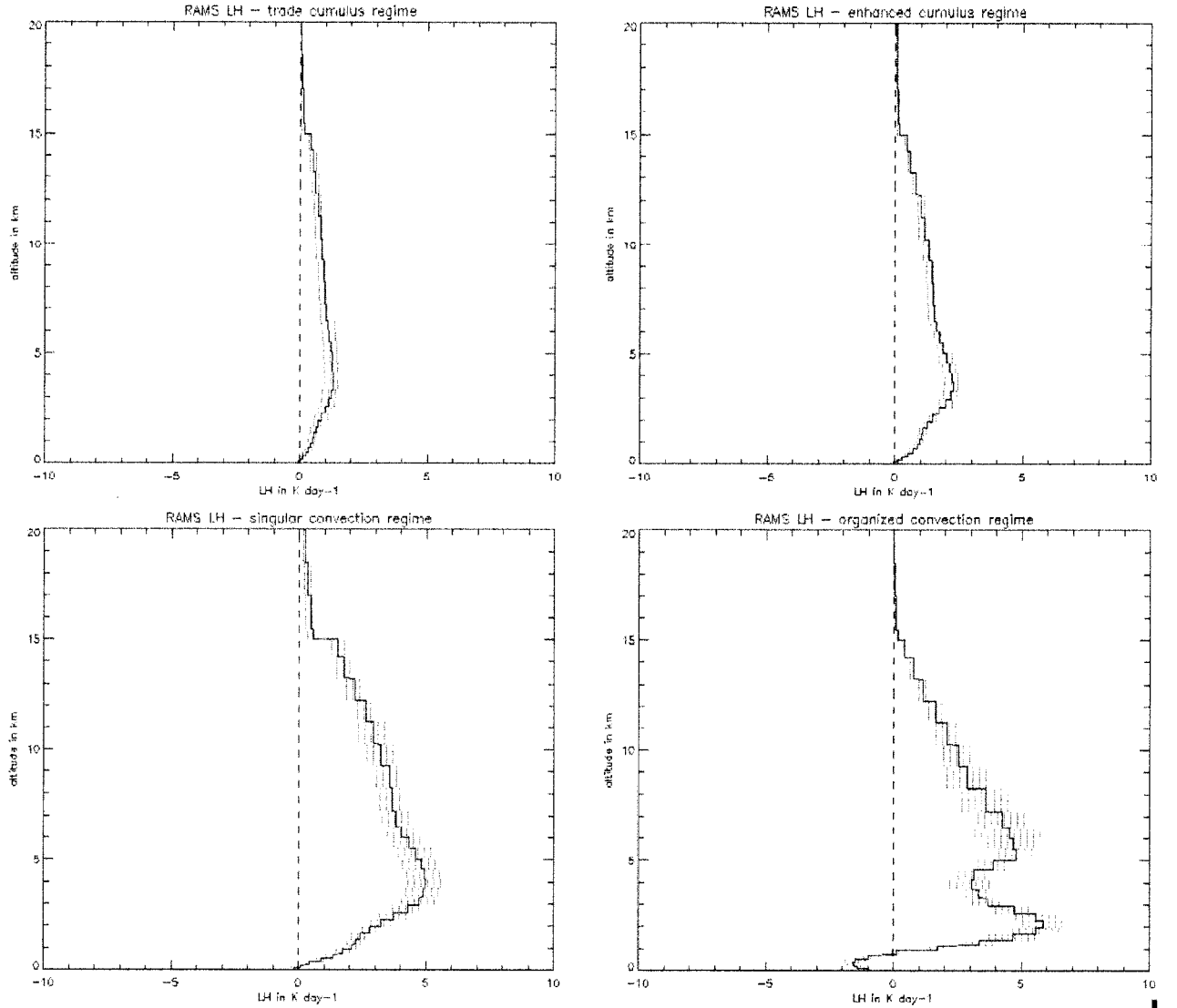


Figure 6.3. Mean profiles of latent heating for the regime-sorted 3D RAMS model run. Clockwise from the upper left, mean profiles of latent heating are presented for the trade cumulus regime, enhanced cumulus regime, organized convection regime, and isolated convection regime. Shading around the profile denotes the standard deviation of the mean for the latent heating at each level.

Using our model generated histograms as input data, we sort model observations into cloud regimes using the cluster centroids found in Chapter 5. We then compute regime- and domain-averaged profiles of latent heating, which we present as Figure 6.3. The first notable results of Figure 6.3 is that the profiles of latent heating for the convectively-suppressed regimes (the top two panels in the figure) are not entirely convection-free, as non-negligible heating exists above the nominal cloud tops for shallow, suppressed clouds in a regime-averaged sense. It is important to remember that the profiles used for this analysis are 960km in length – it is almost inevitable that one or two convective towers will be present in any observation of this size, even when the rest of the scene is generally convectively-suppressed. When averaged over an entire regime, it is the contribution of this occasional convection in otherwise convectively suppressed regimes that supplies the upper-level seen in Figure 6.4. While the overall magnitude of this upper-level heating is small (on the order of 2 K/day or smaller), it is non-negligible and must be taken into account regarding any future retrieval algorithm utilizing cluster analysis – most likely, a modified histogram for identifying cloud regimes using a smaller area will address this convective contamination issue. Developing this modified histogram specially for latent heating retrievals is a future goal of this research.

The convective regimes are represented by the bottom two panels in Figure 6.4. As we noted in Chapter 5, the singular convection regime is typified by isolated cumulonimbus, whereas the organized convection regime is composed of more dense convection, and in the RAMS simulations, typically containing a stratiform rain component. Schumacher et al. (2004) describe corresponding latent heat profiles for varying amounts of convective, shallow convective, and stratiform components of

tropical rainfall using TRMM observations. In general, cloud populations with a significant fraction of stratiform rainfall exhibit low-level cooling below 2km, while shallow convection results in a peak heating between 2-4km; deep convection results in peak heating around 8km. With these generalizations in mind, the convectively-active latent heating profiles presented in Figure 6.4 are consistent with our expectations – the singular convection regime, consisting mostly of isolated deep convective cells, exhibits a peak heating rate at around 4km, with a broad ‘nose’ of heating extending upwards of 10km. The organized convection mode exhibits a more complicated structure – the contribution of the stratiform component is seen with cooling below 1km, and two local maxima of heating rates; one at ~3km, representative of more shallow convection, and a second peak at approximately 6km, again with a broad ‘nose’ of heating extending to nearly 15km.

The relative fraction of each cloud type represented in each regime, and therefore the contribution of each cloud type to the aggregate heating rate for each regime, is very much a function of the dynamics and microphysics of the model responsible for cloud generation. The ability of cluster analysis to separate these important cloud types (namely, shallow convection, deep convection, and stratiform regions) within cloud observations will greatly facilitate the computation of total latent heating profiles of a cloud scene composed of a variety of cloud types. Clearly, much work remains to develop the reflectivity-to-latent heating forward model most applicable to such a retrieval; in the framework of this feasibility study, however, the ability of cluster analysis to separate out relevant parts of observations is evident, and we feel confident

that the application of this dissertation to such problems will result in improved retrieval techniques and analysis of the latent heating of the tropical atmosphere by convection.

6.3 Discussion

Derivation of the vertical heating properties of the tropical atmosphere remains a crucial task, as the implications of convectively-induced heating are thought to have a significant impact on convective organization and control in the tropics (Schumacher et al. (2004), Kiladis et al. (2005)). In order to derive the latent heating properties of convection, it is necessary to understand the vertical profiles and necessary derivatives of the cloud hydrometeors involved, and assessing these properties remains a significant challenge due to the complex processes involved. The utility of cluster analysis, we feel, is in its ability to separate like convective systems into discrete regimes which simplify the assessment of the cloud properties required for these kinds of retrievals, and should prove quite useful in the development of advanced latent heating retrievals.

Chapter 7

Conclusion and Future Research

7.1 Review of Research

The tropical climate systems represents one of the most important physical components in the earth-atmosphere system, and understanding its complexities represents one of the key challenges in the field of atmospheric science at this time. Increased understanding the mechanisms behind the tropical variation in humidity, rainfall, and cloud cover would lead to improvements not only in tropical forecast models, but in global climate models as well, which in turn would better our understanding of the dominant feedback cycles of the atmosphere in the context of climate change.

What is apparent from observation of the tropics is the organization of the tropical hydrologic cycle into distinct phases of convective activity (Johnson et al. 1999, Jakob and Tselioudis (2003), Rossow et al. (2005)) that are related to the important thermodynamics of the tropical troposphere (Nakazawa (1988), Mapes and Houze (1993), Kiladis et al. (2005)). Using the statistical technique of *k*-means analysis as applied to ISCCP histograms of cloud type, Jakob and Tselioudis (2003) described four tropical cloud regimes, with follow-up studies (Jakob et al. (2005), Rossow et al. (2005)) describing more properties. This research confirmed the results of Jakob and Tselioudis (2003) using MODIS observations of cloud top pressure and cloud optical depth, and observed the same tropical cloud regimes, at least as defined by the 2D analysis presented in the literature.

With the understanding that cloud populations within each regime are necessarily representative of the physical processes going on in that regime, and with the idea that perhaps a cluster analysis based on a more comprehensive dataset that includes more information about the cloud properties within each cloud population, we developed a new, 3D histogram for use with *k*-means analysis. We applied this new technique to MODIS observations of cloud-top pressure and cloud optical depth, along with AMSR/E observations of rainfall, to recategorize the tropical cloud regimes. The 3D-derived regimes share many common features with previous analyses, but also have key differences that allow for detailed comparison between the cluster analysis and other analyses of the tropical hydrological cycle.

We then apply our regime analysis to a time series dataset of 3D histograms to determine the occurrence in time of each regime, finding a regular cycle of progression in convective activity that agrees with observations of the MJO (Madden and Julian, (1971)). With this knowledge, we use this index of regime occurrence in the tropical West Pacific to compute regime-averaged properties using reanalysis and satellite observations. We found that the properties of our 3D regimes and the progression in time of these properties as the regimes proceed from one to another agree with other theories of tropical organization (Stephens et al. (2003)). As such, it became possible to analyze the effects of the small-scale interactions between cloud regimes and their larger environment.

Application of our analysis technique to large-scale cloud resolving model underscored the utility of this technique, and found compelling similarities and interesting differences between the regimes computed using model analysis compared to

passive-only observations. With these results in mind, it seems natural to extend the results of the research technique to further analyze the properties of the tropical hydrologic cycle. We now discuss some tantalizing possibilities for use of this particular research tool.

7.2 Future Work

We have already discussed quantitative assessments of large-scale cloud resolving models and performed a feasibility study regarding use of the cluster analysis technique in developing a latent heat retrieval. Obviously, continued research along these lines are expected to bear much fruit. The results of this dissertation suggest additional enhancements to the technique that may prove beneficial to these efforts. In this work, we have focused on the need to develop meaningful cluster input, namely, 3D histograms that accurately represent the relevant cloud properties of the environment in an effort to produce similarly meaningful clusters. For future work, the primary enhancement to histogram development will be to adapt the 3D sorting histogram to use a blend of active- and passive sensors, such as described in Chapter 5. Ideally, a two- or three-sensor approach using CloudSat and MODIS observations, perhaps in conjunction with AMSR-E observations, will provide for us the ideal sorting histogram, capable of accurately representing the key cloud properties of the environment and facilitating a more complete and accurate regime separation.

Applying this histogram to both model analysis and observations will allow for more data to be collected about the important regimes of the tropical system, and with improved regime statistics, which can be assimilated by both the modeling and remote retrieval communities. Using the results of an ongoing cluster computation, instruments

of the A-Train will have a real-time capability to diagnose the regime properties of current observations, which will prove useful in future retrieval algorithms such as the latent heat retrieval discussed in Chapter 6.

7.3 Conclusions

The key conclusions of this research are as follows:

1.) *Cluster analysis is a mature technique, capable of identifying unique and meaningful environmental properties computed using cloud properties as identifying markers.* With careful selection of a sorting histogram, and with judicious application of the stability, confidence, and significance criteria we have developed, the results of cluster analysis can provide a wealth of information regarding the phases of several environmental properties – certainly not limited only to tropical cloudiness estimates. The three tests we developed to assess the validity of cluster analysis; namely, our tests of stability, confidence, and significance, give us a quantitative method for assessing the objective regimes found in cluster analysis.

2.) *Previous cluster analysis techniques using geostationary data can successfully be replicated using other observations.* The ability to verify the results of a MODIS/AMSR-E-based cluster analysis compared to earlier ISCCP-based work gives us faith in the robust nature of tropical cloud regimes, in that we are able to observe these regimes independent of observing platform choice.

3.) *Tropical cloud regimes identified through cluster analysis appear to occur in regular patterns whose governance are likely related to their thermodynamic and dynamic properties.* Through our 3D cluster analysis, we have identified four tropical cloud regimes whose regime-average properties suggest a self-regulating mechanism of

the tropical hydrological cycle through the interplay of radiation, moist physics, and stability of the atmosphere. These regimes have been identified through other means (such as ISCCP data and OLR analyses) as well – cluster analysis provides us with new and unique quantitative methods to assess the properties of tropical cloud regimes, and perhaps more importantly, diagnose cloud population evolution as a function of time.

4.) *Cluster analysis of large-scale cloud resolving model results provides a meaningful way of assessing both model results, and offers quantitative new tools to analyze model results compared with observations.* By comparing the cloud regimes from a large-scale cloud resolving model run to observations, we encounter important differences that underscores the challenges inherent to both modeling and observation of cloud properties, and provides the impetus for using satellite observations of the tropics in new and useful ways.

5.) *Application of the cluster analysis technique to tropical convection allows us to separate important cloud properties necessary for new retrieval techniques.* The ability to distinguish key properties of convective regimes (namely, shallow convection, deep convection, and stratiform regions) using cluster analysis greatly enhances our ability to define the aggregate latent heating properties of these regimes, especially in the context of a radar-based latent heating profile retrieval algorithm.

We hope that in this work we have addressed some of the remaining issues facing improved analysis of the tropical environment, and namely that our theory of analyzing large-scale tropical evolution by focusing on the results of the small-scale interactions between cloud populations (as viewed through the framework of cluster analysis) has perhaps given us insight into the larger-scale properties of tropical cloud regimes.

Furthermore, it is our sincere hope that this insight might be of use to other researchers who can take these results and apply them to physical models of tropical interaction.

It is clear that our understanding of the tropical atmosphere is improving, although much yet remains to be done before we can truly claim to have any mastery over the complicated mechanisms which make up our atmosphere. With the continual improvement in methodology and technology that allows us to observe our world, combined with improved computational ability to analyze the results, we look forward to continued improvements in our understanding of what may well be the most crucial element of the earth-atmosphere system.

Bibliography

- Allan, R.P. and Soden, B.J. 2007. Large discrepancy between observed and simulated precipitation trends in the ascending and descending branches of the tropical circulation. *Geophysical Research Letters* **34**: 10.1029/2007GL031460.
- Anderberg, M.R., 1973: *Cluster Analysis for Applications*. Academic Press, 359 pp.
- Arawaka, A. and W.H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment. Part I. *J. Atmos. Sci.*, **31**, 674-701
- Back, L. E., and C. S. Bretherton (2006), Geographic variability in the export of moist static energy and vertical motion profiles in the tropical Pacific, *Geophys. Res. Lett.*, **33**, L17810, doi:10.1029/2006GL026672.
- Bony, S., and J.-P. Duvel (1994), Influence of the vertical structure of the atmosphere on the seasonal variation of precipitable water and greenhouse effect, *J. Geophys. Res.*, **99**(D6), 12,963–12,980.
- Cotton, W.R., R.A. Pielke Sr., R.L. Walko, G.E. Liston, C.J. Tremback, H. Jiang, R.L. McAnelly, J.Y. Harrington, M.E. Nicholls, G.G. Carrio, and J.P. McFadden, 2003: RAMS 2001: Current status and future directions. *Meteorol. Atmos. Phys.* **82**, 5-29
- Dudhia, J., Moncrieff, M. W., So, D. W. K., 1987: The two-dimensional dynamics of West African squall lines. *Quart. J. Roy. Meteor. Soc.*, **113**, 121–146.
- Grabowski, W. W., and M. W. Moncrieff, 2004: Moisture-convection feedback in the Tropics. *Quart. J. Roy. Met. Soc.*, **130**, 3081–3104
- Grabowski, W. W., and P. K. Smolarkiewicz, 1999: CRCP: A Cloud Resolving Convection Parameterization for Modeling the Tropical Convecting Atmosphere. *Physica D*, **133**, 171–178. (*Special Issue: Predictability: Quantifying Uncertainty in Models of Complex Phenomena, 18th Annual Conference of the Center for Nonlinear Studies, Los Alamos, NM, USA, 11-15 May 1998*)
- Harrington, J. Y., 1997: The effects of radiative and microphysical processes on simulated warm and transition-season Arctic stratus. Ph.D. dissertation, Colorado State University, Fort Collins, CO, 270 pp..
- Haynes, J.M., R.T. Marchand, Z. Luo, A. Bodas-Salcedo, and G.L. Stephens, 2007: A multi-purpose radar simulation package: QuickBeam. *Bull. Amer. Meteor. Soc.*, **88**, 1723-1727.

- Heymsfield A. J., A. Bansemar, S. Lewis, J. Iaquinta, M. Kajikawa, C. Twohy, and M. Poellot, 2002: A general approach for deriving the properties of cirrus and stratiform ice cloud particles. *J. Atmos. Sci.*, **59**, 3–29.
- Hill, G.E., 1974: Factors controlling the size and spacing of cumulus clouds as revealed by numerical experiments. *J. Atmos. Sci.*, **31**, 646
- Jakob, C. and G. Tselioudis, 2003: Objective Identification of cloud regimes in the Tropical West Pacific. *Geophysical Research Letters*, **30**, #21, 2082
- Jakob, C. G. Tselioudis, and T. Hume, 2005: The Radiative, Cloud and Thermodynamic Properties of the Major Tropical Western Pacific Cloud Regimes. *J. of Climate*, **18**, 1203-1215
- Johnson, R. H., and X. Lin, 1997: Episodic Trade-Wind Regimes over the Western Pacific Warm Pool. *J. Atmos. Sci.* **54**, 2020-2034.
- Johnson, R.H., T.M. Rickenbach, S.A. Rutledge, P.E. Ciesielski, and W.H. Schubert, 1999: Trimodal Characteristics of Tropical Convection. *J. of Climate*, **12**, 2397-2418
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, Roy Jenne, and Dennis Joseph, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Met. Soc.*, **77**, 437-471
- Khairoutdinov, M. F., and D. A. Randall (2001), A cloud resolving model as a cloud parameterization in the NCAR community climate system model: Preliminary results, *Geophys. Res. Lett.*, **28**(18), 3617–3620.
- Khairoutdinov, M., D. Randall, and C. DeMott, 2005: Simulations of the Atmospheric General Circulation Using a Cloud-Resolving Model as a Superparameterization of Physical Processes. *J. Atmos. Sci.*, **62**, 2136–2154.
- Kiladis, G.N., K.H. Straub, and P.T. Haertel, 2005: Zonal and Vertical Structure of the Madden-Julian Oscillation. *J. Atmos. Sci.*, **62**, 2790-2809
- Li, Z., and H. G. Leighton, 1993: Global climatologies of solar radiation budgets at the surface and in the atmosphere from 5 years of ERBE data. *J. Geophys. Res.*, **98**, 4919–4930
- Li, Z., M. C. Cribb, and A. Trishchenko, 2002: Influence of surface inhomogeneity on solar radiative transfer under overcast conditions. *J. Geophys. Res.*, 107.

- Lilly, D. K., 1962: On the numerical simulation of buoyant convection, *Tellus* **14**, 148–172.
- Madden, R.A. and P.R. Julian, 1971: Detection of a 40-50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *J. Atmos. Sci.*, **28**, 702-708
- Madden, R. A., and P. R. Julian, 1994: Observations of the 40-50 day tropical oscillation: a review. *Mon. Wea. Rev.*, **122**, 814-837.
- Manabe, S., and R.F. Strickler, 1964: Thermal Equilibrium of the Atmosphere with a Convective Adjustment, *J. Atmos. Sci.*, **21**, 361-385
- Manabe, S., and R.T. Wetherald, 1967: Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity, *J. Atmos. Sci.*, **24**, 241-259
- Mapes, B.E. and R.A. Houze, 1993: Cloud Clusters and Superclusters over the Oceanic Warm Pool. *Mon. Weather Rev.*, **121**, 1398-1415
- Mapes, B. E., and R. A. Houze, 1995: Diabatic divergence profiles in western Pacific mesoscale convective systems. *J. Atmos. Sci.*, **52**, 1807-1828.
- Mapes, B.E., T.T. Warner, Mei Xu, and D.J. Gochis, 2004: Comparison of cumulus parameterizations and entrainment using domain-mean wind divergence in a regional model. *J. Atmos. Sci.* **61**, No. 11, pp. 1284–1295
- Menzel, W. P., W. L. Smith, and T. R. Stewart, 1983: Improved cloud motion wind vector and altitude assignment using VAS. *J. Appl. Met.*, **22**, 377-384.
- Meyers MP, Walko RL, Harrington JY, Cotton WR, 1997: New RAMS cloud microphysics parameterization. Part II: The two-moment scheme. *Atmos. Res.* **45**: 3–39
- Moncrieff M. W., 1992: Organized convective systems: Archetypal dynamical models, mass and momentum flux theory, and parameterization. *Quart. J. Roy. Meteor. Soc.*, **118**, 29–50.
- Nakazawa, T., 1988: Tropical Super Clusters within Intraseasonal Variations over the Western Pacific. *J. of the Met. Soc. Of Japan*, **66**, 823-839
- Platnick, S., M.D. King, S.A. Ackerman, W.P. Menzel, B.A. Baum, J.C. Riedi, and R.A. Frey, 2003: The MODIS cloud products: Algorithms and examples from Terra, *IEEE Trans. on Geoscience and Remote Sensing*, 41(2), 459-473.
- Riehl, H.: 1954, *Tropical Meteorology*, McGraw-Hill Book Co. Inc., N. Y. Toronto, London. 1954, 392pp.

- Rossow, W.B., and R.A. Schiffer, 1991: ISCCP cloud data products. *Bull. Amer. Meteorol. Soc.*, **71**, 2-20.
- Rossow, W.B., G. Tselioudis, A. Polak, and C. Jakob, 2005: Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures. *Geophys. Res. Lett.*, **32**, L21812, doi:10.1029/2005GL024584.
- Saleeby, S.M., and W.R. Cotton, 2004: A large droplet mode and prognostic number concentration of cloud droplets in the Colorado State University Regional Atmospheric Modeling System (RAMS). Part I: Module descriptions and supercell test simulations. *J. Appl. Meteor.*, **43**, 182-195.
- Schumacher, C. and R. A. Houze, Jr., 2003: The TRMM Precipitation Radar's view of shallow, isolated rain. *J. Appl. Meteor.*, **42**, 1519-1524.
- Schumacher, C., R.A. Houze Jr., and I. Kraucunas, 2004: The tropical dynamic response to latent heating estimates derived from the TRMM precipitation radar. *J. Atmos. Sci.*, **61**, 1341-1358
- Simpson, J., 1992: Global circulation and tropical cloud activity. *The Global Role of Tropical Rainfall*, J. S. Theon et al., Eds., A. Deepak Publishing, 77-92..
- Smagorinsky J., 1963: General circulation experiments with the primitive equations. Part I: The basic experiment. *Mon Wea Rev* **91**: 99- 164
- Stephens, G.L., P.J. Webster, R.H. Johnson, R. Engelen, and T. L'Ecuyer, 2003: Observations evidence for the mutual regulation of the Tropical Hydrologic Cycle and Tropical Sea Surface Temperature. *J. of Climate*, **17**, 2213-2224
- Stephens, Graeme L., Deborah G. Vane, Ronald J. Boain, Gerald G. Mace, Kenneth Sassen, Zhien Wang, Anthony J. Illingworth, Ewan J. O'Connor, William B. Rossow, Stephen L. Durden, Steven D. Miller, Richard T. Austin, Angela Benedetti, Cristian Mitrescu, and CloudSat Science Team, 2002: The CloudSat mission and the A-train: A new dimension of space-based observations of clouds and precipitation. *Bull. Amer. Meteorol. Soc.*, **83**, 1771-1790
- Tao, W.-K., E.A. Smith, R.F. Adler, A.Y Hou, R. Meneghini, J. Simpson, Z.S. Haddad, T. Iguchi, S. Satoh, R. Kakar, T.N. Krishnamurti, C.D. Kummerow, S. Lang, K. Nakamura, T. Nakazawa, K. Okamoto, S. Shige, W.S. Olson, Y. Takayabu, G.J. Tripoli, and S. Yang, 2006: Retrieval of Latent Heating from TRMM Measurements. *Bull. Amer. Met. Soc.*, **87**, #11, 1555-1572
- Tulich, S. N., D. A. Randall, and B. E. Mapes, 2007: Vertical-mode and cloud decomposition of large-scale convectively coupled gravity waves in a two-dimensional cloud-resolving model. *J. Atmos. Sci.*, **64**, 1210-1229.

- Walko R.L., C.J. Tremback, R.A. Pielke , and W. R. Cotton, 1995: An interactive nesting algorithm for stretched grids and variable nesting ratios. *J Appl Meteor* **34**: 994–999
- Walko RL, L.E. Band, J. Baron, T.G.F. Kittel, R. Lammers, T.J. Lee, D. Ojima, R.A. Pielke, C. Taylor, C. Tague, C.J. Tremback, P.J. Vidale, 2000: Coupled atmosphere-biophysics- hydrology models for environmental modeling. *J Appl Meteor* **39**: 931–944
- Wheeler, M., G. N. Kiladis, and P. J. Webster, 2000: Large-scale dynamical fields associated with convectively coupled equatorial waves. *J. Atmos. Sci.*, **57**, 613–640.
- Wielicki, B, R. D. Cess, M. King, D. A. Randall and E. Harrison, 1995: Mission to Planet Earth: Role of Clouds and Radiation in Climate. *Bull. Amer. Met. Soc.*, **76**, 2125-2153.
- Wilheit, T., C. D. Kummerow, and R. Ferraro, 2003: Rainfall algorithms for AMSR-E, *IEEE Trans. Geosci. Rem. Sens.*, **41**, 204-214.
- Winker, D.M., Pelon, J., and McCormick, M.P. (2006). Initial Results from CALIPSO. *ILRC 2006*, Nara, Japan.
- Wylie, D. P. and W. P. Menzel: Eight years of global high cloud statistics using HIRS. *J. Climate* ,**12** , 170-184, 1999.
- Yanai, M., S. Esbensen, and J.-H. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.*, **30**, 611–627
- Zhang, Y. S. Klein, G.G. Mace, and J. Boyle, 2007: Cluster analysis of tropical clouds using CloudSat data. *Geophys Res. Lett.*, **34**, L12813