Title
Cluster analysis of cloud properties : a method for diagnosing cloud-climate feedbacks

Author
Gordon, Neil D.

Publication Date
2008

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Cluster Analysis of Cloud Properties: A Method for Diagnosing Cloud–Climate Feedbacks

A Dissertation submitted in partial satisfaction of the Requirements for the degree Doctor of Philosophy

in

Earth Sciences

by

Neil D. Gordon

Committee in charge:

- Joel R. Norris, Chair
- Dan Lubin
- Keiko K. Nomura
- John O. Roads
- Rick Salmon
- Christopher P. Weaver
- Guang Zhang

The Dissertation of Neil D. Gordon is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2008

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to recognize the help and guidance of my advisor Joel Norris. His help over these many years has been invaluable. I would also like to thank my wife, Dora Dalton, for all her editing help, as well as the immeasurable moral support. Without that, I surely would not have been able to finish.

Chapter 3 is, in part, a reprint of material published in the Journal of Geophysical Research as Gordon, N. D., J. R. Norris, C. P. Weaver, and S. A. Klein, 2005: Cluster analysis of cloud regimes and characteristic dynamics of midlatitude synoptic systems in observations and a model. J. Geophys. Res. -Atmos., 110, D15S17, doi: 10.1029/2004JD005027. Copyright 2005 American Geophysical Union. The dissertation author was the primary author of this journal article.

VITA

2000	Bachelor of Science, University of Chicago
2000-2001	Research Assistant, Lamont-Doherty Earth Observatory
2003	Master of Science, University of California, San Diego
2008	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Weaver, C.P., J.R. Norris, N.D. Gordon, and S.A. Klein, 2005: Dynamical controls on sub-global climate model grid-scale cloud variability for Atmospheric Radiation Measurement Program (ARM) case 4. *J. Geophys. Res.*, 110, D15S05, doi:10.1029/2004JD005022.

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

Seager, R., R. Murtugudde, N. Naik, A. Clement, N. Gordon, and J. Miller, 2003: Air-sea interaction and the seasonal cycle of the subtropical anticyclones. *J. Clim.*, 16 (12), 1948-1966.

Seager, R., D.S. Battisti, J. Yin, N. Gordon, N. Naik, A.C. Clement, and M.A. Cane, 2002: Is the Gulf Stream responsible for Europe’s mild winters?. *Q. J. Roy. Met. Soc.*, 128 (586), 2563-2586.

ABSTRACT OF THE DISSERTATION

Cluster Analysis of Cloud Properties: A Method for Diagnosing Cloud–Climate Feedbacks

by

Neil D. Gordon

Doctor of Philosophy

University of California, San Diego, 2008

Joel R. Norris, Chair

Climate models are an important tool in understanding the mechanisms of the climate system and for predicting future climate. However, clouds remain poorly simulated by models. Clouds play an important role in the climate system by reducing the amount of shortwave radiation reaching the surface and by trapping longwave radiation leaving the surface. On average, this balance leads to more cooling than warming, but this is dependent on cloud type and location. Improving the understanding of clouds and how they can be simulated will lead to more accurate models and better predictions of future climate. This study builds a method whereby clouds are grouped into dynamical regimes in order to gain insight into the connection between large-scale dynamics and cloud properties. This method can also be used to understand the sources of modeling errors and help discern ways to improve these models. An additional aspect of the study addresses the lack of understanding of how the balance that clouds create between cooling from reflection of shortwave radiation and warming from absorption of longwave radiation might change as the atmosphere warms due to anthropogenic increases in greenhouse gases like carbon dioxide. This study builds a method whereby data analysis can assist in estimates of the cloud-climate feedback.

INTRODUCTION

Clouds are a critical element of the climate system. They are a physical manifestation of the water cycle, which is essential in the vertical and poleward transport of both water and energy. In perhaps their most familiar role, clouds bring us precipitation. Clouds also play an integral role in regulating the flux of radiation through the atmosphere. It is this last role on which this study will focus.

By intercepting incoming shortwave radiation from the sun and reflecting it back to space, clouds cool Earth's surface. Clouds warm the surface by absorbing longwave radiation emitted from the surface and the atmosphere, radiation that would otherwise escape to space. This subtle balance of cooling and warming is governed by the type and location of clouds. Previous studies have shown that midlatitude clouds can reflect as much as 200 W/m^2 more shortwave radiation than clear sky [*Weaver and Ramanathan*, 1997].

Global climate models (GCMs) are frequently employed to understand the properties of the climate system. A GCM divides the atmosphere into gridboxes, then a series of equations is applied to understand the motion in the atmosphere. These models are effective in that they can be altered to understand the response of the model to various forcings, simulating climate from the past or a proposed future climate scenario. A drawback of GCMs is that they are not able to explicitly simulate all motions in the atmosphere; the size of the gridbox for the model essentially determines the smallest motion simulated. All processes on scales smaller than the gridbox need to be parameterized. Clouds occur on scales both greater than and less than a gridbox, so are both explicitly resolved and parameterized. Previous studies [e.g., *Klein and Jakob,* 1999;

Norris and Weaver, 2001; *Tselioudis and Jakob*, 2002; *Lin and Zhang*, 2004] have shown the inability of GCMs to accurately simulate clouds associated with midlatitude cyclones. Some of this inability arises from the lack of subgrid-scale processes and their effect on gridbox-mean cloud properties. As a result of all of these factors, we cannot look to models to accurately understand the role of clouds in climate.

In addition to aiding the understanding of the mechanisms of the climate system, GCMs are also used to understand how anthropogenic increases in greenhouse gases are affecting the climate. However, the representation of clouds in climate models continues to be the largest source of uncertainty in simulations of future climate [*IPCC*, 2007]. Instead of relying on computer models to understand the changes to cloud properties that we would expect in a warmer atmosphere, we can turn to direct observations of clouds. We can diagnose changes by examining the sensitivity of clouds to temperature fluctuations observed in our current climate.

In approaching improved modeling of clouds, we naturally seek a representation of average cloud properties. To view the average accurately, it is essential to determine what type of clouds we observe and the state of the atmosphere that is coincident with them. A traditional method for examining cloud properties is to take averages over large spatial or temporal areas, for example, taking annual or seasonal means over zonal means or large temporal scales. This process, while necessary for analyzing clouds, is not sufficient to understand the full variability of clouds that we observe in the climate system.

Jakob and Tselioudis [2003] propose a method whereby clouds are grouped into dynamic regimes. By averaging over many cases with similar dynamics, we can understand the dynamics that are an essential part of the clouds they produce. By better understanding the dynamics and thermodynamics of the atmosphere, we can use data to better understand the sensitivity of the cloud properties to temperature. Thus a k-means clustering algorithm is used to group cloud regimes. K-means is also an effective tool for diagnosing errors in cloud properties simulated in GCMs that are specific to certain cloud regimes.

In order to improve our simulation of future climate in relation to clouds, we need a number of improvements: better understanding of the connection between large-scale dynamics of the atmosphere and the resulting cloud properties; a method for diagnosing the ability of GCMs to accurately simulate cloud properties; a process by which we can estimate the effect that a warmer atmosphere will have on cloud properties; and a method to calculate the changes in radiative flux in the atmosphere that come from changes in cloud properties. This study attempts to address all of these.

This study lays out a method whereby a k-means clustering algorithm is applied to satellite-derived cloud properties. This statistical routine builds groups of cloud scenes that all have similar properties. By looking at how these cloud properties vary within these relatively homogeneous groups, we can better understand the sensitivity of cloud regimes to changes in the atmosphere.

Chapter 1 details the clustering algorithm used in this study and its strengths and weaknesses. Chapter 2 details the k-means clustering algorithm as applied to a single location, in the central region of the United States. This region was chosen as it coincides with a dense atmospheric observation network. In addition to describing the clustering routine in detail, this study examines the ability of a single-column model to reproduce the frequency and character of the clusters found from observations, as well as the large-scale atmospheric forcing that corresponds to relatively large or small amounts of subgrid variability in cloud properties.

Chapter 3 expands on this clustering technique for a larger area, providing information about the spatial and seasonal variability of the cloud regimes observed. Additionally, we examine the effect that each cloud regime has on radiative flux in the atmosphere. Chapter 4 uses the clustering technique from the previous two chapters and lays out a technique to measure the sensitivity of these cloud regimes to changes in atmospheric temperature so as to understand how clouds are changing as the climate changes, and how such changes to clouds might mitigate or exacerbate warming imposed by an increase in atmospheric CO_2.

CHAPTER 1

Clustering is a statistical tool that allows sets of data to be grouped together based on similar properties. There are many different types of clustering routines; the one used in this study is known as k-means clustering (KMC). KMC requires the user to specify the number of groups, or clusters, of data to be analyzed. The routine then objectively determines the most efficient partitioning of the data into groups, based on the relative difference between the data points. This chapter will describe KMC, along with its strength and weaknesses.

If clustering groups data based on similar properties, an important aspect of clustering is the decision of how to measure this similarity of data. Our analysis uses satellite-derived measurements of cloud properties, which have specific values. Thus, it is natural to use the Euclidean distance calculation between the data points. In chapter 2, we deal with data that has non-uniform bounds. To correct for this, all data is normalized so that it varies between 0 and 1. The data in chapters 3 and 4 are already bounded by 0 and 1, so no normalization is needed.

Before we can begin KMC, we need to specify the number of clusters we want to examine; this number is known as k. At each time point, we have m observations, so each data point is an m-dimensional vector. The k-means clustering process began with random selection of k data points as initial seeds. For every other point in the data set, the distance is calculated to each of the initial seeds and then assigned to the initial seed it was closest to in a Euclidean sense. This creates k groups of data, each associated with the initial seed that was used to produce it. The number of elements in a cluster divided by the total number of elements is the frequency of occurrence of the cluster, and the average of all elements in the cluster is the centroid. Now, each cluster can be represented by the average values of the points that make up that cluster, its centroid. These cluster centroids are then used as new seeds to reinitialize the clustering routine. This process is repeated until the centroids converge. Depending on the random points initially chosen, the solution after convergence is non-unique. This is resolved by repeating the clustering process with different seeds a large number of times (50 tends to be sufficient), and choosing the solution with the least total variance of the points around the centroid that they are associated with. The different possible solutions are discussed in each chapter.

An important process in the use of KMC is the choice of k. An empirical method employed in chapters 3 and 4 is the calculation of correlation between the centroids in the clustering solution. The KMC routine is run to completion for each value of k starting at k=2 and increasing until the maximum correlation between the centroids exceeds 0.9, where k=l. This indicates that at least two clusters are sufficiently similar that the number of clusters chosen is too high. The best number of clusters is then chosen to be l-1, or one fewer than the solution with correlation exceeding 0.9. This process for choosing the proper number of clusters is laid out by *Rossow et al.* [2005]. This is one method used for the selection of clusters in chapters 3 and 4.

An additional method for choosing the correct number of clusters is the use of additional data sets. In this study, KMC is applied to satellite-derived data of cloud properties. This creates sets of observations that necessarily have similar cloud properties, but the purpose of clustering in this context is to find distinct cloud regimes. To do this, we compare the average meteorology as derived from the National Center for Environmental Prediction (NCEP) Reanalysis. This allows us to ensure that the clusters not only have distinct cloud properties from the satellite data, but also have distinct dynamics as derived from a separate data set, providing encouragement that the clusters found are not purely artifacts of the data but instead represent distinct dynamic regimes with a connection to the large-scale circulation of the atmosphere.

CHAPTER 2

Clouds remain the primary source of uncertainty in the modeling of current climate and predictions of future climate [*IPCC*, 2001]. Many previous studies evaluated global climate model (GCM) reliability by comparing large spatial and temporal averages of simulated and observed cloud properties, such as zonal and seasonal means [e.g., *Weare et al.*, 1996]. One shortcoming of this approach is that it can obscure the presence of compensating errors. For example, *Norris and Weaver* [2001] demonstrated that overprediction of shortwave cloud radiative forcing under conditions of ascent largely balanced underprediction of shortwave cloud radiative forcing under conditions of descent such that a realistic midlatitude ocean radiation climatology occurred in the National Center for Atmospheric Research (NCAR) Community Climate Model version 3 for the wrong reasons. Moreover, comparisons of long-term averages do not provide much information about reasons for correct or incorrect simulation of cloud properties.

A complementary method of model cloud evaluation is compositing, i.e., averaging cloud properties grouped by similar parameters such as 500-mb vertical velocity or sea-level pressure [e.g., *Bony et al.*, 1997; *Klein and Jakob*, 1999; *Norris and Weaver*, 2001; *Tselioudis and Jakob*, 2002; *Jakob*, 2003; *Bony et al.*, 2004; *Lin and Zhang*, 2004]. This technique typically focuses on the connection between cloud properties and the dynamical processes that affect them. There are, however, several disadvantages associated with compositing on dynamical parameters. One is a lack of reliable data for atmospheric variables important for cloud formation (temperature, humidity, vertical velocity) over much of the globe. Moreover, as was pointed out by *Jakob and Tselioudis* [2003], compositing by dynamical parameter requires prior knowledge about the meteorological processes associated with the particular cloud regimes to be identified.

Jakob and Tselioudis [2003] applied a clustering algorithm to International Satellite Cloud Climatology Project (ISCCP) [*Rossow and Schiffer*, 1999] histograms of cloud optical depth and cloud-top pressure to identify dominant modes of cloud variability in the tropical western Pacific. This study uses a similar procedure to determine typical cloud regimes associated with extratropical cyclones for the ISCCP gridbox centered on the Southern Great Plains (SGP) site of the Atmospheric Radiation Measurement (ARM) Program [*Ackerman and Stokes*, 2003]. Only the cool season (November-March) is examined because that is the time of year when clouds in this region are primarily produced by large-scale synoptic systems. Although our study focuses on clouds over Oklahoma and Kansas due to the dense network of ARM measurements there, the use of globally available ISCCP data allow us to assess the extent to which our results can be generalized to other regions of the world. Lidar and cloud radar retrievals from the ARM SGP site provide information about the vertical distribution of cloudiness that is not available from satellite observations. Other ARM measurements have been used to constrain the water and energy budgets for reanalyzed output from a numerical weather prediction model over approximately a $3.5^{\circ}\times3.5^{\circ}$ area, enabling relatively accurate determination of large-scale horizontal and vertical advection of temperature and moisture in a volume approximately the size of a

GCM grid column. This product, the Constraint Variational Analysis (CVA) [*Zhang et al.*, 2001; *Xie et al.*, 2004], is superior to analyses from numerical weather prediction models but is currently available only over the SGP site during January 1999 – March 2001. Meteorological parameters coincident with the satellite cloud scenes are averaged within cloud clusters to gain a better understanding of the dynamics coincident with various cloud types.

Previous studies [e.g., *Klein and Jakob*, 1999; *Norris and Weaver*, 2001; *Tselioudis and Jakob*, 2002; *Lin and Zhang*, 2004] have demonstrated that GCMs have difficulty correctly simulating clouds associated with extratropical cyclones in large part because subgrid scale processes and especially subgrid scale vertical motion are not adequately represented [*Katzfey and Ryan*, 2000; *Ryan et al.*, 2000]. One common GCM problem is production of frontal clouds that are too thick and too horizontally uniform, presumably due to lack of subgrid vertical motions that in the real world thicken clouds in one part of the gridbox and dry out clouds in another part. The first step to improving GCMs is determining which cloud regimes are most problematic in models, and this study accomplishes that by comparing observed cloud properties averaged within each cloud cluster with those from the Single Column Model (SCM) version of the Geophysical Fluid Dynamics Laboratory (GFDL) Atmospheric Model (AM2) [*GFDL Global Atmospheric Development Team*, 2004]. Since the SCM receives realistic large-scale advective tendencies from the CVA, observed and simulated cloud properties should be largely the same if the GFDL AM2 correctly parameterizes subgrid processes. A second goal of this study is the identification of large-scale meteorological conditions that are associated with small or large subgrid spatial variability in frontal clouds, which will be useful for later parameterization. This is investigated by dividing cloud clusters according to instantaneous spatial variability in cloud reflectivity and cloud top pressure in each ISCCP scene and examining the differences in CVA meteorology and advective forcing for cases of high and low cloud variability. The CVA data, however, do not provide information about meso- and smaller scale dynamical processes directly responsible for generating subgrid cloud variability. These must instead be investigated using a high-resolution model, as reported in *Weaver et al.* [2004].

Clustering of Cloud Data

The primary source of cloud observations for this investigation was the ISCCP D1 equal-area (280km x 280km) data set, originally processed from radiances measured by geostationary weather satellites [*Rossow et al.*, 1996; *Rossow and Schiffer*, 1999]. We examined the single gridbox (35-37.5ºN, 99.3-96.2ºW) most closely collocated with the ARM SGP/Cloud and Radiation Testbed (CART), centered on 36.6ºN, 97.5ºW. The entire ISCCP gridbox is located inside the boundary facilities of the SGP/CART site and the CVA domain. ISCCP data provide gridbox mean cloud fraction, cloud-top pressure and visible cloud optical thickness every three hours during daytime, usually averaged from 50-80 pixels about 4-7 km in size and spaced approximately 30 km apart. Other available data are the spatial standard deviations of pixel cloud top pressure and cloud optical thickness within the gridbox and the relative frequencies of pixels occurring in seven cloud top pressure and six cloud optical thickness intervals (42 categories). We restricted our analysis to solar zenith angles less than 72° because cloud property retrievals may be inaccurate when the sun is close to the horizon. The restriction of our analysis to day-time cases should not be significant due to the small diurnal variability in high clouds over the central US in wintertime [*Wylie and Woolf*, 2002]. The non-linear relationship between radiation flux and optical thickness was taken into account by converting cloud optical thickness values to cloud reflectivity at 0.6 microns using an ISCCP look-up table [corresponding to Fig. 3.13 in *Rossow et al.*, 1996]. Thus, our results more correctly represent cloud effects on gridbox mean visible radiation flux.

ISCCP three-hourly data for the months of November-March were grouped into cloud regimes by applying a k-means clustering algorithm to gridbox mean cloud fraction, cloud reflectivity, and cloud top pressure. The k-means procedure classifies all data elements into a specified number of clusters such that within-cluster variance is minimized [*Hartigan*, 1975]. The only arbitrary parameter needed is the number of clusters; the character of the individual cluster means is then objectively determined by the data. For reasons described in the following paragraph, we chose to calculate six clusters. Values of cloud fraction, cloud reflectivity, and cloud top pressure were all converted to a scale varying linearly from 0 to 1 to ensure each parameter would contribute equally to clustering, and times with no cloudiness at all, approximately 5% of the total, were excluded from clustering. The clustering process began with random selection of six data elements as initial seeds, each element comprising a 3-hourly mean cloud fraction, cloud reflectivity, and cloud-top pressure. Every other element in the data set was then assigned to the initial seed it was closest to in a Euclidean sense. The number of elements in a cluster divided by the total number of elements is the frequency of occurrence of the cluster, and the average of all elements in the cluster is the centroid. These cluster centroids became new seeds to reinitialize the clustering routine, which was repeated until the centroids converged.

The most subjective aspect of the k-means method is specifying the number of clusters. After examining results for various numbers, we chose to use six because that was the minimum number of clusters that had clearly distinct cloud properties and meteorological conditions. Additional clusters overlapped preceding clusters without providing appreciable new information. For example, two of the six original clusters were patchy thin cirrus and extensive thicker cirrostratus, and a seventh cluster was merely cirrus of intermediate optical thickness and horizontal coverage. Inclusion of such intermediate clusters would increase the length of the paper and the number of plots without commensurately enhancing our understanding of dynamical and thermodynamical conditions associated with particular cloud types. Another uncertainty in the k-means method is the convergence of the clustering algorithm to different results for different initial seeds. We resolved this ambiguity by clustering on 100 different sets of random initial seeds and choosing the final cluster set with the least sum of variance around each cluster centroid. Only two alternate realizations occurred, and these were substantially similar to the minimum variance cluster set. The differences entailed the addition of a cirrus-type cluster and the combination of optically thick low-top clouds with optically thick high-top clouds in a single cluster rather than placing them in separate clusters.

Our approach differs from that of *Jakob and Tselioudis* [2003] in that we cluster on three parameters (gridbox mean cloud fraction, cloud-top pressure, and cloud reflectivity) rather than on 42 parameters (cloud fraction within each of seven cloud-top pressure and six cloud optical thickness intervals). We chose to use a three-parameter gridbox mean phase space because it is simpler and can be applied to GCMs that do not produce ISCCP-like output (which has non-negligible computational and storage costs). One disadvantage of aggregation of the 42 parameters to three gridbox mean parameters is loss of information when pixels in the gridbox have widely varying cloud properties (e.g., bimodal distributions), but examination of instantaneous ISCCP scenes indicates that unimodal distributions occur 71% of the time in our domain. We define a scene as unimodal if the gridbox mean cloud-top pressure and optical thickness fall into a cloud-top pressure/optical thickness interval that is the same or adjacent to the cloud-top pressure/optical thickness interval with the most pixels. One disadvantage of clustering on 42 parameters is that each parameter is treated as being equally distant from the others, and adjacent cloud top pressure/optical thickness intervals are grouped together only if they co-occur in instantaneous ISCCP scenes. We found that clustering on 42 parameters did not produce results that were any more dynamically distinct than clustering on three parameters. Moreover, the 42-parameter method converged to a larger number of solutions for different starting seeds than did the three-parameter method.

Although the 1999-2001 time period is of greatest interest because that is when advective forcing from the CVA is available, the clustering algorithm was applied to 14 years (1988-2001) of the ISCCP data for January-March and November-December to reduce sampling uncertainties. Clusters for January 1999-March 2001 were determined by matching elements to the nearest 14-year centroid, with no iterative reclustering. The 1999-2001 centroids were nearly identical to those for the entire 14 years, but the cluster frequencies were slightly different. In order to match the temporal resolution of the CVA, January 1999-March 2001 ISCCP data were linearly interpolated from three-hourly to one-hourly before the application of clustering.

Table 2.1 lists cluster centroids for the 1999-2001 time period, ordered according to relative frequency. Conditions with completely clear sky occurred 6% of the time (in the three-hourly data) and were not clustered. Although, for convenience, we label each cluster with a cloud name, this does not imply that the name is characteristic of every element in the cluster. Pixel frequency distributions as a function of cloud optical thickness and cloud top pressure are plotted for each cluster in Fig. 2.1. The variability in cloud optical thickness and cloud top pressure seen in the histograms results from both the subgrid variability of cloud properties within individual scenes and the variability of the gridbox mean elements around the cluster centroids. The first and second clusters, “extensive cirrus” and “patchy cirrus”, are optically thin with high cloud tops, but have very different cloud fractions. Clusters 3 and 4, “frontal/nimbostratus” and “stratus/stratocumulus”, are optically thick with nearly 100% cloud cover. Although the frontal/Ns reaches into the upper troposphere, St/Sc has the lowest cloud top of any cluster. The fifth cluster is a mixture of clouds at multiple levels in the atmosphere with nearly 100% cloud cover. Examination of individual scenes indicated that clouds occur at a variety of levels at the same time, so the histogram in Fig. 2.1 is not merely an artifact of averaging. This mixed cluster does not separate into high cloud and low cloud clusters when more than six clusters are calculated, suggesting it is indeed a distinct regime. Cluster 6, “cumulus/cirrus,” is characterized by a combination of low-level and high-level optically thin clouds, and in this case the gridbox mean cloud top pressure is not representative of individual pixels.

While satellite observations describe well the horizontal distribution of cloud properties, they supply much less information about the profile of overlapping clouds. Knowledge of the vertical cloud distribution can be obtained from the millimeter cloud radar (MMCR) and ceilometer instruments located at the ARM SGP Central Facility. The MMCR measurements have been combined in the Active Remote Sensing of Cloud Layers (ARSCL) data product [*Clothiaux et al.*, 2000], which identifies the presence of cloud at 45 m vertical resolution and provides the top and bottom heights of every cloud layer directly above the instruments. ARSCL 10-sec data at 45 m resolution were averaged within 250 m vertical intervals and 30-min periods centered on the hourly time points of the CVA to obtain cloud fraction as a function of height. We assume that the frequency of cloudiness above the instrument during a 30-min interval is identical to the spatial cloud fraction within the surrounding local area (18-36 km for 10-20 m s^{-1} advection speeds). This is much smaller than the ISCCP domain (280 km), but averaging over longer time intervals can mix in substantial temporal variability in the cloud field [Appendix of *Kim et al.*, 2004]. Averages of the highest cloud top, lowest cloud base, and integrated thickness of all layers were also calculated. We determined lowest cloud base according to ceilometer measurements (30-sec sampling and 8 m vertical resolution) since they do not misidentify precipitation as cloud like the MMCR. This adjustment was not undertaken for the profiles of cloud fraction at each level, which may lead to a slight overestimate of cloud fraction, especially near the surface. At least 50% of the data in a 30-min interval were required to have valid retrievals in order to construct an average (16% of the data were missing from the entire dataset). The ARSCL 30-minute values were assigned to the same clusters that the coincident ISCCP data were classified in, and then averaged. Although cloud fractions reported by ISCCP and ARSCL may exhibit large disagreement at any given time due to differences in spatial sampling, this effect is random and will be reduced by averaging over many time points in a cluster. Systematic disagreement due to different methods of observation will remain.

Table 2.2 shows mean ARSCL cloud fraction, highest cloud top, lowest cloud bottom, and integrated cloud thickness of all layers. The cloud fraction listed in Table 2.2 is determined by the occurrence of cloud at any level. Cluster cloud top heights reported by ARSCL correspond well with those reported by ISCCP. Less agreement occurs for cloud fraction, although the relative variations between clusters are similar. One reason ISCCP might report larger cloud fraction than ARSCL is that pixels are designated as completely cloudy even if the actual cloud is smaller than the pixel size. Another possible reason is that MMCR fails to detect clouds otherwise seen by laser at heights above 8 km about 10-20% of the time [*Clothiaux et al.*, 2000]. Physical cloud thickness is not directly comparable to optical cloud thickness since the latter also depends on the condensate concentration, effective particle size, and water phase. Figure 2.2 displays average cloud fraction at every level for each cluster. Clouds in the cirrus regimes occur almost entirely above 5 km whereas St/Sc clouds are generally confined to the lowest 3 km. Frontal/Ns cloudiness is horizontally extensive and exists in a deep layer, consistent with having the largest physical and optical thickness of any cluster. The ARSCL profiles for the multilayer clusters (Cu/Ci and mixed) exhibit non-negligible cloud occurrence over a wide range of vertical levels.

Characteristic Dynamics

Dynamical parameters associated with ISCCP 1-hourly interpolated data were obtained from the Constraint Variational Analysis (CVA) [*Zhang et al.*, 2001; *Xie et al.*, 2004], a single-column analysis carried out for a domain approximately the size of a GCM gridbox centered on the ARM SGP Central Facility. The CVA constrains numerical weather prediction (NWP) model output with atmospheric soundings and measurements of precipitation, surface energy fluxes, and top-of-atmosphere energy fluxes. Column-integrated mass, water, energy and momentum are conserved by the application of objective analysis techniques. The resulting product provides vertical profiles of atmospheric conditions (horizontal winds, vertical motion, temperature, relative humidity) as well as the tendencies of temperature and water vapor due to large-scale horizontal and vertical advection. SCMs and Cloud System Resolving Models produce more realistic cloud and precipitation simulations when they are forced by advective tendencies from the CVA rather than from the original NWP output [*Xie et al.*, 2003]. CVA data are currently available at 25 mb spacing between 1000 mb and 100 mb for every hour during January 1999-March 2001.

To provide insight into the atmospheric state and advective forcing associated with the various cloud regimes, we averaged vertical profiles of CVA data over the times corresponding to each cluster. Monthly means were removed from relative humidity (RH) and temperature values prior to averaging to prevent the large basic state decline in RH and temperature with height from dominating the plots. Similarly, advective tendencies of water vapor mixing ratio were divided by the saturation mixing ratio at each level, thus converting them to tendencies in RH under the assumption that temperature remains constant. For consistency, all values of RH and saturation are with respect to liquid water even though saturation with respect to ice may be more applicable in the upper troposphere. We calculated 95% confidence intervals for the cluster means assuming a normal distribution and counting successive hours classified into the same cluster as a single realization. The smallest effective sample size for any cluster is 75 (frontal/Ns). The total number of ISCCP interpolated hourly data contributing to the clusters is 1790. Vertical profiles of perturbation RH are displayed in Fig. 2.3, pressure vertical velocity in Fig. 2.4, and advective tendencies of water vapor mixing ratio in Fig. 2.5.

The mean cloud properties of each cluster are physically consistent with the dynamical forcing. Upper tropospheric RH is higher than normal for the Extensive Ci regime (cluster 1) due to upward vertical motion near the 350 mb level that is increasing the water vapor mixing ratio over time. This positive total advection of water vapor does not occur in the patchy Ci regime (cluster 2), which instead experiences stronger downward motion and drier conditions than extensive Ci. The frontal/Ns regime (cluster 3) is associated with very strong upward motion that is rapidly increasing water vapor mixing ratio by vertical advection and producing a large positive RH perturbation from surface to tropopause. Despite the occurrence of mean ascent throughout the troposphere in the St/Sc regime (cluster 4), a positive anomaly in RH occurs only below the 600 mb level. Negative horizontal advection of water vapor overwhelms the positive vertical advection to cause net drying above the low-level clouds. Weak upward motion in the mixed cloud regime (cluster 5) produces small positive RH anomalies in the middle and upper troposphere through vertical advection of water vapor. Negative horizontal advection of water vapor dominates in the Cu/Ci regime (cluster 6), and the troposphere is anomalously dry except near the surface and tropopause. Despite the similarities between advective forcing and cloud properties described above, it is important to keep in mind that the mean meteorological conditions may not be characteristic of every element in the cluster.

Subgrid Spatial Cloud Variability

Previous studies have found that GCMs have difficulty correctly representing subgrid variability in cloud properties [e.g., *Norris and Weaver*, 2001; *Tselioudis and Jakob*, 2002]. This is particularly the case for frontal cloudiness, which typically is too uniformly high and optically thick under conditions of strong ascent. *Norris and Weaver* [2001] attributed this to the lack of representation of subgrid vertical motions in current GCM parameterizations, since even if gridbox mean vertical motion were upward, subgrid variability could result in stronger ascent in one portion of a gridbox and weak descent in another portion of the gridbox. Because GCMs currently do not consider subgrid variability in vertical motions aside from moist convective parameterizations, gridbox mean ascent tends to produce spatially uniform saturation of the entire grid column. Although the role of subgrid variability in vertical motion is difficult to investigate observationally due to lack of reliable data, high-resolution simulations of two synoptic systems passing over the SGP site indeed indicate a strong connection between the mesoscale distribution of upward motion and the mesoscale distribution of cloudiness [*Weaver et al.*, 2004]. Identification of the large-scale forcing associated with mesoscale variability in vertical motion and cloudiness will aid parameterization of these effects. Since computationally intensive simulations are available only for a few short time periods, it is useful to examine in observations how the large-scale meteorological forcing differs between cases of high and low subgrid cloud variability with the same gridbox mean cloud properties.

We carried this out by partitioning clusters into subsets based on the spatial standard deviations of cloud reflectivity and cloud top pressure of each element. The “high variability” subset of a cluster then consists of those elements whose standard deviations are above the median values of both parameters for the cluster, and the “low variability” subset consists of those elements whose standard deviations are below the median values. It happens to be the case that the spatial standard deviations of cloud reflectivity are dominated by the gridbox means because the standard deviation must be close to zero when the mean value is close to zero. For this reason we divide cloud reflectivity standard deviations by the gridbox means before partitioning the cluster. The high variability and low variability subsets each have slightly more than one quarter of the elements since subgrid variability in cloud reflectivity tends to be positively correlated with subgrid variability in cloud top pressure.

Figure 2.6 shows horizontal, vertical, and total water vapor advection for high variability and low variability subsets of the frontal/Ns regime (cluster 3). Strong ascent produces positive vertical and net positive total water vapor advection for both subsets. The horizontal water vapor advection, however, is negative for high variability cases and positive for low variability cases. The presence of positive vertical and negative horizontal water vapor advection also occurs with the high variability subsets of the St/Sc (cluster 4) and mixed (cluster 5) cloud regimes (not shown). These results suggest that substantial subgrid variability in cloud top pressure and cloud reflectivity may result from subgrid variability in vertical motion that saturates only part of a gridbox otherwise being dried by horizontal advection, although subgrid variability in horizontal moisture advection may also play a role. Contrastingly, horizontal moistening favors much more uniform saturation and cloud properties.

Subsets of high and low subgrid cloud variability are sometimes associated with completely different meteorological conditions. Fig. 2.7 shows this is the case for the St/Sc regime (cluster 4). The low variability subset resembles cold-sector stratocumulus. These clouds occur beneath subsidence that caps a shallow boundary layer and constrains cloud top height and thickness to be relatively uniform. The high variability subset resembles warm-sector stratus. Examination of individual days suggests plumes of moist ascending air advected from the subtropics form stratus clouds of varying heights and thicknesses. High and low variability subsets of the extensive cirrus regime (cluster 1) also occur in different meteorological regimes (Fig. 2.8). The low variability subset is associated with subsidence, a cold troposphere, and a depressed tropopause, presumably in the upper-level trough following the passage of a cold front. The high variability subset is associated with ascent, a warm troposphere, and an elevated tropopause, presumably in an upper-level ridge ahead of an approaching cyclone.

Global Representativeness

The global representativeness of the results at the SGP site can be assessed by measuring the proximity of locally generated cluster centroids to the SGP centroids.

We did so by calculating clusters for 1999-2001 data in each ISCCP gridbox using the 14-year (1989-2001) SGP centroids as initial seeds with no iterative reclustering. The average Euclidian distance between these centroids and the SGP centroids was computed for each gridbox with weighting by cluster frequency. The resulting values then describe how well cloud regimes around the world resemble cloud regimes at the SGP site. To provide insight into the relative importance of differences from the 14-year SGP centroid, we scaled distances from all gridboxes by the distance between the 1999-2001 SGP centroid and the 14-year centroid. Thus, a scaled distance equal to one means clusters at an arbitrary gridbox are as close to the 14-year SGP centroid as clusters calculated from three of those 14 years. Table 2.3 lists 14-year centroids and frequencies at the SGP site, and a comparison with Table 2.1 demonstrates that differences between centroids calculated over 1989-2001 and centroids calculated over 1999-2001 are small. Fig. 2.9 shows that cloud regimes over many midlatitude land regions and especially over the eastern half of the U.S. have similar properties to those at the SGP site. This suggests the atmospheric state and advective forcing documented for each SGP cluster are broadly representative of midlatitude continental cool-season cloudiness.

Model Cloud Comparison

One difficulty with evaluating the quality of GCM cloud simulation is that it is not always clear whether errors in cloud simulation result from incorrect large-scale forcing or an incorrect response of parameterizations to correct large-scale forcing.

This problem is mitigated in the examination of three-hourly SCM cloud output from runs with CVA forcing, which presumably experienced similar forcing as the observed clouds. In this study we will examine output from an SCM implementing the full physics parameterizations, vertical resolution, and time step of the GFDL AM2 [*GFDL Global Atmospheric Model Development Team*, 2004], but our procedure could be applied to any model. The GFDL SCM has 24 vertical levels and a prognostic cloud scheme based upon *Tiedtke* [1993] with stratiform microphysics from *Rotstayn* [1997] and *Rotstayn et al.* [2000]. The cumulus parameterization is Relaxed-Arakawa-Schubert [*Moorthi and Suarez,* 1992] and the turbulence parameterization is based upon *Lock et al.* [2000]. Temperature and moisture forcings specified from the CVA without nudging and the winds specified from observations drive the model. The data analyzed in this study come from hours 12 to 36 of 36-hour SCM forecasts that begin every day with the observed sounding. This minimizes the drifts of temperature and moisture that can develop in such SCM simulations [*Ghan et al.*, 2000].

As was done for the ISCCP data, the three-hourly SCM cloud data were linearly interpolated to one-hourly. Cluster by cluster comparison of observed and simulated cloud properties will establish the specific cloud regimes that are well or poorly modeled. For direct comparability with the satellite data, the SCM output had been converted into frequency distributions of "pixels" with various values of cloud optical thickness and cloud top pressure using the ISCCP simulator [*http://gcss-dime.giss.nasa.gov*; as described in *Klein and Jakob*, 1999; *Webb et al.*, 2001]. The

ISCCP simulator divided the SCM column into 50 subcolumns and randomly assigned cloud cover or clear to each level of each subcolumn such that the model overlap assumption was maintained and the fraction of subcolumns with cloud was the same as the SCM column cloud fraction at each level. The total optical thickness and cloud top pressure of each subcolumn were then calculated from the vertical distribution of condensate amount and effective particle size in a manner consistent with ISCCP retrievals of cloud properties. Each subcolumn was treated as a pixel, and the fraction of pixels in each ISCCP cloud top pressure/optical thickness intervals was calculated. Subcolumn cloud optical thickness was sometimes less than that detectable by satellite, defined in this study as optical thickness less than 0.3. These subvisible cloudy pixels were not included in calculations of mean gridbox cloud fraction, cloud optical thickness, and cloud top pressure. Conditions of solely subvisible cloudiness in the SCM column occurred 10% of the time, and completely cloudless conditions occurred 18% of the time (in the three-hourly data).

Table 2.4 lists average SCM cloud properties and Fig. 2.10 presents SCM pixel frequency distributions for the same times classified into the ISCCP clusters during 1999-2000. Although SCM data were not available for Jan-Mar of 2001, this has little impact on the overall results since differences between simulated and observed clouds are much larger than differences between averages over 1999-2000 and 1999-2001. Comparison with Table 2.1 and Fig. 2.1 shows that model clouds are much more optically thick than observed clouds. This is true even if the subvisible clouds are included in the optical thickness average. SCM cloud fraction, however, is generally less than average ISCCP cloud fraction due to the frequent occurrence of completely clear sky or subvisible cloudiness in the SCM at times when clouds were actually observed. Overprediction of cloud optical thickness and underprediction of cloud fraction are common compensating errors in GCMs. Clusters 3, 4, and 5 in Fig. 2.10 also exhibit the typical GCM behavior of producing clouds that are too optically thick and too high in the atmosphere under conditions of gridbox mean ascent (Fig. 2.4) [*Norris and Weaver*, 2001; *Tselioudis and Jakob*, 2002; *Lin and Zhang*, 2004; *Xie et al.*, 2004; *Xu et al.*, 2004; *Zhang et al.*, 2004]. The St/Sc cloud regime (cluster 4), which exhibits substantial cloudiness in the middle and upper troposphere despite net advective drying at these levels (Fig. 2.5), has the most egregious error. Since the *Tiedtke* [1993] cloud parameterization does not have a way to treat the impacts of horizontal advection differently from the impacts of vertical advection, the SCM always generates cloud water and cloud fraction whenever gridbox mean ascent occurs and gridbox mean RH is greater than 80%. Although substantial variability in cloud properties is evident in the histograms displayed in Fig. 2.10, more variability results from temporal changes in the gridbox mean and less from spatial variability within the gridbox than is the case for ISCCP. Examination of individual scenes indicated that all cloud pixels occur in a single cloud top pressure interval and single cloud optical thickness interval three times as often in the SCM than in ISCCP.

Model cloudiness might differ from coincident observed cloudiness if there were a delay in the SCM response to observed forcing or if the CVA did not correctly represent the observed forcing. To account for this possibility, we carried out a statistical evaluation of the SCM distribution of cloud properties. This was accomplished by classifying the SCM elements into their own clusters by performing the clustering routine on the gridbox mean SCM cloud properties. Table 2.5 lists the resulting SCM cluster centroids, and Fig. 2.11 displays pixel frequency distributions as a function of cloud optical thickness and cloud top pressure for each cluster. Note that the SCM clusters have a different order and frequency than the ISCCP clusters. Patchy Ci and Cu/Ci regimes occur much less frequently in the SCM than in ISCCP, presumably because the SCM too often produces clear sky instead of partial cirrus cloudiness. This underestimation still exists even if subvisible cloudiness is included in the clusters. Although the greater frequency of completely clear sky in the SCM causes the relative frequency of frontal/Ns and St/Sc regimes to increase, they still occur with approximately the same absolute frequency as in the ISCCP clusters. Although SCM cluster cloud fractions are comparable to the ISCCP clusters, the optical thickness values much larger. Fig. 2.11 indicates that most of the SCM clusters have less variability in cloud optical thickness and/or cloud top pressure than do the ISCCP clusters, consistent with the general tendency for lesser subgrid spatial variability in individual SCM scenes.

Summary and Discussion

This study demonstrates how satellite observations of midlatitude cool-season continental cloudiness can be grouped into distinct cloud regimes by application of a k-means clustering algorithm to gridbox mean cloud fraction, cloud reflectivity, and cloud top pressure. These regimes correspond to typical cloud types associated with various synoptic conditions over land during winter: extensive cirrus, patchy cirrus, frontal/nimbostratus, stratus/stratocumulus, cumulus/cirrus, and a mixture of clouds at a variety of levels. Averages of ground-based retrievals of cloud fraction, cloud height, and cloud thickness are consistent with the satellite cloud distributions and provide additional insight into the vertical structure of the cloud regimes. Consistency is also found between cloud properties of each regime and vertical profiles of meteorological parameters averaged over a domain approximately the size of a GCM gridbox and constrained by a dense network of observations to conserve column-integrated mass, water, energy and momentum. In particular, a close relationship is found between mean vertical profiles of water vapor advection, relative humidity, and cloudiness for each regime. We investigated cloud properties over the ARM SGP site since it was one of the few locations on Earth with accurate observations of water vapor advection, but our general results should be applicable to many midlatitude land regions around the globe.

A primary motivation for this study is the diagnosis of errors in GCM simulations of specific cloud regimes. To this end we carried out an analogous classification of cloud output from the SCM version of the GFDL AM2 model and compared properties of the resulting cloud regimes with those of the observations. Since the SCM was forced with realistic large-scale boundary conditions, simulated cloudiness should be similar to coincident observed cloudiness if the model correctly parameterizes subgrid processes. An alternative diagnostic method is to calculate cloud regimes separately for the SCM output and the observations and compare the mean properties of the cloud regimes. Both methods indicate cloud optical thickness is too large and completely clear sky is too frequent in the model, as is the case for other GCMs and SCMs [e.g., *Norris and Weaver*, 2001; *Tselioudis and Jakob*, 2002; *Lin and Zhang*, 2004; *Xie et al.*, 2004; *Xu et al.*, 2004; *Zhang et al.*, 2004]. The SCM appears to reproduce the correct absolute frequencies of frontal/Ns and St/Sc regimes, albeit with clouds that are too bright, but other cloud regimes that include optically thin cirrus are underproduced.

Another feature of the simulated clouds is their lack, relative to the observations, of subgrid spatial variability in cloud optical thickness and cloud top pressure, especially for the frontal/Ns regime. The GFDL SCM and other models tend to saturate the entire grid column under conditions of gridbox mean ascent, thus producing a uniform and very optically thick cloud. Possible reasons for this were explored using observed cloud and meteorological data. Division of the observed frontal/NS regime into subsets of high and low subgrid spatial cloud variability indicates high subgrid variability is associated with negative horizontal water vapor advection whereas low subgrid variability is associated with positive horizontal advection. In both subsets, vertical water vapor advection is positive, a result of large-scale ascent. This finding suggests that models do not sufficiently represent the effects of horizontal advection of dry air in a vertically moistened column. If the vertical moistening is non-uniform, due to subgrid variability in vertical velocity, then substantial subgrid variability in cloudiness results. *Weaver et al.* [2004] showed, using high-resolution simulations, that storms over this site indeed exhibited substantial mesoscale (and smaller-scale) variability in vertical motion and that this variability was strongly correlated with similar variability in clouds. More research must be carried out to identify the large-scale meteorological forcing and subgrid processes responsible for producing various observed cloud regimes and how these can be better represented in GCMs.

Table 2.1 - ISCCP mean cloud properties ± 1 standard deviation for each cluster during Jan-Mar of 1999-2001 and Nov-Dec of 1999-2000.

Cluster #	Mean Cloud Fraction	Mean Cloud-top Pressure (mb)	Mean Reflectivity	Optical Thickness	Frequency (%)	Cloud Type
1	0.82 +/- 0.14	209 +/- 74	0.11 +/- 0.06	0.9	19	Extensive Cirrus
2	0.24 +/- 0.15	190 +/- 98	0.07 +/- 0.05	0.5	17	Patchy Cirrus
3	1.0 +/- 0.03	431 +/- 116	0.68 +/- 0.11	20.9	16	Frontal/Nimbostratus
4	0.92 +/- 0.11	707 +/- 86	0.58 +/- 0.13	13.3	13	Stratus/Stratocumulus
5	0.90 +/- 0.13	432 +/- 103	0.31 +/- 0.07	3.7	12	Mixed
6	0.33 +/- 0.18	657 +/- 155	0.28 +/- 0.10	3.3	6	Cumulus/Cirrus
					6	Clear

Table 2.2 - ARSCL mean cloud properties ± 1 standard deviation for the same times as the ISCCP clusters.

Cluster #	Cloud Fraction	Cloud-top Height (km)	Cloud-base Height (km)	Integrated Cloud Thickness (km)	Cloud Type
1	0.48 +/- 0.66	9.4 +/- 5.0	5.8 +/- 3.9	3.1 +/- 2.7	Extensive Cirrus
2	0.13 +/- 0.43	9.3 +/- 5.1	6.0 +/- 3.2	3.0 +/- 2.1	Patchy Cirrus
3	0.87 +/- 0.72	8.9 +/- 3.8	1.4 +/- 1.7	5.6 +/- 3.0	Frontal/Nimbostratus
4	0.80 +/- 0.86	3.6 +/- 2.9	1.2 +/- 1.7	1.3 +/- 1.1	Stratus/Stratocumulus
5	0.70 +/- 0.81	7.5 +/- 4.2	3.6 +/- 2.7	2.9 +/- 2.3	Mixed
6	0.70 +/- 0.81	5.0 +/- 3.2	2.7 +/- 2.3	1.7 +/- 1.4	Cumulus/Cirrus

Table 2.3 - ISCCP Mean cloud properties for each cluster during Jan-Mar of 1988-2001 and Nov-Dec of 1988-2000.

Cluster #	Mean Cloud Fraction	Mean Cloud-top Pressure (mb)	Mean Reflectivity	Frequency (%)	Cloud Type
1	0.85	227	0.11	20	Extensive Cirrus
2	0.23	218	0.07	18	Patchy Cirrus
3	1.00	429	0.62	12	Stratus/Stratocumulus
4	0.94	698	0.54	15	Frontal/Nimbostratus
5	0.92	451	0.29	17	Mixed
6	0.33	660	0.25	12	Cumulus/Cirrus
				2	Clear

Table 2.4 - SCM mean cloud properties for the same times as the ISCCP cloud clusters.

Cluster #	Mean Cloud Fraction	Mean Cloud-top Pressure (mb)	Mean Reflectivity	Cloud Type
1	0.46 +/- 0.45	291 +/-273	0.27 +/- 0.23	Extensive Cirrus
2	0.28 +/- 0.40	327 +/- 288	0.29 +/- 0.24	Patchy Cirrus
3	0.90 +/- 0.26	371 +/- 186	0.74 +/- 0.24	Frontal/Nimbostratus
4	0.84 +/- 0.32	579 +/- 239	0.73 +/- 0.21	Stratus/Stratocumulus
5	0.72 +/- 0.40	388 +/- 257	0.48 +/- 0.29	Cumulus/Cirrus
6	0.45 +/- 0.44	461 +/- 333	0.41 +/- 0.31	Mixed

Table 2.5 - Mean cloud properties for clusters derived from the SCM output.

Cluster #	Mean Cloud Fraction (%)	Mean Cloud-top Pressure (mb)	Mean Reflectivity	Frequency (%)	Cloud Type
1	0.88 +/- 0.14	148 +/- 75	0.17 +/- 0.08	14	Extensive Cirrus
2	0.23 +/- 0.13	105 +/- 86	0.08 +/- 0.06	8	Patchy Cirrus
3	1.0 +/- 0.02	369 +/- 122	0.87 +/- 0.07	18	Frontal/Nimbostratus
4	0.98 +/- 0.06	795 +/- 96	0.69 +/- 0.17	16	Stratus/Stratocumulus
5	0.93 +/- 0.12	368 +/- 120	0.45 +/- 0.12	10	Mixed
6	0.37 +/- 0.21	780 +/- 110	0.61 +/-0.18	4	Cumulus/Cirrus
				30	Clear/Subvisible

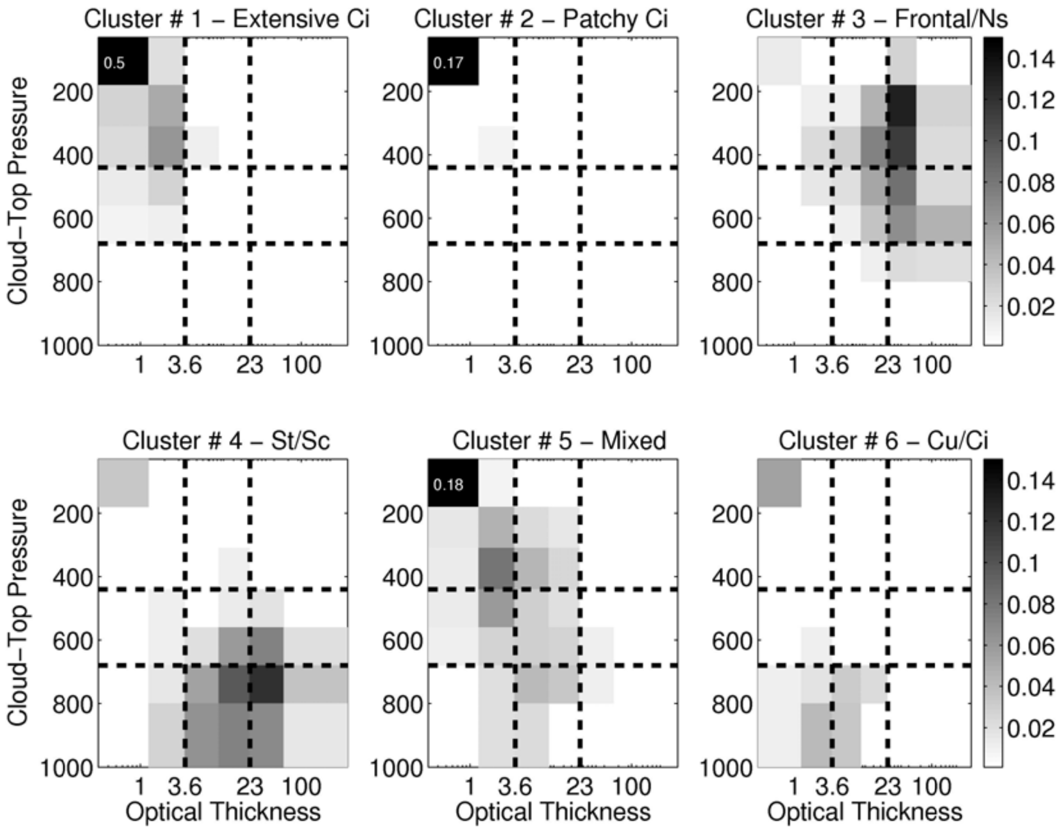


Figure 2.1 - Frequencies of ISCCP pixels in cloud-top pressure and cloud optical thickness intervals for each cluster. The dashed lines indicate the nine ISCCP standard cloud categories.

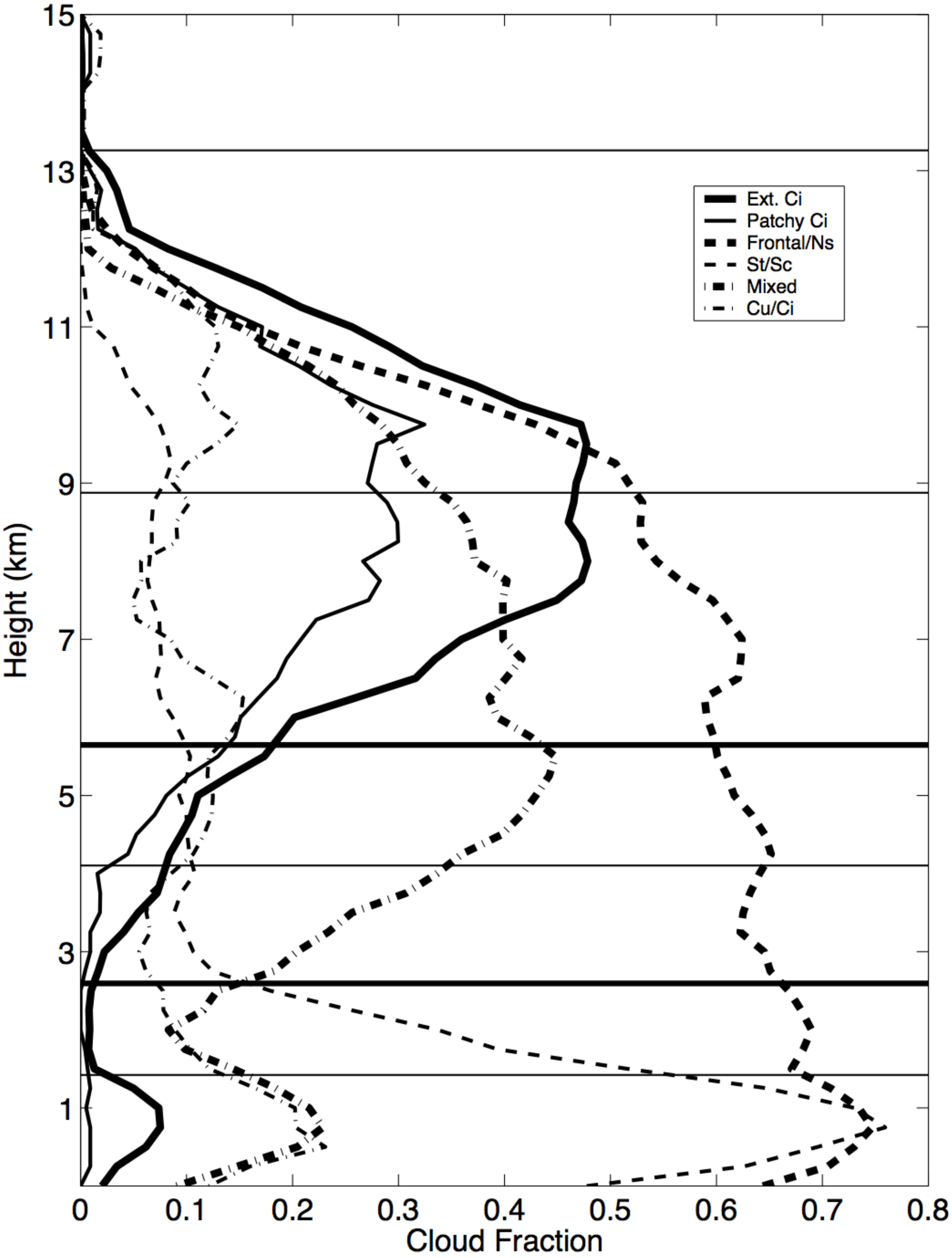


Figure 2.2 - Profiles of ARSCL mean cloud fraction for each ISCCP cluster. The horizontal lines indicate boundaries of ISCCP cloud-top pressure intervals.

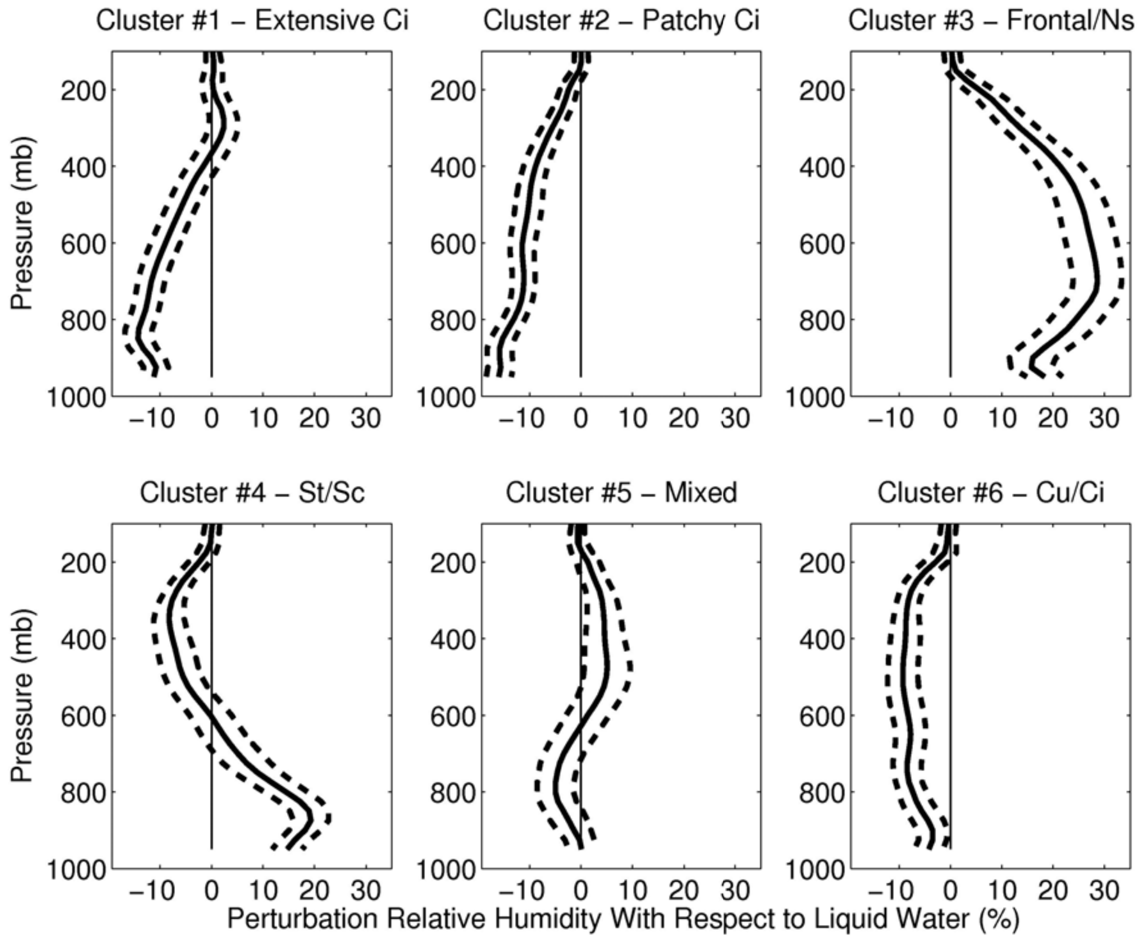


Figure 2.3 - Profiles of mean CVA perturbation RH with respect to water (%) for each cluster. The dashed lines indicate 95% confidence intervals.

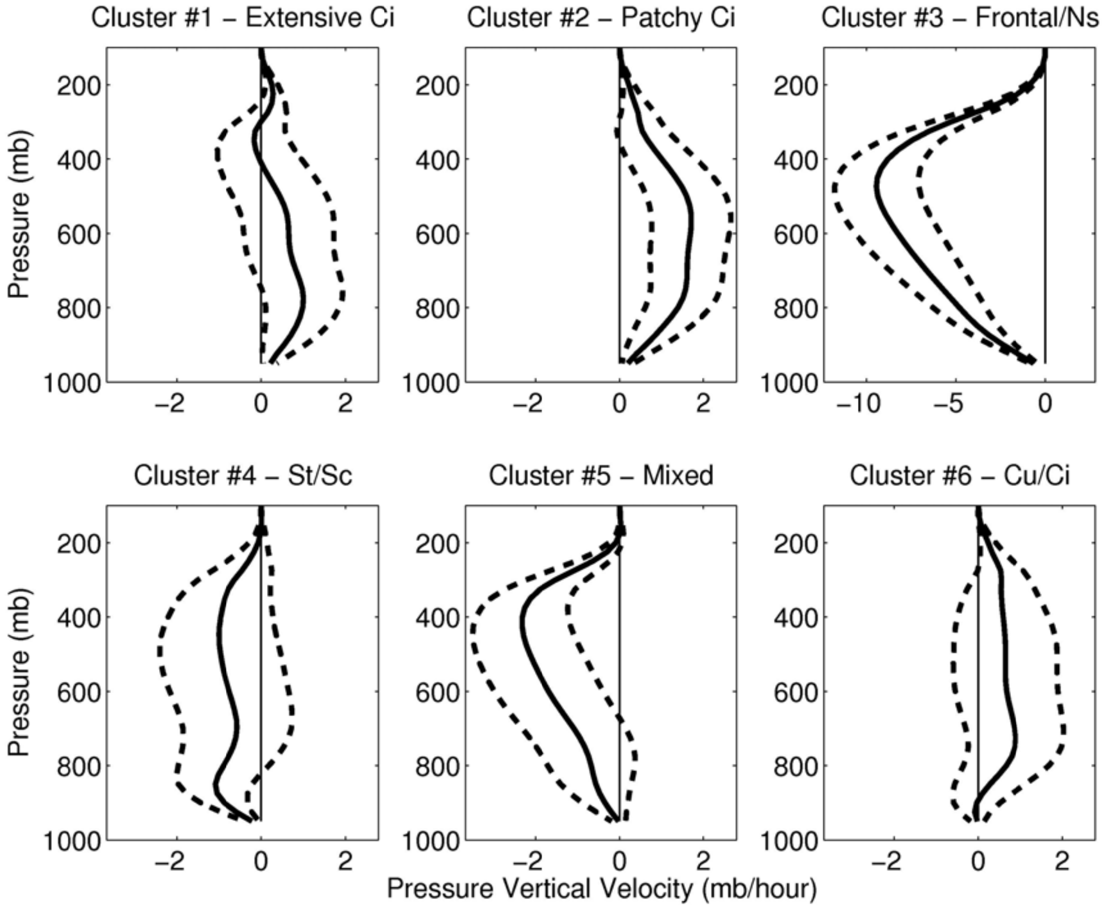


Figure 2.4 - As in Fig. 2.3, but for pressure vertical velocity (mb/hr). Negative values correspond to upward motion.

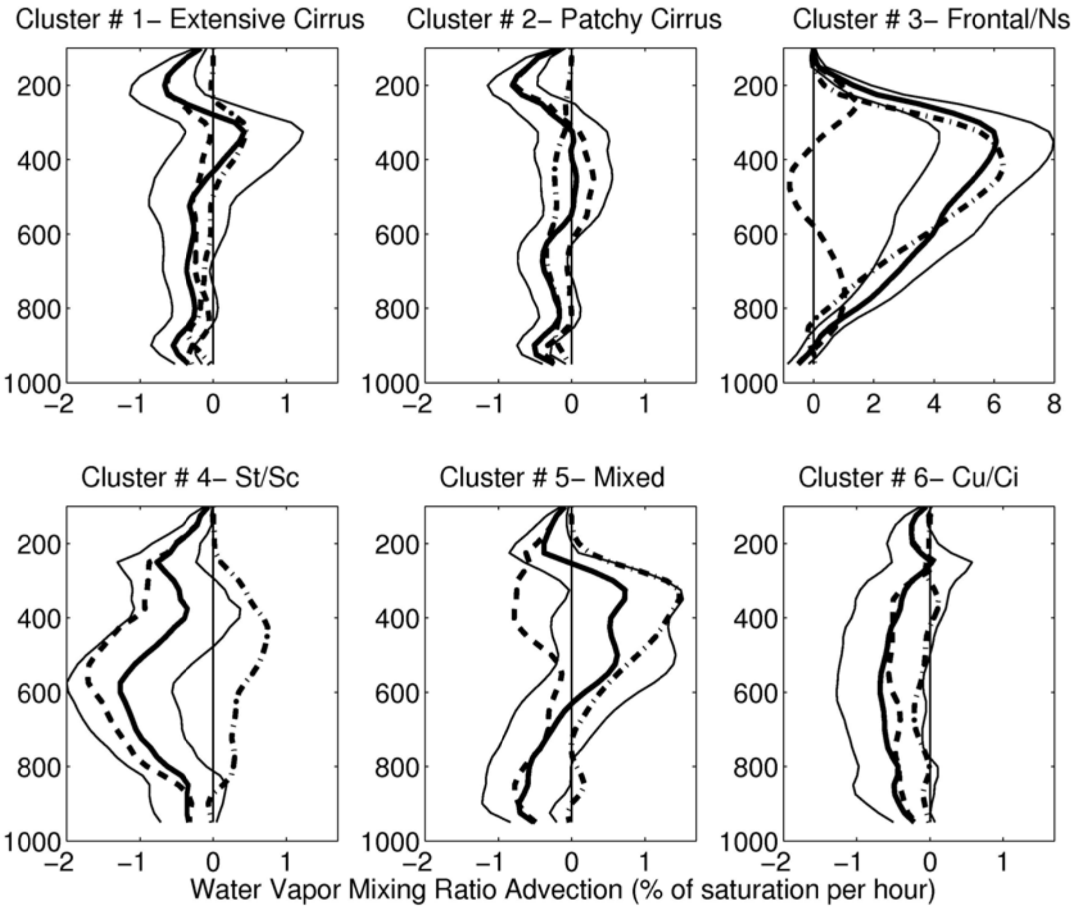


Figure 2.5 - Profiles of mean CVA horizontal (thick dash), vertical (thick dot-dash), and total (thick solid) advection of water vapor mixing ratio for each cluster. Advection values have been normalized by saturation mixing ratio at each level (units are % of saturation/hr). The thin solid lines indicate 95% confidence intervals for total advection.

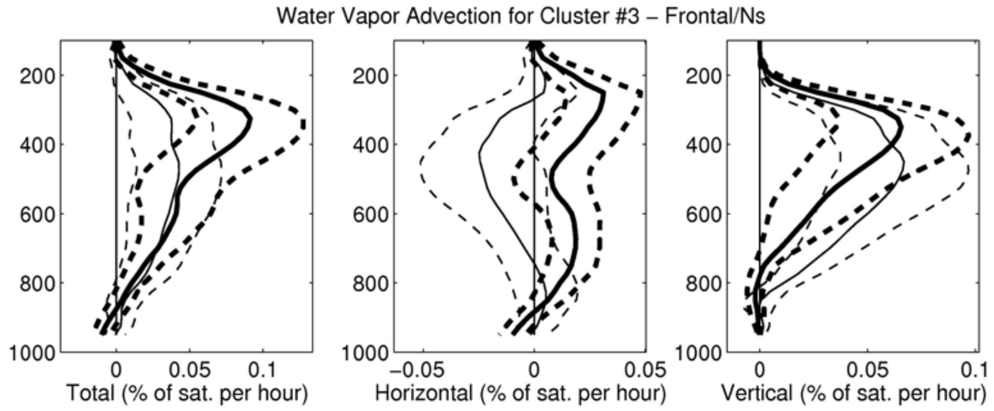


Figure 2.6 - Profiles of mean CVA total, horizontal, and vertical water vapor advection (% of saturation/hr) for subsets of high (thin line) and low (thick line) subgrid cloud variability in cluster 3 (frontal/Ns). The dashed lines indicate 95% confidence intervals

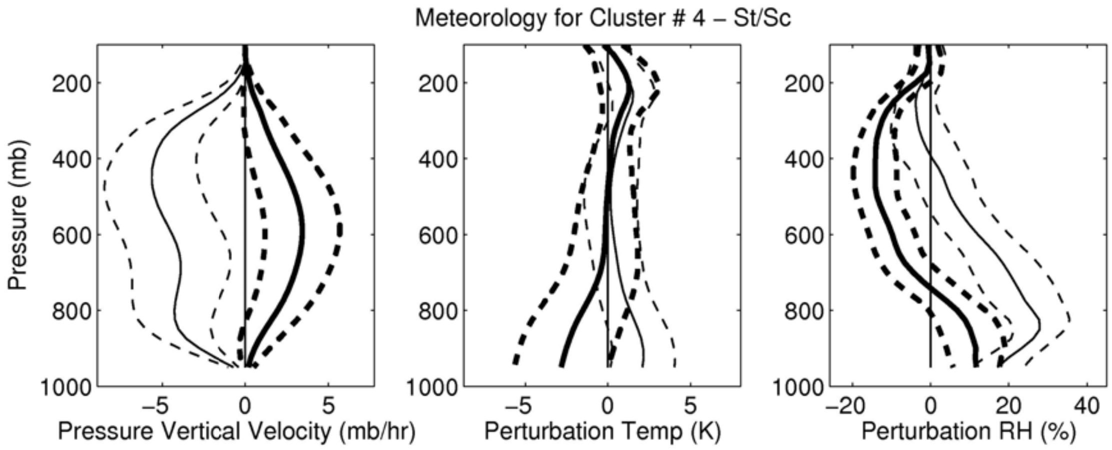


Figure 2.7 - As in Fig. 2.6, except for pressure vertical velocity (mb/hr), temperature (K) and perturbation RH (%) in cluster 4 (St/Sc).

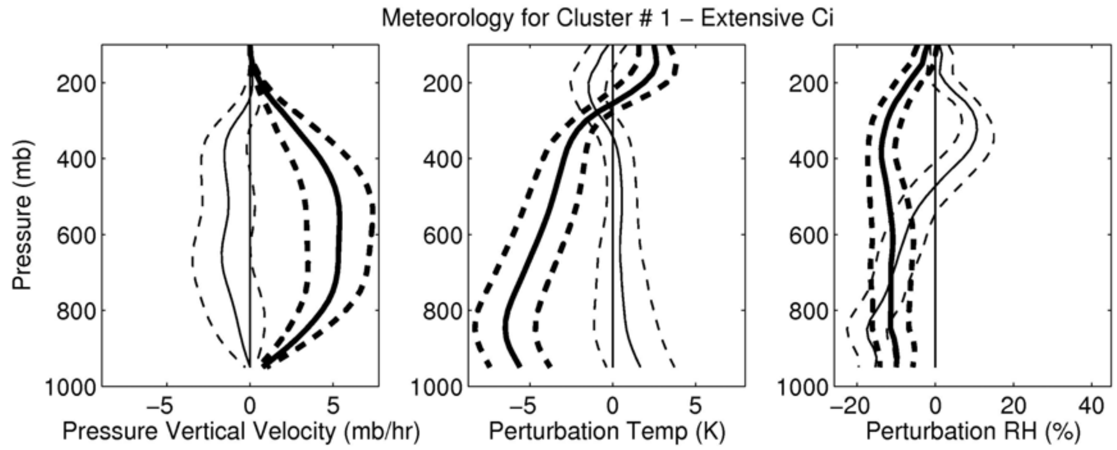


Figure 2.8 - As in Fig. 2.7, except for cluster 1 (Extensive Ci).

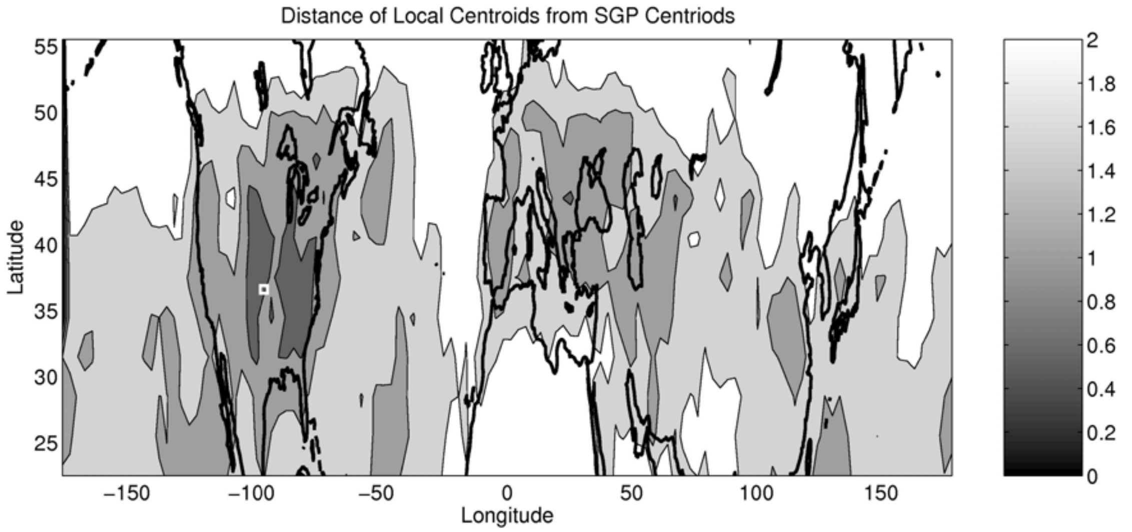


Figure 2.9 - Distances of 1999-2001 centroids at each gridbox from the 1989-2001 centroids at the SGP gridbox, scaled by the 1999-2001 SGP value. Contour interval is 0.5; values less than 1.0 are dark gray and values greater than 2.0 are white. A white square marks the SGP site.

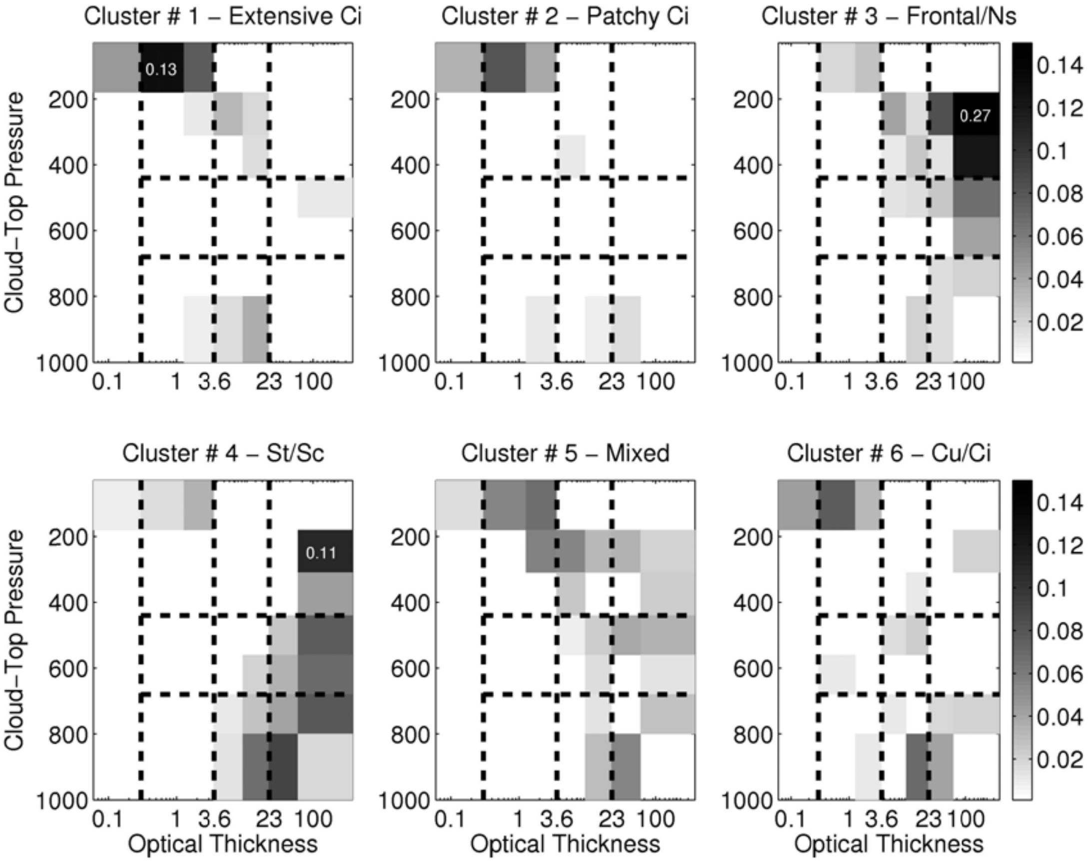


Figure 2.10 - Frequencies of SCM pixels in cloud-top pressure and cloud optical thickness intervals for each ISCCP cluster. The dashed lines indicate the nine ISCCP standard cloud categories and a special category of subvisible cloudiness (optical thickness < 0.3).

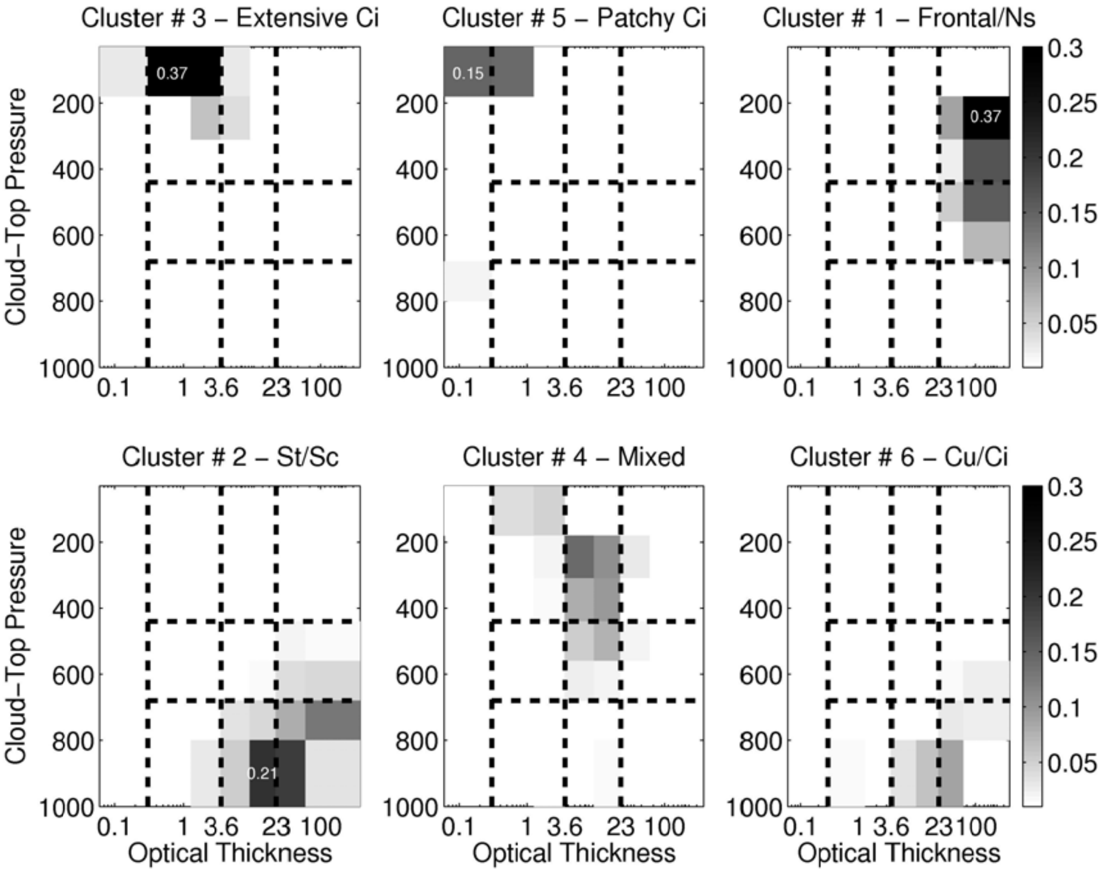


Figure 2.11 - As in Fig. 2.9, but for cluster generated from an independent clustering of SCM mean properties.

CHAPTER 3

The representation of clouds in climate models continues to be the largest source of uncertainty in simulations of future climate [*IPCC,* 2007]. The previous chapter detailed deficiencies in the ability of GCMs to accurately predict cloud properties associated with midlatitude synoptic systems, even when provided with accurate information about the large-scale forcing of a single GCM gridbox. This points to a lack of thorough understanding of the connection between large-scale forcing and cloud properties. To better simulate clouds in GCMs, we must have a better handle on the relationship between large-scale motion in the atmosphere and the clouds that these motions produce. Determining what dynamics and thermodynamics are necessary elements of a given type of cloud is crucial information to fill out our understanding of the atmosphere and to improve the ways that models simulate clouds. Past studies of model fidelity focused on large spatial and temporal scales [*Weare et al.*, 1996], but averaging across disparate dynamical regimes can mask the source of errors. Compensating errors in the simulation of cloud properties are often masked when only zonal or seasonal means of cloud properties are included. *Norris and Weaver* (2001) found that overprediction of cloud properties during instances of upward motion in the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM3) was largely cancelled out by underprediction of cloud properties with downward motion. This study, along with many others, detailed the utility of compositing, whereby cloud scenes are grouped based on a selected property of the atmosphere. Dividing the atmosphere into a series of distinct regimes, each with unique cloud properties, is an effective method for understanding the connections between the dynamics and thermodynamics of the atmosphere and the clouds they produce.

Jakob [2003] proposed compositing as the core of a strategy to improve the parameterization of clouds in Global Climate Models (GCMs). A significant limitation of compositing is that it requires prior knowledge of the dynamics of the atmosphere that are important for cloud formation. A technique based on applying statistical algorithms to cloud data, first laid out by Jakob and Tselioudis (2003), provides a more objective means of grouping clouds into distinct regimes based on their satellite-derived properties.

Our study follows this method by using a k-means clustering algorithm to group cloud scenes from the International Satellite Cloud Climatology Project (ISCCP) based on similar cloud properties. K-means clustering is an effective tool for finding patterns in large amounts of data, and has proven particularly useful for finding distinct meteorological regimes by analyzing satellite-derived cloud data [Jakob and Tselioudis, 2003; *Gordon et al.,* 2005; Jakob et al., 2005]. This study extends the clustering analysis of the previous section to all of the midlatitude ocean regions. We have restricted our study to ocean regions so as to minimize the role that surface features play in cloud forcing.

Jakob and Tselioudis (2003) proposed that clustering cloud properties into dynamical regimes would be helpful in diagnosing regime-dependent errors in the GCM simulation of cloud properties and in understanding cloud feedbacks. The following chapter addresses the latter point. The second part of this paper presents the data sets that we will be using. Section 3 contains details of the clustering algorithm used. Section 4 discusses the cloud, meteorological, and radiative properties of each cluster.

Data Sources

The source of cloud observations for this investigation was the three-hourly ISCCP D1 equal-area (280 km x 280 km) data set, originally processed from radiances primarily measured by geostationary weather satellites, with lesser contribution from polar-orbiting satellites [*Rossow et al.*, 1996; *Rossow and Schiffer*, 1999]. The ISCCP data consist of cloud fractions within a gridbox in three intervals of cloud-top pressure (CTP) and cloud optical thickness (τ), giving us nine ctp-τ categories. Each element of the data is the fraction of individual satellite pixels that exhibit those height and thickness properties. Since we analyzed the optical properties of the clouds, valid data only exists for daytime hours. We restricted our analysis to one time point per day for each satellite gridbox, choosing the value with the smallest solar zenith angle (closest to local noon). This restriction avoided biases associated with more valid data points coming from regions near the equator and from points in the summer hemisphere, where there are a greater number of daylight hours. The satellite pixels used to generate the ctp-τ histograms are approximately 4-7 km in size and spaced approximately 30 km apart, with up to 80 pixels per gridbox.

Our analysis spans nearly the entire available record of ISCCP, 21 years (1984-2004), and incorporates all ocean points between 30° and 50° in both hemispheres, representing 1,444 gridboxes. The ISCCP data consisted of nearly 10 million observations, each consisting of a CTP-τ histogram, which is a nine-dimensional vector of cloud fraction. Clear-sky observations, which are infrequent (occur less than 1% of the time), and for which all elements of the vector are exactly 0, are excluded from the clustering but retained in the analysis to examine the meteorology associated with this dynamical regime.

In addition to mean cloud properties, we examined the radiative flux data derived from the ISCCP data [*Zhang et al.*, 2004]. The flux data consists of upwelling and downwelling, shortwave and longwave radiative flux for both clear and cloudy parts of the gridbox. This data is provided at the surface, the top of the atmosphere (TOA), and at three levels within the atmosphere (680 mbar, 440 mbar, and 100 mbar). To complement the satellite-derived properties of the cloud regimes, we also analyzed data from the Extended Edited Cloud Report Archive (EECRA), which provides surface-observer reports of frequency of occurrence of different cloud types.

In order to obtain information about the dynamics and thermodynamic structure of the atmosphere, we utilized National Center for Environmental Prediction (NCEP) NCAR Reanalysis [*Kalnay et al.*, 1996]. This data set provided standard meteorological parameters as well as the advective tendencies of moisture and temperature derived from the large-scale gradients and atmospheric motions, which are important to the formation of clouds. We have restricted our analysis to midlatitudes because much of the dynamical forcing that leads to cloud formation in these regions is at or above the spatial scale of the satellite gridboxes, and the best quality reanalysis data from NCEP-NCAR is in the midlatitudes.

Cluster Analysis

This data was grouped into cloud regimes by applying a k-means clustering algorithm. The k-means procedure classifies all data elements into a specified number of clusters such that within-cluster variance is minimized [*Hartigan*, 1975; *Jakob and Tselioudis*, 2003]. The only arbitrary parameter needed is the number of clusters; the character of the individual cluster means is then objectively determined by the data. The clustering process began with random selection of k data elements as initial seeds, each element comprising a nine-element vector of cloud fraction in each height-thickness category. All other elements in the data set were then assigned to the initial seed to which they were closest in a Euclidean sense. The number of elements in a cluster divided by the total number of elements is the frequency of occurrence of the cluster, and the average of all elements in the cluster is the centroid. These cluster centroids became new seeds to reinitialize the clustering routine, which was repeated until the centroids converged.

A significant uncertainty in the k-means method is the convergence of the clustering algorithm to different results for different initial seeds. We resolved this ambiguity by clustering on 50 different sets of random initial seeds and choosing the final cluster set with the least sum of variance around each cluster centroid (the other possible solutions will be discussed later). Specifying the number of clusters is the most subjective aspect of the k-means method. After examining results for various numbers, we chose to use seven clusters, as that was the minimum number of clusters that had clearly distinct cloud properties and meteorological conditions. Additional clusters overlapped preceding clusters without providing appreciable new information; inclusion of such intermediate clusters would have increased the number of plots without commensurately enhancing our understanding of dynamical and thermodynamical conditions associated with particular cloud types.

Our approach differs from that of *Gordon et al.* [2005] in that we cluster on the ISCCP histograms as opposed to the gridbox mean cloud fraction, cloud-top pressure, and cloud reflectivity. Ours is similar to the approach used by *Jakob and Tselioudis* [2003], except that instead of using 42 parameters (cloud fraction within each of seven cloud-top pressure and six cloud optical-thickness intervals), we average these 42 parameters to nine to correspond with the ISCCP-defined cloud types. The 42 parameters do not provide significantly more information than the nine parameters chosen.

Table 3.1 lists the mean properties of the cluster centroids for the 1984-2004 time period, ordered according to relative frequency. The nonlinear relationship between radiation flux and optical thickness was taken into account by converting cloud optical thickness values to cloud reflectivity at 0.6 microns using an ISCCP look-up table [corresponding to Fig. 3.13 in *Rossow et al.*, 1996]. The mean reflectivity was then converted back to cloud optical thickness using the same table.

Thus, our results more correctly represent cloud effects on grid-box-mean visible radiation flux.

Figure 3.1 displays the mean ISCCP histograms for each of the seven clusters. Cloud fraction in each category of the mean histograms that is less than 2% is not displayed. The first three clusters are all predominantly low clouds; cluster #4 is the only midlevel cloud cluster and the last three clusters are all high clouds. The justification for the names given to each cluster in Table 3.1 will be described in the following sections.

Ramanathan et al. [1989] defined cloud forcing as the difference in the radiative flux between cloudy areas and cloud-free areas:

$$SWCRF = (\uparrow SW_{clear}^{TOA} - \uparrow SW_{all-sky}^{TOA})$$
$$LWCRF = (\uparrow LW_{clear}^{TOA} - \uparrow LW_{all-sky}^{TOA})$$

The ISCCP cloud data we used are from the points closest to local noon, the time of day when downwelling shortwave radiation at TOA and the upwelling longwave radiation at the surface are diurnal maximums and have large variability over season and location. In order to avoid a radiative weighting that would bias towards summertime and low-latitude points, we normalize our radiation parameters by the downwelling SW at TOA for SWCRF and by the upwelling LW radiation at the surface. Thus our cloud radiative forcing parameters become:

$$SWCRF_{norm} = (\uparrow SW_{clear}^{TOA} - \uparrow SW_{all-sky}^{TOA})/(\downarrow SW_{TOA})$$
$$LWCRF_{norm} = (\uparrow LW_{clear}^{TOA} - \uparrow LW_{all-sky}^{TOA})/(\uparrow LW_{surface})$$

To get a diurnal average of the cloud forcing in units of W/m^2, as opposed to a noontime value, we multiplied the normalized cloud forcing by the diurnally averaged value of downwelling shortwave flux at TOA for SWCRF$_{norm}$ or the upwelling longwave flux at the surface for LWCRF$_{norm}$. Both of these values were determined by averaging the ISCCP flux data for all three-hourly data points during a day, yielding:

$$SWCRF_{diurnal} = (\downarrow SW_{TOA}^{diurnal}) * (\downarrow SW_{TOA} - \downarrow SW_{TOA}^{clear}) / (\downarrow SW_{TOA})$$
$$LWCRF_{diurnal} = (\uparrow LW_{surface}^{diurnal}) * (\downarrow LW_{TOA} - \downarrow LW_{TOA}) / (\uparrow LW_{surface})$$

The values for these properties are listed in Table 3.2. Since the parameters are calculated as cloud minus clear, the SW CRF are positive, even though they represent cooling. The converse is true for the LW parameters. The clusters with the largest SWCRF$_{diurnal}$ and the largest negative LWCRF$_{diurnal}$ are those that are vertically extensive, namely the Weak and Strong Frontal clusters. The Cirrus cluster represents where the cooling from reflected solar radiation is nearly cancelled out by the trapping of longwave radiation from the surface. Averaging these properties, weighting by their relative frequencies, resulted in an estimate of the cloud radiative forcing of all midlatitude clouds. Our calculation found a net cooling of -39 W/m^2, which is a slightly greater cooling effect than was found globally by other studies (-17 W/m^2 by *Ramanathan et al.*, 1989; and -27 W/m^2 by *Ardanuy et al.*, 1991).

In addition to the satellite-derived cloud properties, we also looked at average cloud properties derived from the EECRA [*Hahn and Warren*, 1999]. This data, which is comprised of frequency of occurrence of cloud types as reported by a surface observer. Over the ocean, this data is primarily provided by merchant ships, therefore data over southern hemisphere oceans is sparse. As a result, the mean surface observations in Table 3.3 are for the northern hemisphere points only. This should not bias the results appreciably, since no cluster is primarily restricted to the southern hemisphere, and mean cloud properties and dynamics are similar for each cluster in either hemisphere (not shown).

Table 3.3 shows the frequency of occurrence for 11 different primarily low-cloud categories for all clusters, and average low-cloud and total cloud amount. Each cell of the table shows the mean value for that cluster and the anomaly over the mean for all clusters. The relative values of the surface observations correspond with the satellite-derived cloud properties in a gross sense, but because of the different spatial scale and method of observation, we would naturally expect there to be differences. The surface observations are particularly useful in distinguishing the differences between the low-cloud clusters and providing information on cloud extent underneath thick clouds.

The Small Cu cluster shows the highest occurrence of clear-sky observations, 6.3% of the time, more than twice as frequent as for other clusters. This cluster also has a higher than average occurrence of small cumulus (18.5%) as well as moderate and large cumulus (16.8%). The Large Cu cluster has an even larger enhancement in this category (18.6 %), with large amounts of mixed cumulus and stratocumulus (22.5%) and ordinary stratocumulus (18.6%). The Sc/St cluster is fairly evenly divided between mixed cumulus and stratocumulus (20.2%), ordinary stratocumulus (20.6%), and fair-weather stratus (15.4%). There is an increase in low cloud amount going from cluster #1 to #3 in both the surface observations and the satellite data (Table 3.3). The large low-level cloud amount for the Deep As (80.9%), Weak Frontal (79.9%), and Strong Frontal clusters (86.4%) suggests that these clouds most often extend down to the surface, even though the satellite retrievals are not able to provide that information.

Characteristic Dynamics

Dynamical parameters associated with ISCCP data were obtained from the NCEP-NCAR Reanalysis, a global gridded product available in six-hour increments. To provide insight into the atmospheric state and advective forcing associated with the various cloud regimes, we averaged vertical profiles of NCEP data over the times corresponding to each cluster. Monthly means for each spatial point and each vertical level were removed from all meteorological parameters to prevent spatial and seasonal biases from affecting the profiles. The tendencies of water-vapor mixing ratio are converted to tendencies in relative humidity (RH) by dividing by the saturation mixing ratio at each level. For consistency, we chose all values of RH and saturation with respect to liquid water even though saturation with respect to ice may be more applicable in the upper troposphere. Additionally, the meridional wind for all points in the southern hemisphere was multiplied by -1 before averaging so that horizontal flow is in a poleward sense. Vertical profiles of perturbation RH are displayed in Fig. 3.2, temperature in Fig. 3.3, pressure vertical velocity in Fig. 3.4, horizontal advective tendencies of water-vapor mixing ratio in Fig. 3.5, and horizontal advective tendencies of temperature in Fig. 3.6. Profiles of vertical advection tendencies of both temperature and water vapor (not shown) are similar to the profile of vertical motion (water vapor being the opposite sign).

The mean cloud properties of each cluster are physically consistent with the dynamical forcing. The low cloud clusters (1-3) occur in regions of gridbox mean descent, with relatively high moisture confined to the lower troposphere. Cluster 1 has the weakest average dynamical forcing, with near-mean profiles in temperature (Fig. 3.3) and both zonal and meridional wind (not shown), as well as horizontal advection of moisture (Fig. 3.5) and horizontal advection of temperature (Fig. 3.6). Clusters 2 and 3 have very similar dynamics, with the exception of their temperature profiles. Between the two clusters, there were different temperature reversals near 300 mbar. In the Large Cu cluster, the temperature profile switches from anomalously cold below 300 mbar to anomalously warm above that level. This temperature profile indicates a depressed tropopause. The Sc/St cluster, on the other hand, had an opposite temperature reversal—warm below and cold above 300 mbar, suggesting a slightly elevated tropopause (Fig. 3.3). Additionally, there is a temperature inversion in the lower troposphere for this cluster.

The Weak and Strong Frontal clusters exhibit signs of being between the trough and the ridge in a midtropospheric synoptic wave. Both have strong upward motion (Fig. 3.4) and a very moist troposphere (Fig. 3.2), as well as a relatively southwesterly flow (in a northern hemisphere sense) near the tropopause (not shown). Both frontal clusters exhibit cloud cover extending down to the surface (table 3.3), suggesting that the Weak Frontal cluster is not solely high-level clouds advected off from the Strong Frontal cluster.

The profiles of dynamics for the Deep As cluster are similar in character but smaller in magnitude than clusters 6 and 7. The RH profile for the Ci cluster shows significant moisture above 500 mbar (Fig. 3.2) and the temperature profile (Fig. 3.3) suggests that it coincides with an elevated tropopause. The large positive horizontal moisture advection (Fig. 3.5), suggests that some of these clouds are formed as blow-off from deep convective clouds, while the small upward motion in the upper troposphere suggests that some of these clouds may be dynamically generated in the upper atmosphere (Fig. 3.4).

Figures 3.7a-g show the spatial distribution of annual mean frequency of each cluster. The most frequent cluster (Small Cu) is restricted to the equatorward and coastal regions of our analysis (Fig. 3.7a) and has the lowest cloud fraction of the seven clusters. The histogram (Fig. 3.1) shows that this cluster is predominantly low clouds, but the mean cloud-top pressure of less than the other low clouds implies that small amounts of high clouds are mixed in. The second cluster (Large Cu) is more extensive than the first, occurs more often in the center of the ocean basins, and is more frequent in the southern hemisphere (Fig. 3.7b). The final low-cloud cluster (Sc/St) has a very distinctive geographical distribution. The region of highest frequency is the subtropical anticyclone region in the eastern Pacific Ocean. Other regions of frequent Sc include the northern Pacific Ocean and off the west coast of

Australia (Fig. 3.7c). Other anticyclone regions that frequently have Sc are too far equatorward to be included in our analysis.

The only predominantly midlevel cluster (Deep As) has extensive cloud cover and is primarily located in the higher latitude regions of the analysis domain (Fig. 3.7d). These clouds tend to be restricted to higher latitudes and have an increased frequency in northern hemisphere summer (Fig. 3.11), suggesting that these are weakly forced synoptic storms. From surface observations, the low-cloud frequency for this cluster is 81%, suggesting that the clouds extend to the surface and are frequently precipitating (Table 3.3). The cirrus cluster (Ci) has the smallest cloud reflectivity and several regions of occurrence. The first is in the western Atlantic Ocean, seemingly high-level blow-off from decayed synoptic storms crossing the Americas. The other region of increased frequency is in the central Pacific, possibly the cirrus advected from the deep convective towers of the west Pacific equatorial warm pool (Fig. 3.7e). The final two clusters (Weak Frontal and Strong Frontal) are fixtures of the storm track, with the Strong Frontal cluster more focused in the western half of the ocean basins (Figs. 3.7f-g). These two clusters have similar mean cloud fraction and cloud-top pressure, but the Strong Frontal cluster has much higher optical thickness (Table 3.1).

As well as looking at the spatial distribution of each cluster, we can also compare the relative location of each cluster. Figure 3.8 shows the relative frequency of clusters in the region surrounding the Strong Frontal cluster. Each box shows the average frequency of occurrence in a Around the Strong Frontal cluster, the location of the other clusters is consistent with their placement in the midlatitude synoptic wave. The Large Cu cluster is most frequently found to the west of the Strong Frontal, while the Ci cluster is predominantly to the east and northeast (Fig. 3.8).

As mentioned previously, our clustering algorithm has the ability to converge to a different solution, depending on the initial seeds provided. We resolved this by taking the case with the smallest total variance. Other than the set with the lowest sum of variance, there are two sets of clusters to which the solution can converge. The only difference is the inclusion of either an additional low-level or midlevel cluster, both with the loss of one of the Frontal clusters. In analyzing clustering results for values of k higher than seven, we often found cases with more than three low-cloud or more than one midlevel cluster. In both of these instances, the inclusion of the additional cluster did not improve the analysis, as the cluster with intermediate cloud properties also exhibited intermediate meteorological properties.

Seasonal Cycle

In order to gain a more in-depth understanding of the dynamics of each cluster, it is useful to examine the seasonal cycle of each cluster’s spatial distribution. In cluster 1, the spatial distribution of each season is nearly identical to that of the annual mean (not shown). Figs. 3.9a and 3.9b show the spatial distribution of the Large Cu cluster for the December-January-February (DJF) and June-July-August (JJA) seasons, respectively. In DJF, the clouds are predominant in the North Pacific, with extensive occurrence in the western half, suggesting that these clouds are the result of cold air advecting over warmer water behind a frontal system. There is a marked reduction in these clouds in the North Atlantic off the east coast of North America, possibly due to the warmer water of the Gulf Stream.

The Sc/St cluster has a very strong seasonal cycle (Fig. 3.10a-b). In DJF, this cluster is primarily restricted to the southern hemisphere storm track region, most likely associated with post-frontal Sc clouds (Fig. 3.10a). In the boreal summer, these clouds dominate the north and east Pacific Ocean and, to a slightly lesser extent, the north and east Atlantic Ocean. Sc clouds in the eastern Pacific anticyclone region are most extensive at this time of year (Fig. 3.10b).

For the most part, the distributions of clouds in the Deep As and Ci clusters for each season are fairly similar to the mean distributions (not shown). The exception for the As cluster is for JJA in the North Pacific (Fig. 3.11). The latter is evidence that these clouds are associated with relatively weak frontal storms that are confined to higher latitudes and most frequent in the summer months. For the Ci cluster, there is a region of enhanced frequency in the western North Pacific Ocean near the southern extent of our domain in JJA (Fig. 3.12). These high clouds may be the result of greater convection in the western equatorial warm pool, leading to more blow-off from the convective towers.

Conclusion

This study demonstrates how clouds can be grouped into distinct regimes based on a k-means clustering algorithm applied to satellite-derived cloud data. The data used for this study were cloud fraction in three bins of cloud-top pressure and three bins of cloud optical thickness. Atmospheric dynamics, derived from NCEP-NCAR Reanalysis, allowed us to consider these cloud clusters as part of the synoptic environment in which they reside. Our findings were also consistent with the spatial distribution of each of the clouds, allowing us to label these cloud clusters as forced by the synoptic meteorology.

The deep, thick cloud clusters (6 and 7) exhibit dynamics indicative of being part of a synoptic wave: strong ascent, southwesterly flow, a moist troposphere, and an elevated tropopause. Additionally, the frequency of occurrence of these cloud clusters indicates that they occur in the storm-track region and preferentially in the winter season. The low cloud clusters (1, 2, and 3), with weak mean descent, moist lower troposphere, and weak forcing, represent the dynamics that would be associated with low clouds over the ocean. The unique spatial distribution of these cloud clusters reinforces the interpretation of their dynamics, whether it is the prevalence of low clouds in the subtropical anticyclone region of the Pacific Ocean or in the north and west Pacific during winter months.

Clustering provides a tool for examining large amounts of data and extracting information about patterns within the data. The clustering of cloud properties can group observations with similar large-scale dynamics, an effective method of diagnosing the ability with which GCMs will accurately predict cloud properties. Instead of requiring that models reproduce zonal or seasonal means, we can learn whether a model can recreate the observed cloud clusters with the same frequency and spatial distribution as is observed in nature. Also, if certain clusters are not simulated well, we can compare the model-simulated dynamics for a given cluster to the observed dynamics to better determine the source of a model’s deficiency.

Additionally, since within each cluster there are relatively uniform dynamics, we can determine the thermodynamic change in cloud properties due to surface warming (Part II). By keeping dynamics constant, we are able to obtain an estimate of the partial derivative of cloud properties with surface temperature from observations rather than from computer models.

Table 3.1 – Mean ISCCP properties for each cluster.

	1 - Small Cu	2 - Large Cu	3 - Sc/St	4 - Deep As	5 - Ci	6 - Small Frontal	7 - Large Frontal
Frequency	0.275	0.184	0.165	0.14	0.113	0.077	0.043
Mean CF	54.1	77.8	92.9	97.5	87.4	99.0	99.4
Mean CTP	658.2	781.0	776.4	584.3	431.8	382.6	347.6
Mean Tau	3.63	2.89	7.19	8.30	2.30	8.90	23.08
emiss_cld	0.905	0.847	0.991	0.997	0.834	0.999	1.000
CTT	267.9	272.4	274.4	259.7	246.8	240.8	236.7

Table 3.2- Mean ISCCP flux data for each cluster.

	1 - Small Cu	2 - Large Cu	3 - Sc/St	4 - Deep As	5 - Ci	6 - Small Frontal	7 - Large Frontal
SWCRF	-39.01	-40.46	-96.96	-112.89	-55.04	-123.05	-168.38
LWCRF	14.81	10.32	13.17	40.59	46.71	78.26	87.02

Table 3.3 – Mean surface-reported cloud properties for northern hemisphere points.

Obsevation		Cluster #						
		1	2	3	4	5	6	7
Clear-sky Frequency	Mean	6.8	2.2	1.1	0.4	3.5	0.4	0.2
	Anom	3.7	-0.9	-2.0	-2.6	0.4	-2.7	-2.9
Obscured-sky Frequency	Mean	2.2	2.1	10.5	9.3	1.8	9.1	12.5
	Anom	-3.3	-3.4	5.0	3.8	-3.8	3.6	6.9
Total Cloud Amount	Mean	58.2	68.5	87.1	90.7	67.4	92.6	95.9
	Anom	-16.5	-6.3	12.4	15.9	-7.4	17.9	21.1
Low-level Cloud Amount	Mean	47.2	59.7	79.9	80.9	48.7	79.9	86.4
	Anom	-16.6	-4.1	16.1	17.1	-15.1	16.1	22.6
Rain and Snow Frequency	Mean	4.2	6.4	4.6	11.5	3.0	13.1	30.1
	Anom	-3.3	-1.1	-2.9	4.0	-4.5	5.6	22.6
No-low-cloud Frequency	Mean	16.3	6.4	4.9	4.7	17.8	5.6	3.9
	Anom	6.2	-3.7	-5.3	-5.4	7.7	-4.5	-6.2
Small Cumulus Frequency	Mean	18.5	13.7	5.1	4.5	15.7	4.7	2.5
	Anom	7.1	2.3	-6.3	-6.9	4.4	-6.7	-8.8
Moderate and Large Cumulus Frequency	Mean	16.8	18.6	6.9	6.5	14.1	5.8	3.8
	Anom	4.7	6.5	-5.3	-5.7	1.9	-6.4	-8.3
Mixed Cumulus and Stratocumulus Frequency	Mean	16.9	22.5	20.2	19.3	18.8	18.1	12.3
	Anom	-1.8	3.7	1.5	0.6	0.1	-0.6	-6.4
Ordinary Stratocumulus Frequency	Mean	9.6	18.6	20.6	18.3	12.1	16.7	13.7
	Anom	-4.4	6.5	6.6	4.4	-1.8	2.8	-0.3
Fair-weather Stratus Frequency	Mean	5.5	5.8	15.4	15.7	6.3	16.2	17.9
	Anom	-4.6	-4.3	5.2	5.6	-3.9	6.1	7.8
Bad-weather Stratus Frequency	Mean	6.1	7.5	11.4	16.0	6.2	18.5	28.0
	Anom	-4.5	-3.2	0.8	5.4	-4.4	7.9	17.3

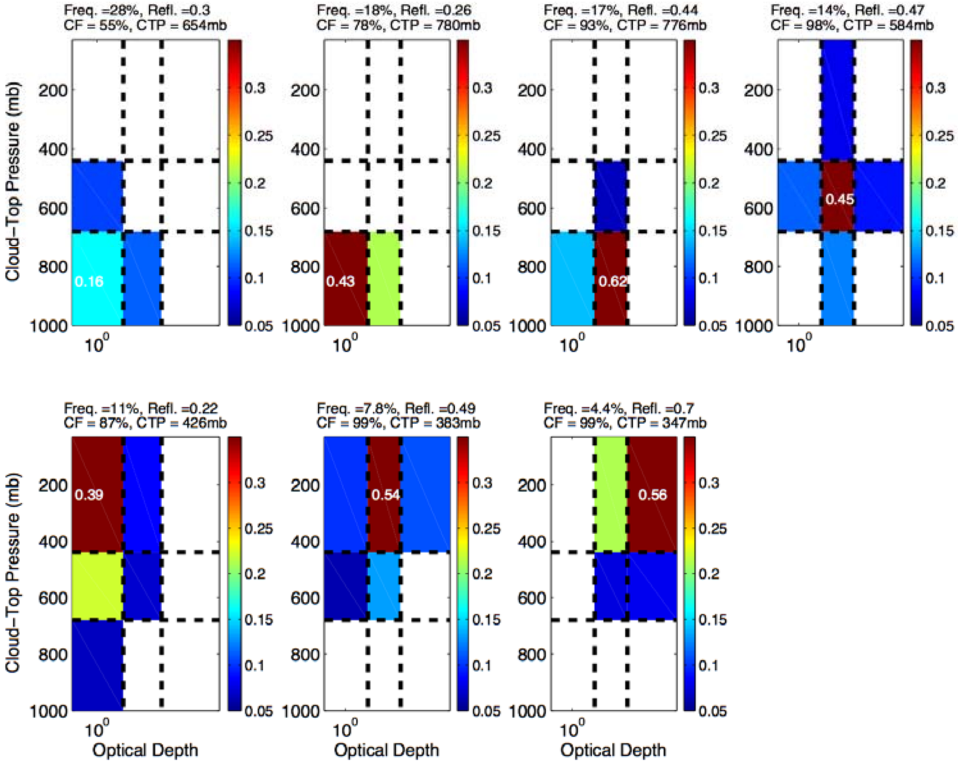


Figure 3.1 – Mean ISCCP histograms of cloud-top pressure and optical thickness.

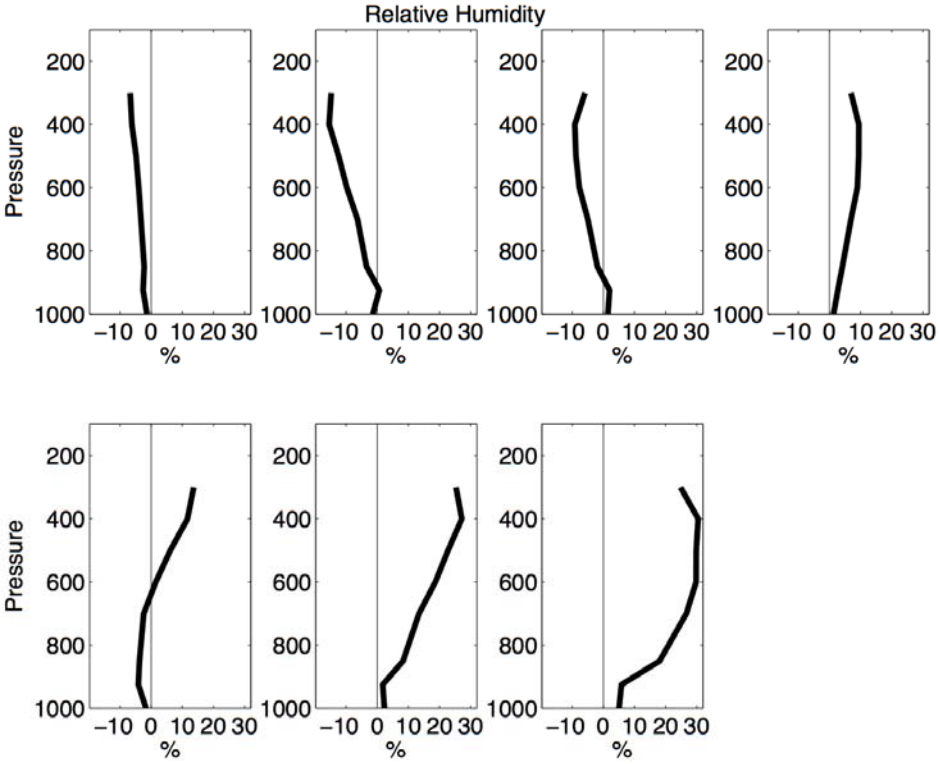


Figure 3.2 – Profiles of mean relative humidity for each cluster from NCEP Reanalysis.

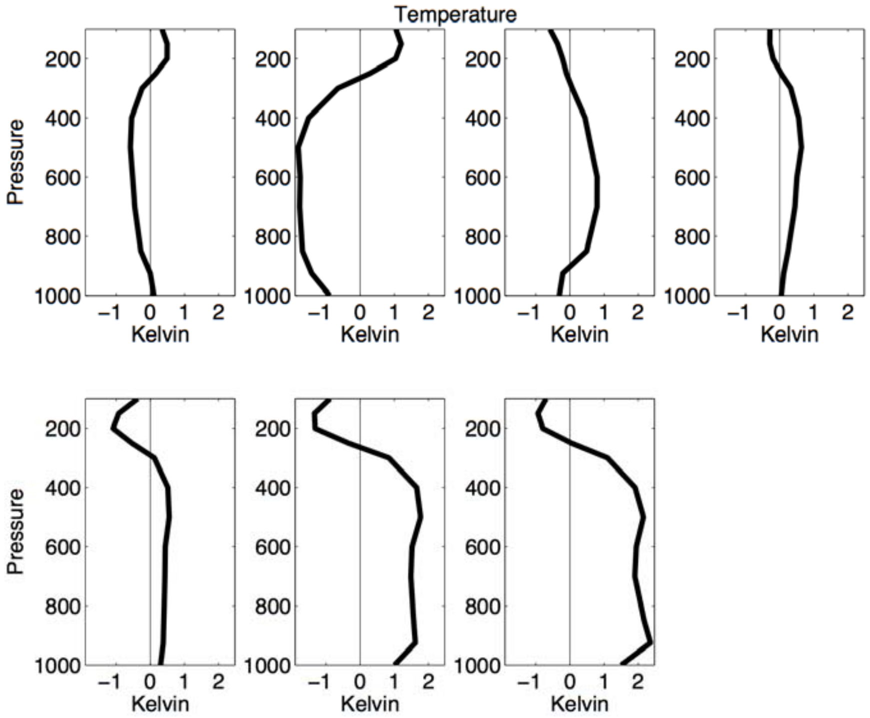


Figure 3.3 – Mean temperature profiles from NCEP Reanalysis.

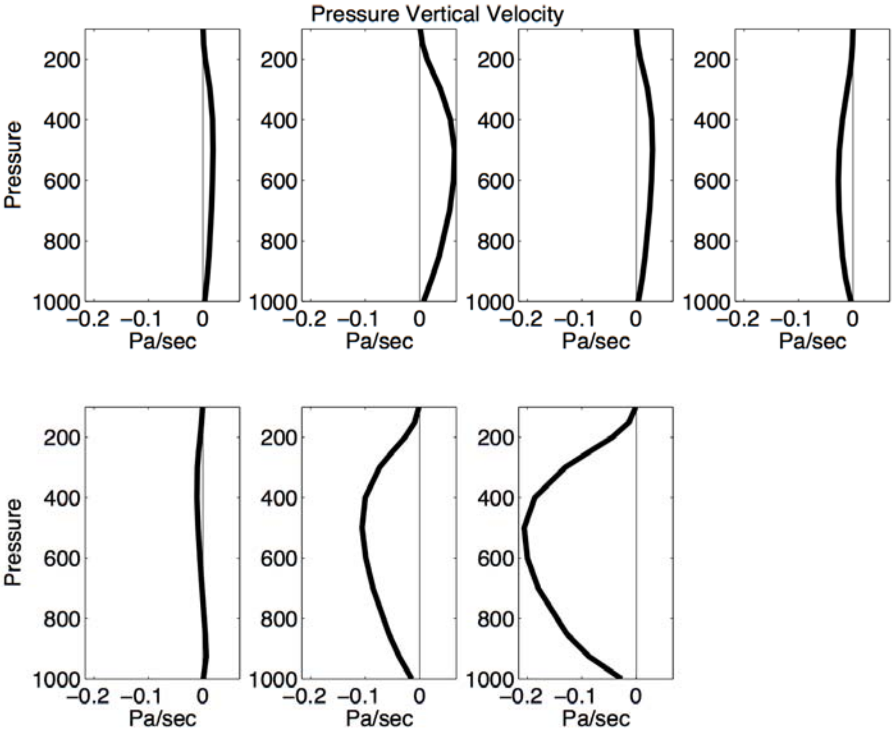


Figure 3.4 – Mean pressure vertical velocity.

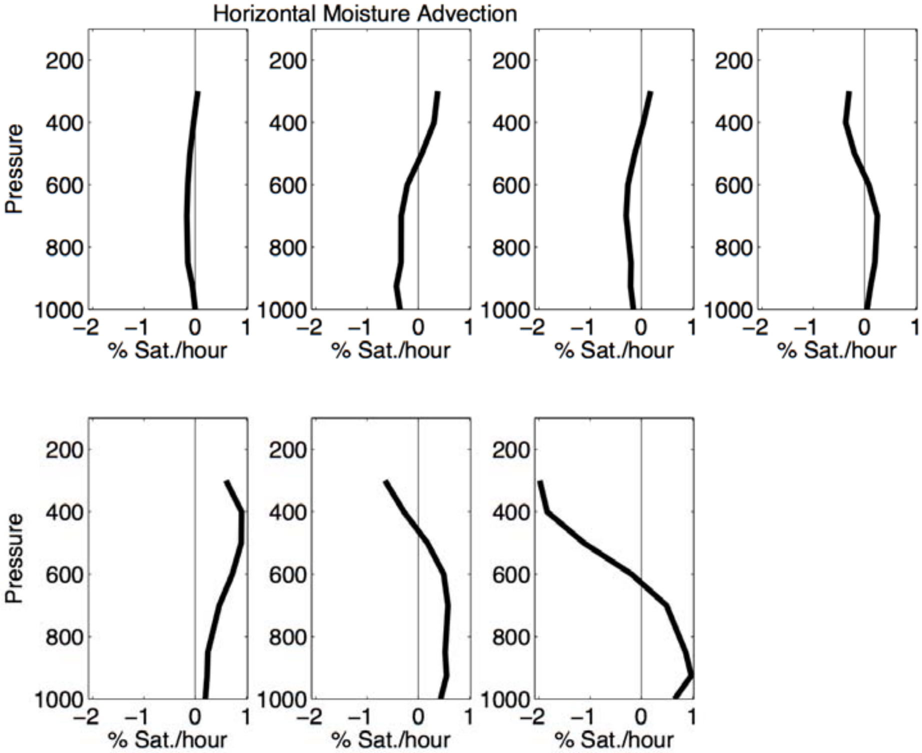


Figure 3.5 – Mean horizontal moisture advection.

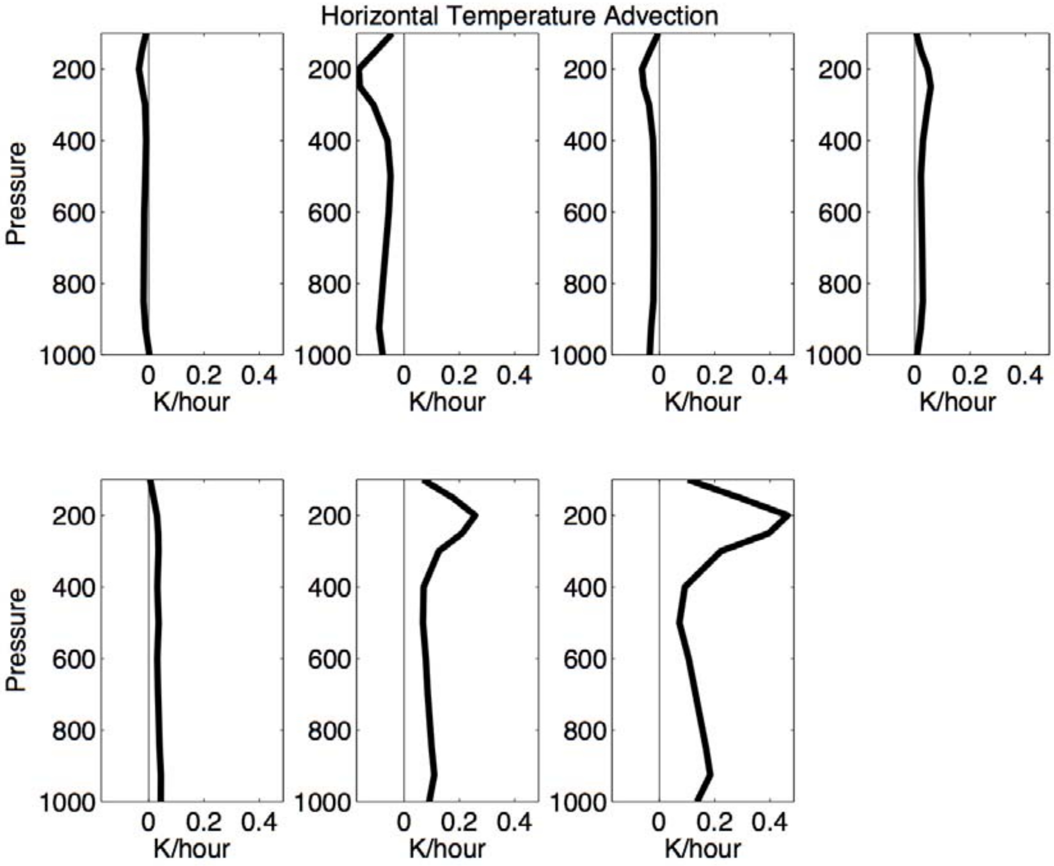
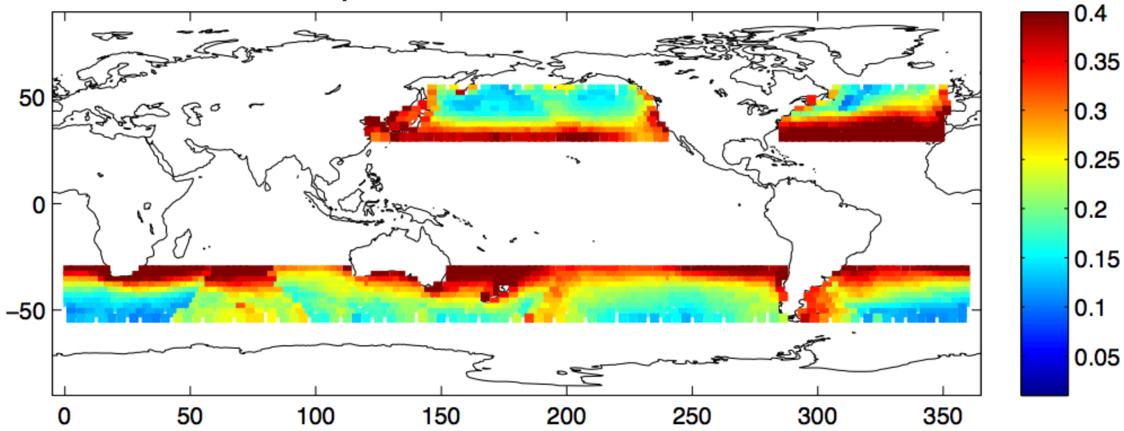
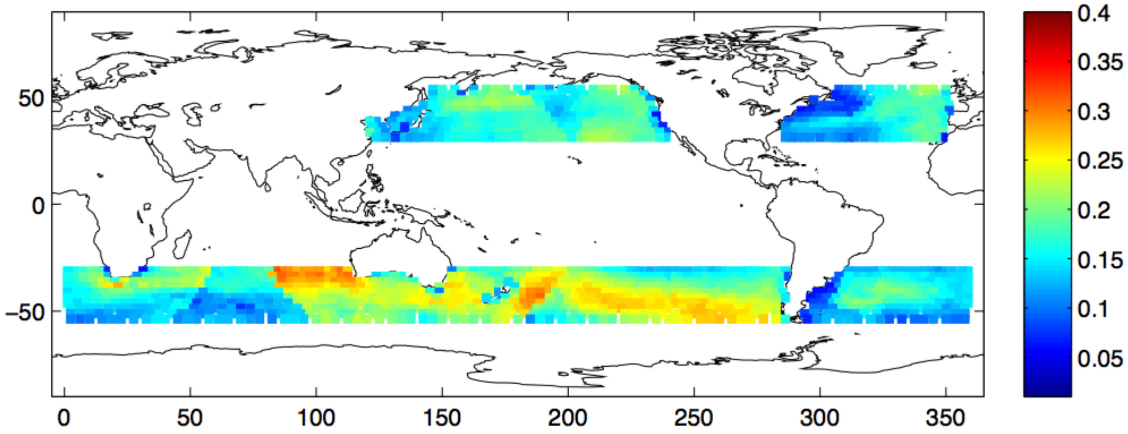


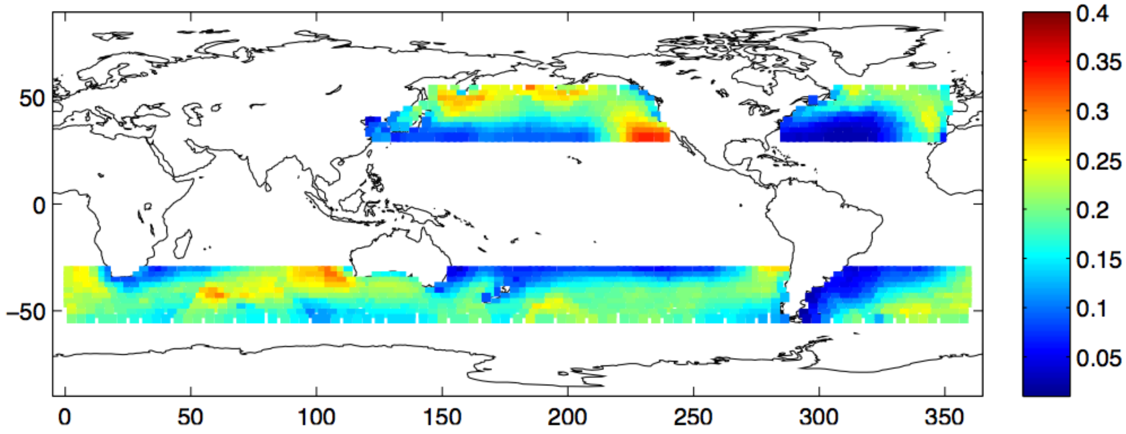
Figure 3.6 – Mean horizontal temperature advection.



(a) – Patchy Cumulus

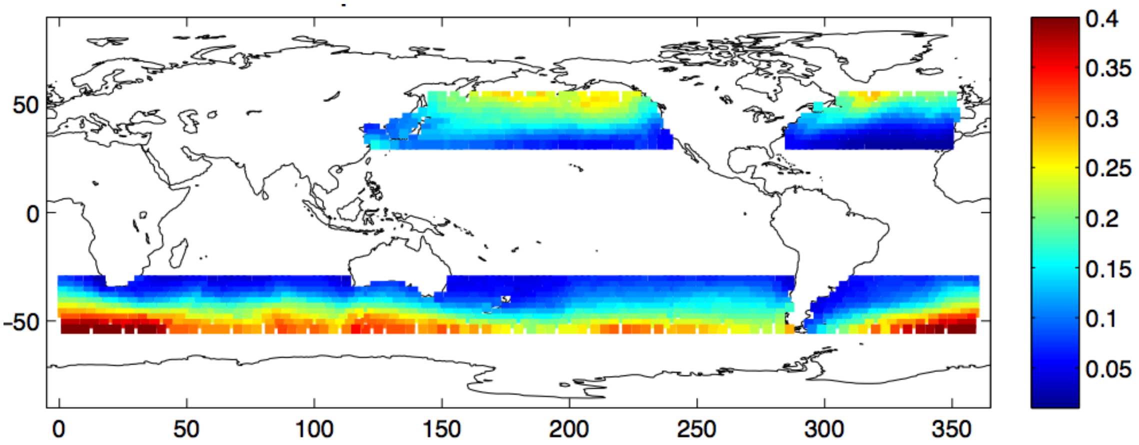


(b) – Large Cumulus

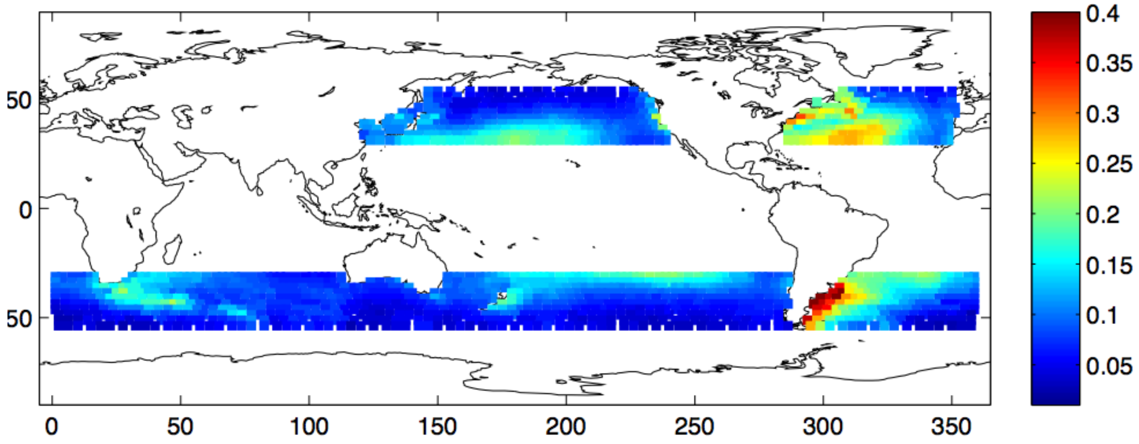


(c) – Stratocumulus/Stratus

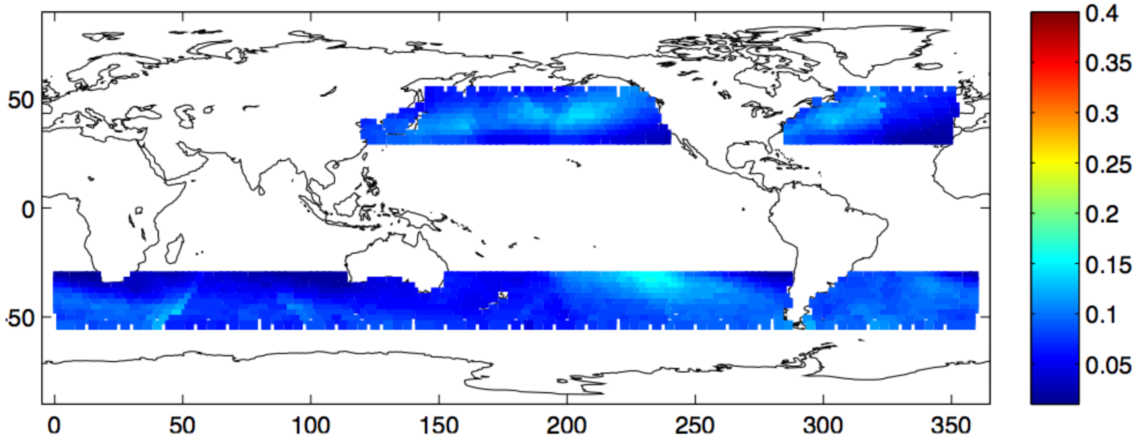
Figure 3.7 – Annual spatial distribution for each cluster.



(d) – Deep Altostratus

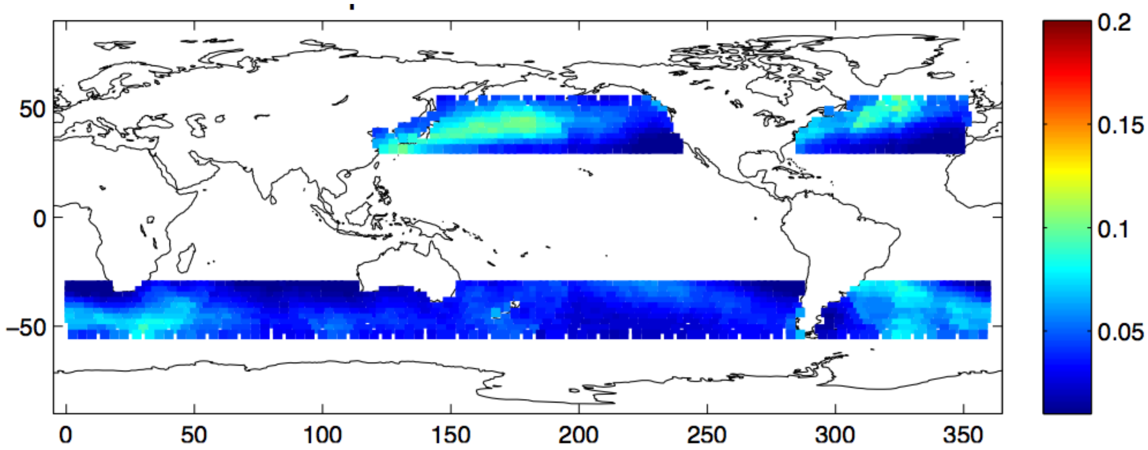


(e) – Cirrus



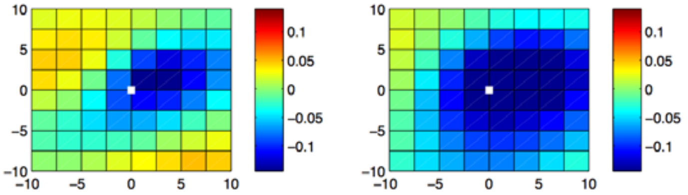
(f) – Weak Frontal

Fig 3.7 (cont)



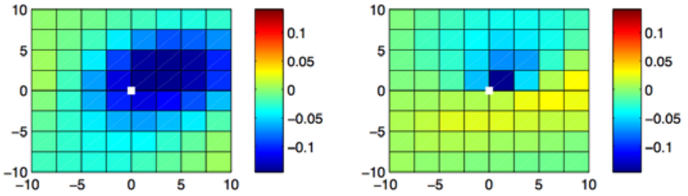
(g) – Strong Frontal

Fig 3.7 (cont)



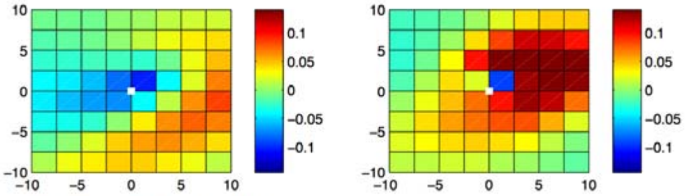
Distribution of Cluster #3 around Cluster #7

Distribution of Cluster #4 around Cluster #7



Distribution of Cluster #5 around Cluster #7

Distribution of Cluster #6 around Cluster #7



Distribution of Cluster #7 around Cluster #7

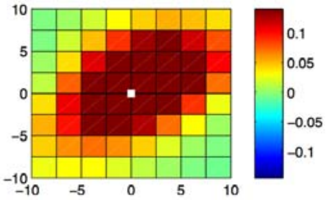
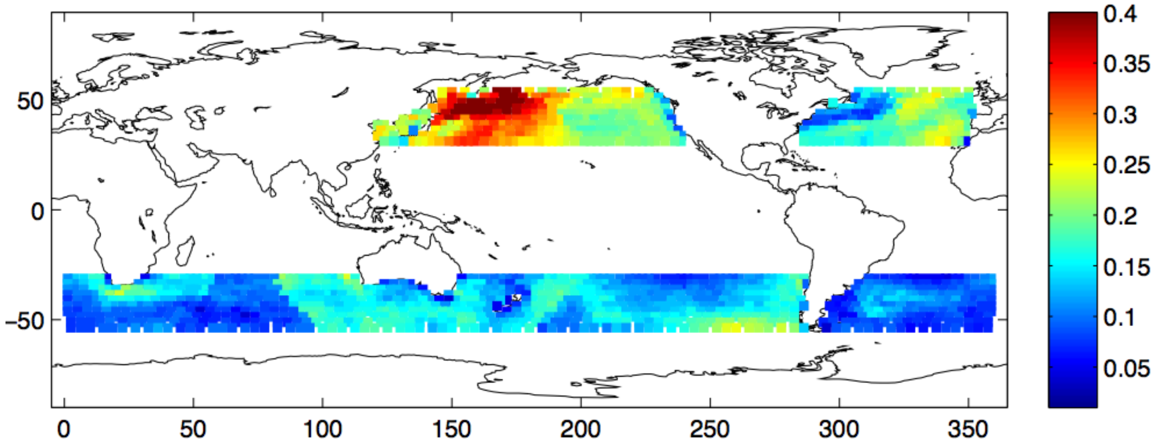
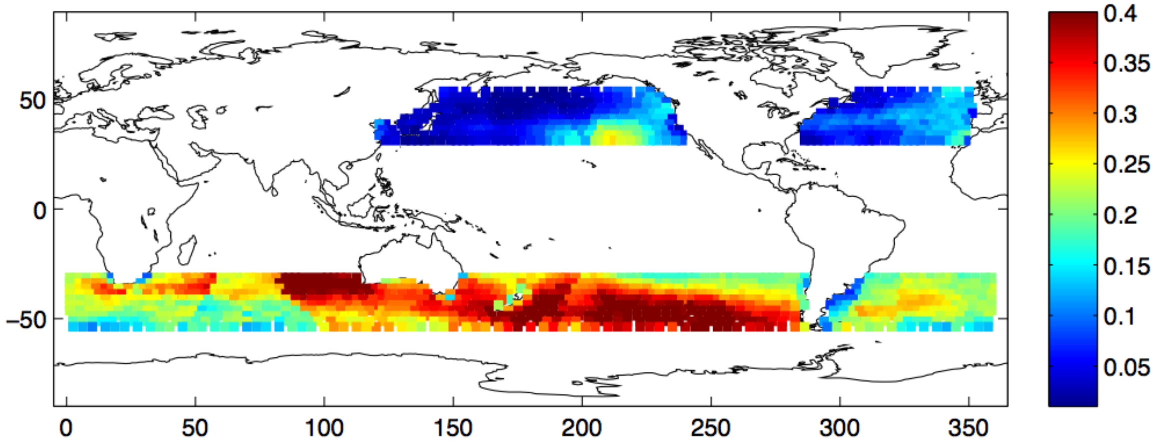


Figure 3.8 – Spatial distribution of the seven clusters around the occurrence of the Strong Frontal cluster.

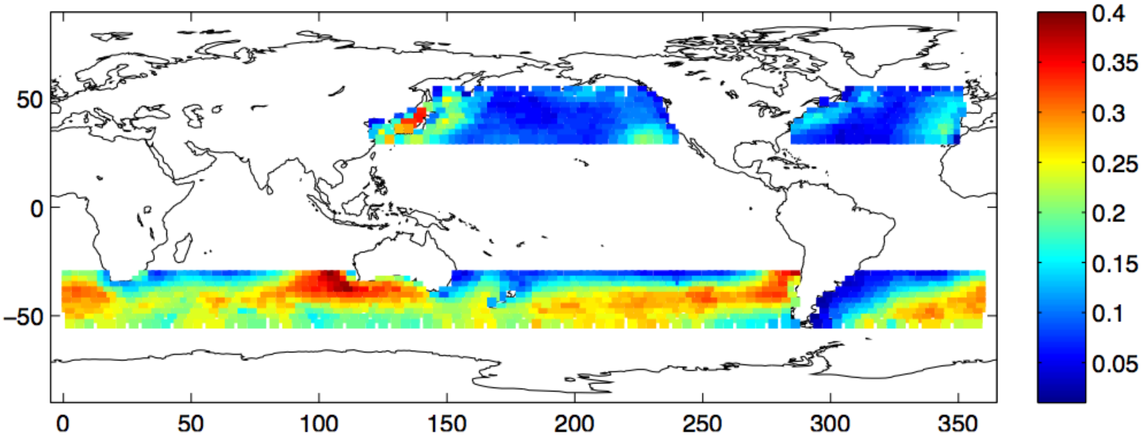


(a) – DJF

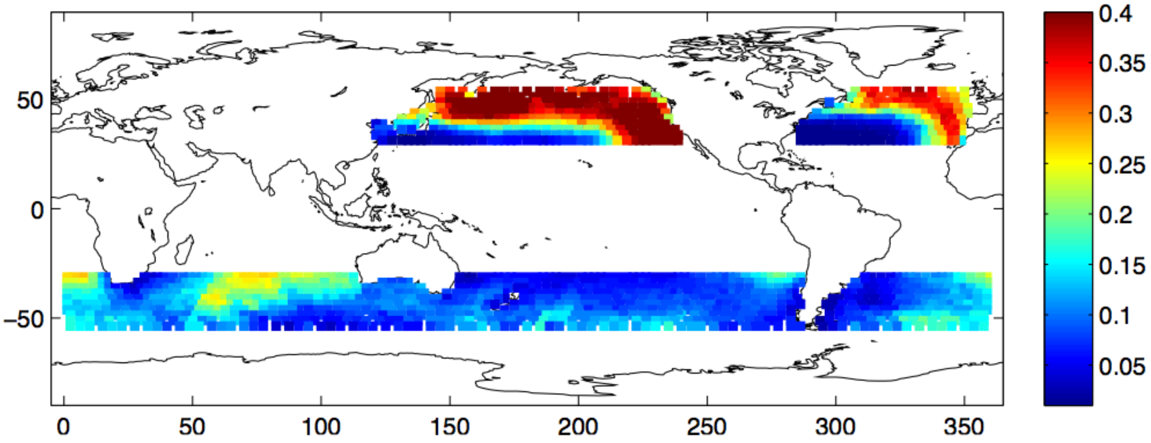


(b) – JJA

Figure 3.9 – Seasonal distribution of Large Cumulus cluster.



(a) – DJF



(b) – JJA

Figure 3.10 – Seasonal distribution of Stratocumulus/Stratus cluster.

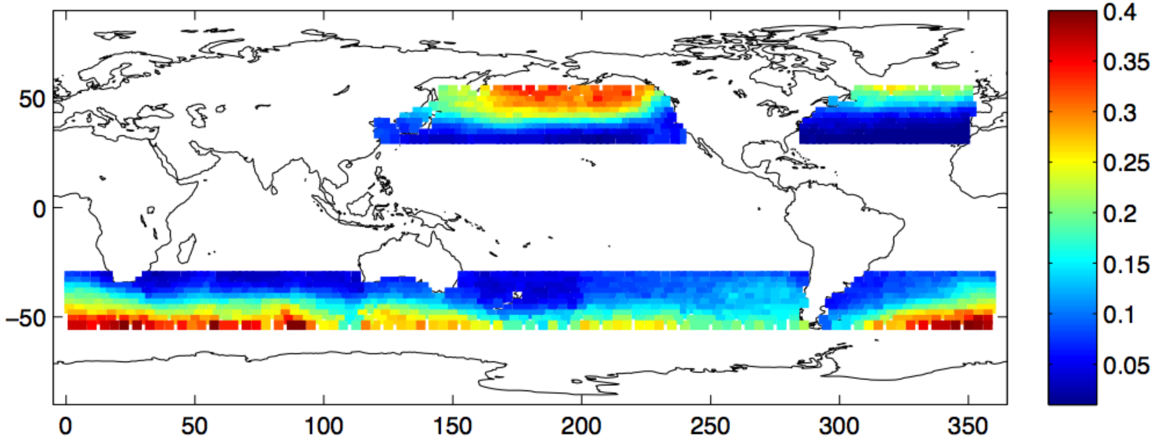


Figure 3.11 – Distribution of Deep Altostratus cluster for JJA.

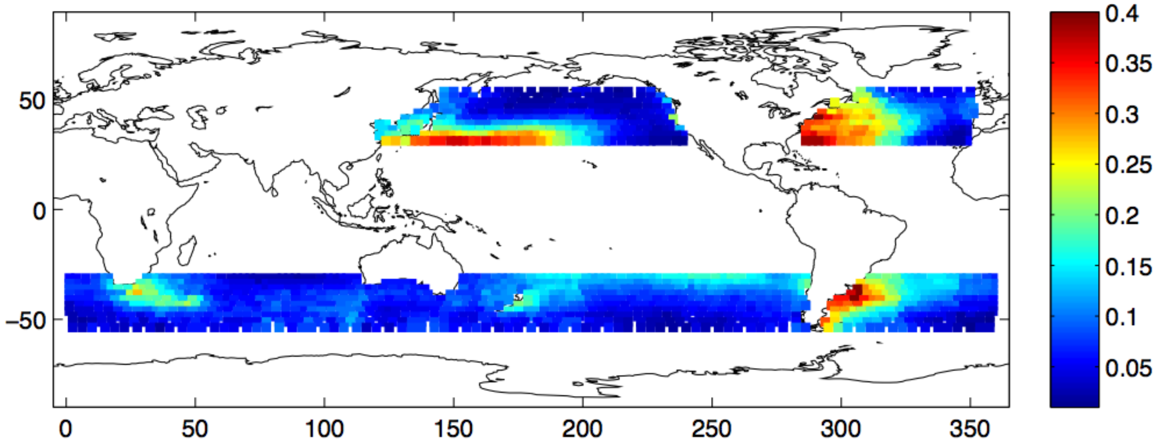


Figure 3.12 – Distribution of Cirrus cluster for JJA.

CHAPTER 4

In order to fully understand the implications for Earth's climate system of increased carbon dioxide, we must understand how all elements of climate might change. Clouds play an integral part in the climate system by reflecting solar radiation and absorbing terrestrial radiation. The balance between cooling and warming is an important control on the temperature of Earth's surface. Clouds reflect as much as 200 W/m^2 more shortwave radiation than clear sky [*Weaver and Ramanathan*, 1997], while the total direct radiative forcing of a doubling of atmospheric CO_2 weighs in at less than 2 W/m^2. Since clouds are such a important control on the balance of radiation in the atmosphere, it is important to understand how clouds might respond to an initial warming from increased CO_2; this is known as the cloud-climate feedback.

A first step in understanding cloud feedbacks is to determine the sensitivity of cloud properties to temperature. *Jakob et al.* [2005] suggest that linking distinct cloud regimes to specific characteristics of the atmosphere can be particularly useful in understanding cloud feedbacks. The present study builds on the classification system laid out in the previous chapter to diagnose the sensitivity of cloud properties—particularly their ability to affect the flux of radiation in the atmosphere—to changes in atmospheric temperature.

A k-means clustering algorithm is applied to International Satellite Cloud Climatology Project (ISCCP) data to group cloud regimes based on similar cloud properties. Many other studies have shown this to be an effective method for separating satellite data into cloud regimes accompanied by distinct dynamics and thermodynamics [*Jakob and Tselioudis*, 2003; *Gordon et al.*, 2005; *Jakob et al.*, 2005]. The use of the clustering algorithm allows us to group cloud scenes with similar atmospheric dynamics and to divide this relatively homogeneous group into relatively warm and cold groups, providing information about thermodynamic changes to clouds and the resulting effect on the balance of radiation in the climate system. Section 2 describes the data used in this study. The third section describes the clustering routine and the method for distinguishing warm and cold cases. Section 4 discusses the changes in the radiative properties of clouds as a result of atmospheric warming.

Data Sources

The source of cloud observations for this investigation was the three-hourly ISCCP D1 equal-area (280 km x 280 km) data set, originally processed from radiances primarily measured by geostationary weather satellites, with lesser contribution from polar-orbiting satellites [*Rossow et al.*, 1996; *Rossow and Schiffer*, 1999]. The ISCCP data consist of cloud fractions within a gridbox in three intervals of cloud-top pressure (CTP) and cloud optical thickness (τ), giving us nine ctp-τ categories. Each element of the data is the fraction of individual satellite pixels that exhibit those height and thickness properties. Since we analyzed the optical properties of clouds, valid data only exist for daytime hours. We restricted our analysis to one time point per day for each satellite gridbox, choosing the value with the smallest solar zenith angle (closest to local noon). This restriction avoided biases associated with more valid data points from regions near the equator and from points in the summer hemisphere, where there are a greater number of daylight hours. The satellite pixels used to generate the ctp-τ histograms are approximately 4-7 km in size and spaced approximately 30 km apart, with up to 80 pixels per gridbox.

Our analysis spans nearly the entire available record of ISCCP, 21 years (1984-2004), and incorporates all ocean points between 30° and 50° in both hemispheres, representing 1,444 gridboxes. The ISCCP data consisted of nearly 10 million observations, each consisting of a CTP-τ histogram, which is a nine-dimensional vector of cloud fraction. Clear-sky observations, which are infrequent (occur less than 1% of the time), and for which all elements of the vector are exactly 0, are excluded from this analysis.

In addition to mean cloud properties, we examined the radiative flux data derived from the ISCCP data [*Zhang et al.*, 2004]. The flux data consists of upwelling and downwelling, shortwave and longwave radiative flux for both clear and cloudy parts of the gridbox. This data is provided at the surface, the top of the atmosphere (TOA), and at three levels within the atmosphere (680 mbar, 440 mbar, and 100 mbar). In addition to the radiances from the ISCCP data, the profiles of temperature and humidity are derived from TIROS operation vertical sounder (TOVS). This with column ozone abundance, ISCCP derived profiles of cloudiness, and a climatological distribution of aerosol are input into the radiative transfer model of the Goddard Institute for Space Studies (GISS) GCM.

In order to obtain information about the dynamics and thermodynamic structure of the atmosphere, we utilized National Center for Environmental Prediction (NCEP) National Center for Atmospheric Research (NCAR) Reanalysis [*Kalnay et al.*, 1996]. This data set provided standard meteorological parameters as well as the advective tendencies of moisture and temperature derived from the large-scale gradients and atmospheric motions, which are important to the formation of clouds. We have restricted our analysis to midlatitudes because much of the dynamical forcing that leads to cloud formation in these regions is at or above the spatial scale of the satellite gridboxes, and the best quality reanalysis data from NCEP-NCAR is in the midlatitudes.

Clustering and Temperature Restriction

The clustering algorithm used in this study is discussed in detail in the companion study (Part I). The k-means procedure classifies all data elements into a specified number of clusters such that within-cluster variance is minimized [*Hartigan*, 1975; *Jakob and Tselioudis*, 2003]. The only arbitrary parameter needed is the number of clusters; the character of the individual cluster means is then objectively determined by the data. The clustering process began with random selection of k data elements as initial seeds, each element comprising a nine-element vector of cloud fraction in each height-thickness category. All other elements in the data set were then assigned to the initial seed to which they were closest in a Euclidean sense. The number of elements in a cluster divided by the total number of elements is the frequency of occurrence of the cluster, and the average of all elements in the cluster is the centroid. These cluster centroids became new seeds to reinitialize the clustering routine, which was repeated until the centroids converged.

Before the clustering algorithm was applied to the data, some quality control was used to remove unreliable data points. All points with any sea ice, as reported by the satellite, or any points with anomalously high clear-sky albedo were excluded. The normal range of clear-sky albedo was determined for bins of solar zenith angle (SZA) by calculating the difference between the 1st percentile and the median value. A data point has a valid value for clear-sky albedo if it is less than the sum of the median and the difference between the median and the 1st percentile value. This allows clear-sky albedo to vary uniformly above and below the median. This calculation is done independently for bins of SZA, as there is a strong relationship between α_{clear} and SZA (not shown). We also compared the skin temperature as reported by the satellite (T_{skin}) to the sea-surface temperature (SST) derived from NCEP reanalysis. We exclude points in which T_{skin} was less than 271 K and cases where the difference between SST and T_{skin} was greater than -4K and less than 8K. We noticed that T_{skin} tended to be warmer than SST by about 2K for most observations over the midlatitude oceans. All of these restrictions removed less than 1% of the initial data, but was necessary to ensure there were no spurious changes contaminating our analysis.

To best approximate temperature changes due to climate change, we used the mean temperature of the troposphere to divide cases into warm and cold, and did not restrict the temperature difference to the surface. In order for a case to be considered relatively warm (or cold), it had to be above (or below) the median of tropospheric mean temperature for each ISCCP gridbox, calendar month, and cluster number. This is done to ensure that no geographical or temporal biases are introduced. As observed by *Norris and Iacobellis* [2006], a large contributor to local temperature is temperature advection. In order to eliminate this dynamical influence, we require that all cases analyzed be in the middle portion (between 25th and 75th percentile) of horizontal and vertical temperature advection. This is conducted independently for three different layers of the atmosphere, which corresponded to the layers of the ISCCP cloud histograms (1000-680mbar, 680-440mbar, and 440-100mbar).

Two more dynamic differences to be controlled for are the lower-tropospheric static stability and the tropopause height. The former has particular influence on the dynamics of the stratocumulus cluster, while the latter influences the three high-cloud clusters. The change in temperature is greater in the mid-troposphere than at the surface, so cases that are relatively warm have a higher stability than cold cases. To correct for the dynamic influence of changes in lower-tropospheric stability, we require that the temperature difference between 1000mbar and 700mbar be in the middle portion for the data. For the high-cloud clusters, warm cases tend to have a higher tropopause, which allows for clouds to extend higher in the atmosphere. Since we are interested in non-dynamical effects of temperature on clouds, we wanted to minimize this effect. To do so, we require that the temperature difference between 200mbar and 400mbar be in the middle portion. The temperature difference between these two levels acts as an indication of tropopause height because of the strong temperature reversal that occurs between 400mbar and 200mbar. Again, all of this is conducted independently for each ISCCP gridbox, calendar month, and cluster number, and we ensure that each group has the same number of warm and cold points.

The initial data contained just under 10 million points, and after all these restrictions we are left with about 75,000 points designated as warm and the equal number as cold. In order to ensure that our study is relevant, we must strike a balance between being overly selective on dynamic constraints and keeping a sufficiently large number of points.

To understand the differences between cases that are warm or cold, we must analyze the variability of cloud and meteorological properties within each group. To determine this, we calculate the 95% confidence interval for the cloud and meteorological parameters in the warm and cold groups of each cluster. For each cluster, we calculate the effective number (N_{eff}) of points that are designated warm or cold. Points are considered independent observations if they are not adjacent gridboxes and separated by more than one day in time. From all points that are either warm or cold, we randomly select N_{eff} points, without replacement, and label them as warm, and the equal number as cold. We then have a selection of points that has the same effective number as our original warm and cold groups, but with a random mixture of warm and cold points. We repeat this 1,000 times. For mean cloud or meteorological properties to be different at the 95% confidence level, the difference must be greater than 975, or fewer than 25 of the random realizations described above.

Results

Figure 4.1 shows the mean ISCCP histograms for each of the seven clusters that were described in Part I of this study. These histograms represent the frequency that clouds in the domain exhibit the specified category of cloud-top pressure and cloud optical thickness. Since we are interested in changes to clouds as a result of surface warming, we will focus on differences in cloud properties between the warm and cold cases.

Figure 4.2 shows the difference between the ISCCP histograms for the warm cases and cold cases. Unshaded areas of the histograms represent regions where the difference was not significant at the 95% confidence level as described previously. Additionally, Table 4.1 shows the warm-cold differences in mean cloud properties for each cluster per degree change in temperature. Again, numbers are shown only for those values that are different than zero at the 95% confidence level.

In Table 4.1 we see that there is a fairly consistent change between the different clusters. Reduction in cloud fraction, increase in cloud-top pressure (lowering of cloud-top), and increase in optical thickness are the most common pattern going from cold to warm. The Patchy Cu cluster has the largest decrease in cloud fraction at 2.3%/K, accompanied by increases in cloud-top pressure and optical thickness. The other low-cloud clusters (Large Cu and Sc/St) have smaller reductions in cloud fraction, but have large increases in cloud-top pressure at 6.9 and 9.1 mb/K, respectively. The Large Cu cluster's increase in optical thickness (0.09/K) is similar to that of the Patchy Cu (0.13/K), while the Sc/St cluster shows a reduction in optical thickness of 0.05/K. The Deep As cluster shows little change in cloud fraction or cloud-top pressure, but exhibits the largest change in optical thickness of 0.33/K. A reduction in cloud fraction of 0.9 %/K is accompanied by a reduction in cloud-top (-3.8 mb/K), with a small change in optical thickness (0.03/K). For the weak frontal cluster, the only significant change is an increase in optical thickness of 0.25/K, while the strong frontal cluster has no significant changes (Table 4.1).

Figure 4.2 shows the difference between the mean ISCCP histograms between the warm and cold cases for each cluster. The first two clusters exhibit a general shift to optically thicker clouds, with reductions in the occurrence of clouds in the thinnest category and increases in the middle-thickness category. This is mirrored in the increase in mean optical thickness for these two clusters. Similar shifts toward the higher optical thickness can be seen for the Deep As cluster as well as the Weak Frontal cluster. A decrease in mean cloud-top pressure for the Cirrus cluster can be seen from the changes in the histogram. A general decrease in clouds is exhibited in the lower two levels of the atmosphere with an attendant increase in cloud fraction in the upper level.

To aid in the understanding in the cloud changes, we can look at the differences in the mean profiles of meteorology derived from NCEP Reanalysis. Figures 4.3-4.7 show the average anomaly profile of temperature, relative humidity, vertical velocity, and specific humidity (respectively). The similarity in the profiles of vertical velocity give confidence that the dynamics in the warm and cold subgroups in the cluster are similar.

In order to better understand the effect of changes in cloud properties on the climate system, it is important to look at changes in the fluxes of radiation in the atmosphere. Upwelling and downwelling long- and shortwave radiation data are provided by the ISCCP flux data at the surface, top-of-atmosphere, and three levels in the atmosphere (100mb, 440mb, and 680mb). These levels correspond to the boundaries between low, middle, and high clouds in the ISCCP histograms. In order for our study to be applicable to the climate change scenario of warming from increased CO_2, we must be careful in our analysis of radiative fluxes. Since we are only selecting one satellite observation per day for each ISCCP gridbox, and it is the observation closest to local noon, we must normalize the shortwave fluxes by the top-of-atmosphere insolation. For each observation, the shortwave flux at any point in the atmosphere is divided by the insolation for that point. The shortwave fluxes are then multiplied by the diurnally averaged insolation for that cluster. This allows us to calculate the mean radiative effect of the changes in cloud properties. Table 4.2 shows the mean shortwave cloud radiative forcing (SWCRF) for warm and cold cases defined as:

$$SWCRF = (SW \uparrow^{TOA}_{clear,normalized} - SW \uparrow^{TOA}_{all-sky,normalized}) * SW \downarrow^{TOA}_{diurnal}$$

The shortwave fluxes must be normalized by the instantaneous insolation, so that more shortwave forcing is not primarily a result of more incoming solar radiation. To convert SWCRF back to units of W/m^2, we multiply the average difference between the clear and all-sky SW fluxes by the diurnally averaged insolation for each cluster.

This allows us to calculate average changes in cloud properties that are not biased by season or latitude.

In order to understand the changes in SWCRF, we can break the changes into components. Another definition of SWCRF:

$$SWCRF = -f * \alpha * SW \downarrow^{TOA}$$
$$\alpha = \alpha_{overcast} - \alpha_{clear} = \frac{SW \uparrow^{TOA}_{overcast} - SW \uparrow^{TOA}_{clear}}{SW \downarrow^{TOA}}$$

This allows us to attribute changes in SWCRF to observed changes in cloud properties, either cloud fraction (f) or α, which is the difference between the albedo of an overcast scene and the clear-sky albedo. For most clusters, the changes to clear-sky albedo are smaller than the changes to the albedo of the overcast scene (Table 4.2). To break up the SWCRF changes, we define:

$$\Delta SWCRF_{CF} = -\Delta f * \overline{\alpha} * SW \downarrow^{TOA}$$
$$\Delta SWCRF_{\alpha} = -\overline{f} * \Delta\alpha * SW \downarrow^{TOA}$$

Therefore, the term $\Delta SWCRF_{CF}$ represents changes in the SWCRF resulting only from changes in cloud fraction while $\Delta SWCRF_{\alpha}$ represents changes only from changes in α. These changes are primarily a result of changes to the cloud, as opposed to clear-sky changes (Table 4.2). All calculations are conducted such that a positive number represents a net warming for the warm cases relative to the cold cases.

For the Patchy Cu cluster, there is a large warming of 1.3 W/m^2*K associated with a reduction in cloud fraction, letting more solar radiation in, while only partially balanced by a cooling associated with a small increase in reflected shortwave radiation

(-0.3 W/m^2*K) as a result of an increase in optical thickness (Table 4.1). There is a near cancellation in the SW effects for the Large Cu cluster, with a small warming (0.3 W/m^2*K) due to a decrease in cloud fraction nearly balanced by a change in SWCRF of -0.4 W/m^2*K from the increased opacity of these clouds. The changes to the Sc/St cluster warm the climate system, with warming resulting from less cloud fraction (0.5 W/m^2*K) and a 0.7 W/m^2*K warming from a modest increase in optical thickness of 0.05 (Table 4.1).

The largest SW change among any of the clusters is due to a cooling of the Deep As cluster. All of the -1.5 W/m^2*K change in SWCRF is a result of the increase in optical thickness of 0.33/K. The warming of 0.3 W/m^2*K for the Ci cluster is a result of the reduction of cloud fraction, while changes in optical thickness have negligible impact. Conversely, the change in optical depth produces the large cooling of -1.0 W/m^2*K for the Weak Frontal cluster. There is little change in the SW properties of the Large Frontal cluster.

For the changes in longwave flux, calculating the difference between the average LWCRF for the warm and cold subgroups is not directly applicable to the climate change scenario, since the temperature difference that we observe is not identical to what we would expect from a doubling of CO_2. The temperature change is largest in the mid-troposphere, with a much smaller difference near the surface. This is in contrast to a much more uniform warming with height projected in $2xCO_2$ model simulations (*IPCC*, 2007). Our observations will underestimate the upwelling longwave flux from the surface, relative to the more uniform warming associated with climate change. In order to understand how the longwave fluxes might change, we can examine the components of the longwave flux:

$$LW_{all-sky}^{TOA} = f * LW_{overcast}^{TOA} - (1-f) * LW_{clear}^{TOA}$$
$$LWCRF = LW_{clear}^{TOA} - LW_{all-sky}^{TOA} = f * (LW_{clear}^{TOA} - LW_{overcast}^{TOA})$$

We also can examine the terms that go into the top-of-atmosphere upwelling LW flux for overcast sky and the LWCRF, where g_{ac} is the above-cloud greenhouse parameter, ε_{cld} is the cloud emissivity, and LW_{bc} is the LW flux coming from beneath the cloud:

$$LW_{overcast}^{TOA} = (1-g_{ac}) * [\varepsilon * \sigma * T_{CTT}^{4} + (1-\varepsilon) * LW_{bc}]$$
$$LWCRF = f * LW_{clear}^{TOA} - f * (1-g_{ac}) * [\varepsilon * \sigma * T_{CT}^{4} + (1-\varepsilon) * LW_{bc}]$$

The cloud greenhouse parameter accounts for the reduction in the emission from cloud-top or from the atmosphere and surface below the cloud. For clouds that have near-unit emissivity, namely the Sc/St, Deep As, and the two Frontal clusters, the below-cloud longwave flux will be negligible. In these cases, nearly all radiation emitted from below the cloud-top is absorbed by the cloud. The cloud-top temperature (T_{CT}) is calculated from the cloud-top pressure and the mean temperature profile is derived from the NCEP reanalysis, so as to focus on the effect that changes in cloud properties, like cloud-top pressure, had on the TOA radiation balance.

Our analysis leaves us with one equation and two unknowns, namely g_{ac} and LW_{bc}. Yet, since the below-cloud longwave flux goes away for clusters with $\varepsilon \approx 1$, we can calculate a value of g_{ac} for those four clusters. Fig. 4.8 shows that the value of g_{ac}

is highly dependent on cloud-top pressure. We assume that this is the only factor that controls the value of this parameter. We use the values obtained for the two Frontal clusters and the mid-level cluster to deduce the value of g_{ac} for the cirrus cluster. Additionally, we use the value of g_{ac} for the St/Sc cluster and the Deep As, along with their mean cloud-top pressure, to determine the value of above-cloud greenhouse parameters for the other two low-cloud clusters.

Now that we have an estimate for the fraction of radiation absorbed by the atmosphere above the cloud for each cluster, we can use the above formula to calculate the below-cloud longwave flux for the three clusters with $\varepsilon < 1$ (Table 4.3). We have introduced these formulae to attribute changes in LWCRF to specific changes in the mean cloud properties. By utilizing the above equations, we can observe the change in LWCRF as a result of changes in cloud fraction, emissivity, and cloud-top temperature:

$$\begin{aligned}
LWCRF_f &= \Delta f * \overline{(LW^{TOA}_{clear} - LW^{TOA}_{overcast})} \\
LWCRF_\varepsilon &= -\overline{f} * (1 - g_{ac}) * \Delta\varepsilon[\sigma * \overline{T}^4_{CT} - LW_{bc}] \\
LWCRF_{CTT} &= -\overline{f} * (1 - g_{ac}) * [4 * \overline{\varepsilon} * \sigma * \overline{T}^3_{CT} * \Delta T_{CT}]
\end{aligned}$$

Values for each of these is displayed in Table 4.4. The changes in LWCRF as a result of the consistent reduction in cloud fraction for all clusters are all negative, which is a cooling on the climate. The changes resulting from increases in emissivity are relatively small, with the largest change for the cirrus clouds, which is also the cluster with the smallest mean emissivity. The largest changes in LWCRF come from changes to cloud-top temperature, with changes of both signs. For the Cirrus cluster, there is a strong warming associated with the changes in cloud-top (1.1 W/m^2 per K), while changes from emissivity (0.4 1 W/m^2 per K) and cloud fraction (-0.5 1 W/m^2 per K) nearly balance out. This is the only cluster in which changes in cloud properties act as a positive feedback for both the longwave and shortwave flux.

Conclusion

This study has presented a method by which satellite-derived data on cloud properties can be used to determine the character and magnitude of cloud feedbacks in our climate system. A clustering routine is applied to 21 years of ISCCP satellite data in order to partition the data into seven groups, each representing a different cloud regime with a unique suite of dynamics. These clusters are then used as the initial kernel for producing a set of observations whose dynamics are relatively homogeneous. In addition to cluster membership, other dynamics were factored in so as to minimize variability in dynamics that are important for cloud formation. We can then isolate changes in cloud properties resulting from thermodynamic changes of the atmosphere.

Across clusters, there is a consistent reduction in cloud fraction and an increase in optical depth for the warm clouds. The changes in cloud-top pressure are less consistent. The inclusion of the ISCCP flux data allows us to examine how these changes in cloud properties affect the radiative flux in the atmosphere. If our study is to be applicable to the climate change scenario, where an external forcing results in a near-uniform warming of the troposphere, we must take caution in the analysis of our results. Instead of taking the difference in LWCRF between the warm and cold subgroups, we instead examine how the LWCRF is sensitive to individual cloud properties. The sum of these changes can then provide information about the change in LWCRF in the climate change scenario.

For the changes in the shortwave flux, it was important to carefully normalize the TOA fluxes by the insolation, so as to consider mean radiative effects and so as to not introduce excess influence from summertime and equatorward points. The changes in the SWCRF could then be divided into those resulting from changes in cloud fraction or cloud albedo. For all clusters except the Sc/St, the two effects were competing, with a reduction in cloud fraction leading to a warming while an increase in cloud albedo leading to a cooling.

Table 4.5 shows the sum of the changes to the shortwave and longwave radiation. This partial cancellation of shortwave effects along with strong changes in cloud-top temperature resulting from a lowering of cloud-top, leads to an average cloud forcing of -0.47 W/m^2 per degree K, which suggests that clouds act as a negative feedback on the climate system. *Williams and Tselioudis* [2007] use a clustering algorithm on output from a GCM to suggest that changes in cloud properties will act to enhance the warming brought about by the doubling of CO_2.

Table 4.1 – Mean cloud properties for warm and cold subgroups, and the difference between the two divided by the temperature change for each cluster.

	1 - Small Cu	2 - Large Cu	3 - Sc/St	4 - Deep As	5 - Ci	6 - Small Frontal	7 - Large Frontal
Frequency	0.275	0.184	0.165	0.14	0.113	0.077	0.043
Mean CF	54.1	77.8	92.9	97.5	87.4	99.0	99.4
Warm	45.5	75.9	91.9	97.7	84.8	99.3	99.5
Cold	50.7	77.1	92.9	97.9	87.0	99.4	99.5
Diff (per K)	-2.28	-0.55	-0.44	-0.07	-0.92	-0.04	0.00
Mean CTP	658.2	781.0	776.4	584.3	431.8	382.6	347.6
Warm	679.2	799.2	795.0	588.1	425.2	389.1	350.9
Cold	668.7	783.8	774.5	589.5	434.1	388.5	349.0
Diff (per K)	4.6	6.9	9.1	-0.5	-3.8	0.3	0.8
Mean Tau	3.63	2.89	7.19	8.30	2.30	8.90	23.08
Warm	3.32	2.89	7.07	8.30	2.12	9.06	22.63
Cold	3.03	2.69	7.19	7.46	2.06	8.44	22.63
Diff (per K)	0.13	0.09	-0.05	0.33	0.03	0.25	0.00
emiss_cld	0.905	0.847	0.991	0.997	0.834	0.999	1.000
Warm	0.884	0.847	0.990	0.997	0.808	0.999	1.000
Cold	0.860	0.825	0.991	0.995	0.799	0.999	1.000
Diff (per K)	0.0106	0.0096	-0.0004	0.0009	0.0039	0.0002	0.0000
CTT	267.9	272.4	274.4	259.7	246.8	240.8	236.7
Warm	270.7	274.6	276.4	260.6	246.5	242.1	237.8
Cold	268.2	272.0	273.5	258.7	246.0	240.2	235.8
Diff (per K)	1.11	1.17	1.27	0.72	0.21	0.80	0.85

Table 4.2 – SWCRF for warm and cold subgroups, the difference between the two (per degree temperature change) and the changes in SWCRF resulting from changes in cloud fraction and emissivity

t_diff	2.25	2.23	2.25	2.58	2.37	2.45	2.37
	1 - Small Cu	2 - Large Cu	3 - Sc/St	4 - Deep As	5 - Ci	6 - Small Frontal	7 - Large Frontal
SWCRF	-39.01	-40.46	-96.96	-112.89	-55.04	-123.05	-168.38
Warm	-30.69	-37.85	-95.33	-113.48	-50.16	-118.72	-162.83
Cold	-32.16	-37.51	-97.99	-109.70	-50.87	-116.23	-162.94
Diff (per K)	0.66	-0.15	1.18	-1.46	0.30	-1.02	0.05
SWCRF_cf							
Warm	-25.68	-36.67	-95.29	-111.13	-47.31	-117.29	-162.77
Cold	-28.57	-37.26	-96.32	-111.35	-48.54	-117.42	-162.76
Diff	1.285	0.264	0.457	0.086	0.516	0.055	-0.006
SWCRF_alpha							
Warm	-27.46	-37.40	-94.97	-113.21	-47.97	-118.66	-162.69
Cold	-26.80	-36.53	-96.64	-109.26	-47.88	-116.06	-162.84
Diff	-0.293	-0.390	0.742	-1.534	-0.042	-1.060	0.064

Table 4.3 – Below-cloud longwave flux.

	1 - Small Cu	2 - Large Cu	5 - Ci
LW_bc	305.01	357.71	337.99

Table 4.4 – LWCRF changes as a result of changes in cloud fraction, cloud-top pressure, and emissivity.

	1 - Small Cu	2 - Large Cu	3 - Sc/St	4 - Deep As	5 - Ci	6 - Small Frontal	7 - Large Frontal
del_LWCRF(f) (per K)	-0.61	-0.07	-0.06	-0.03	-0.48	-0.04	0.00
del_LWCRF (T) (per K)	-0.53	-0.92	-1.38	0.14	1.09	-0.10	-0.32
del_LWCRF (e) (per K)	0.05	0.25	0.00	0.00	0.38	0.00	0.00

Table 4.5 – Total SWCRF and LWCRF changes for each cluster and the average from the midlatitude based on relative frequency of occurrence.

	1 - Small Cu	2 - Large Cu	3 - Sc/St	4 - Deep As	5 - Ci	6 - Small Frontal	7 - Large Frontal	Global
Frequency	0.275	0.184	0.165	0.14	0.113	0.077	0.043	
sum del_LWCRF	-1.09	-0.74	-1.44	0.11	0.99	-0.14	-0.32	
del_SWCRF (total)	0.65	-0.15	1.18	-1.47	0.30	-1.01	0.05	
SUM	-0.43	-0.89	-0.25	-1.36	1.29	-1.15	-0.27	-0.47

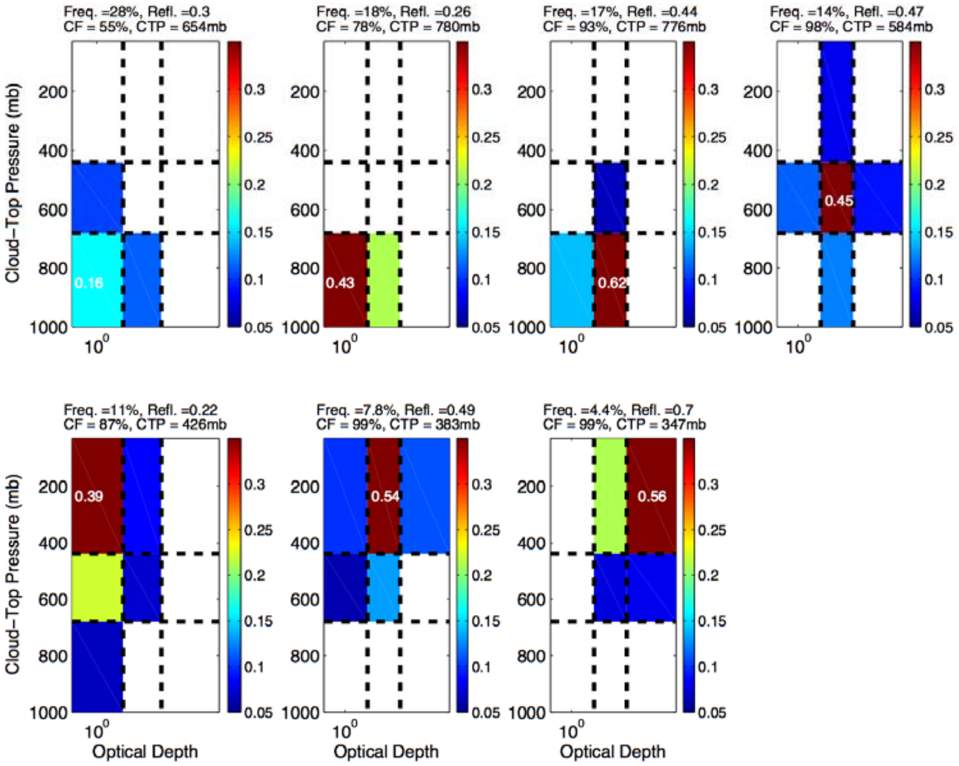


Figure 4.1 – Mean ISCCP histograms of cloud-top pressure and optical thickness.

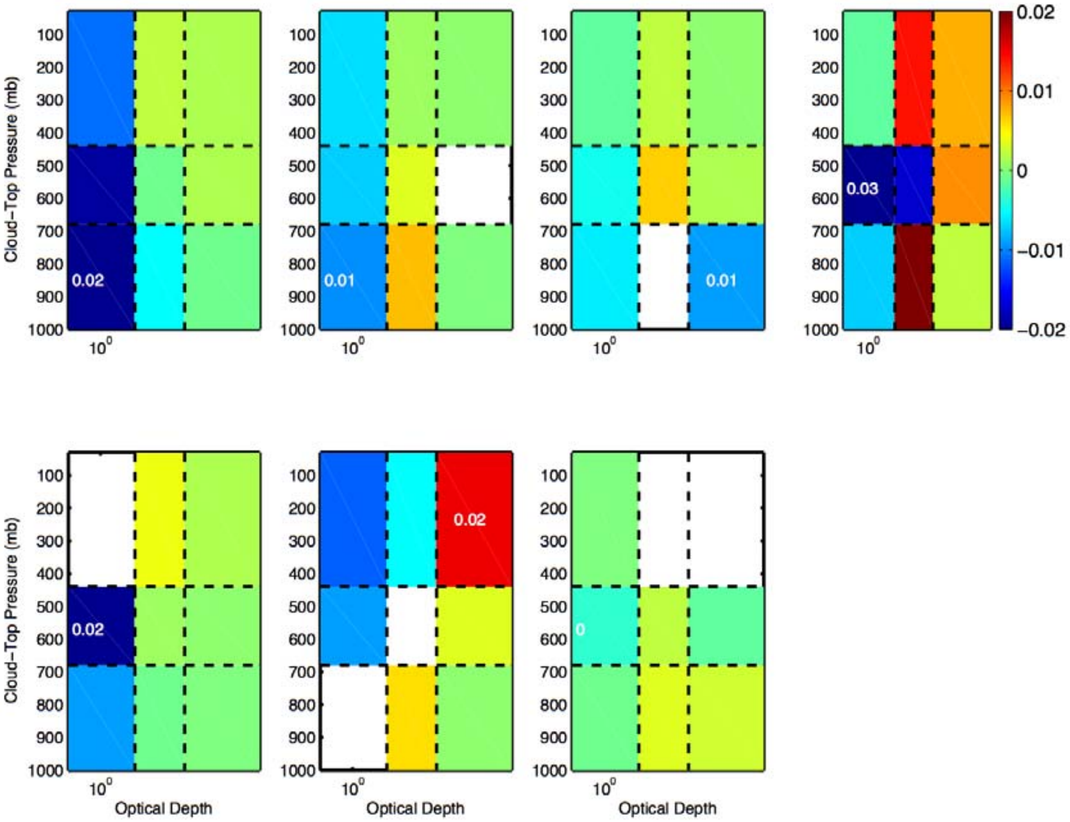


Figure 4.2 – Differences in ISCCP histograms between warm and cold subgroups for each cluster.

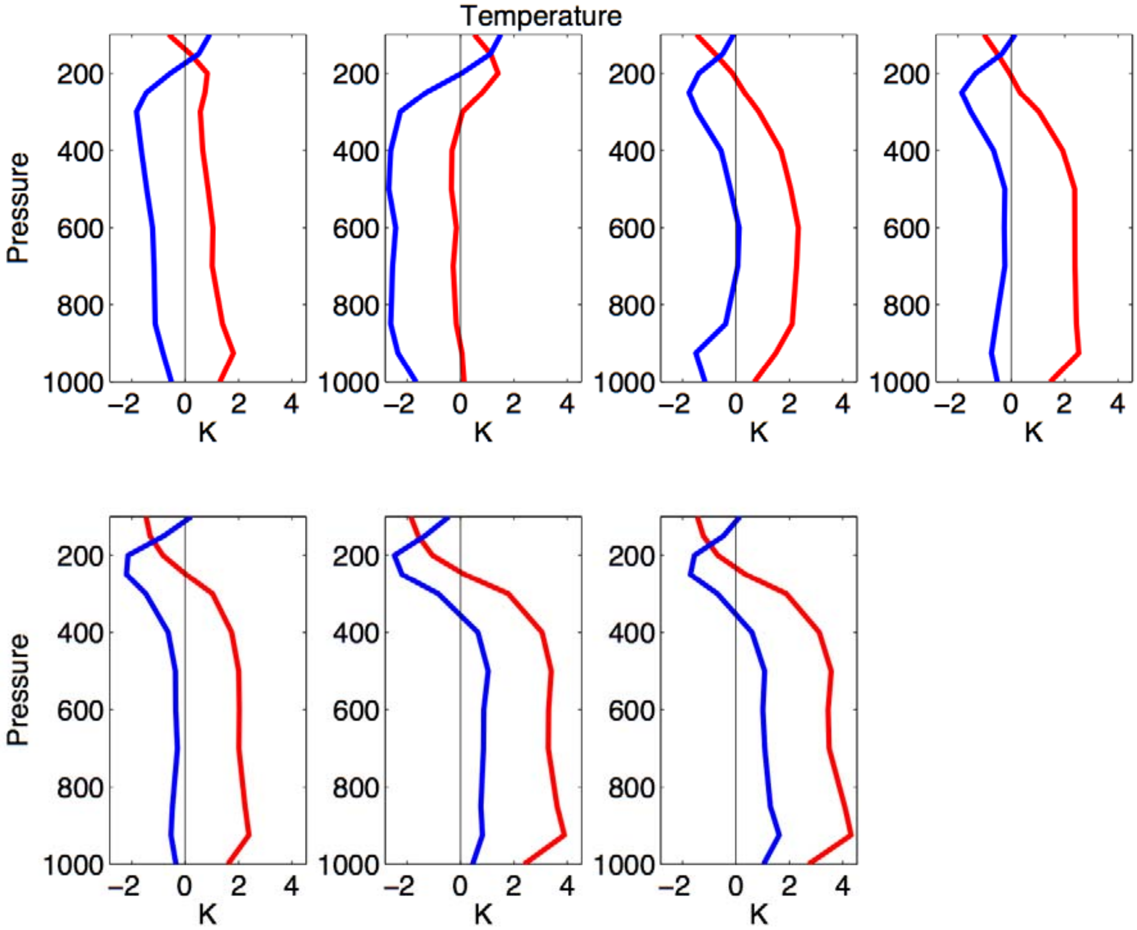


Figure 4.3 – Mean profile of anomaly temperature for warm (red) and cold (blue) subgroups of each cluster.

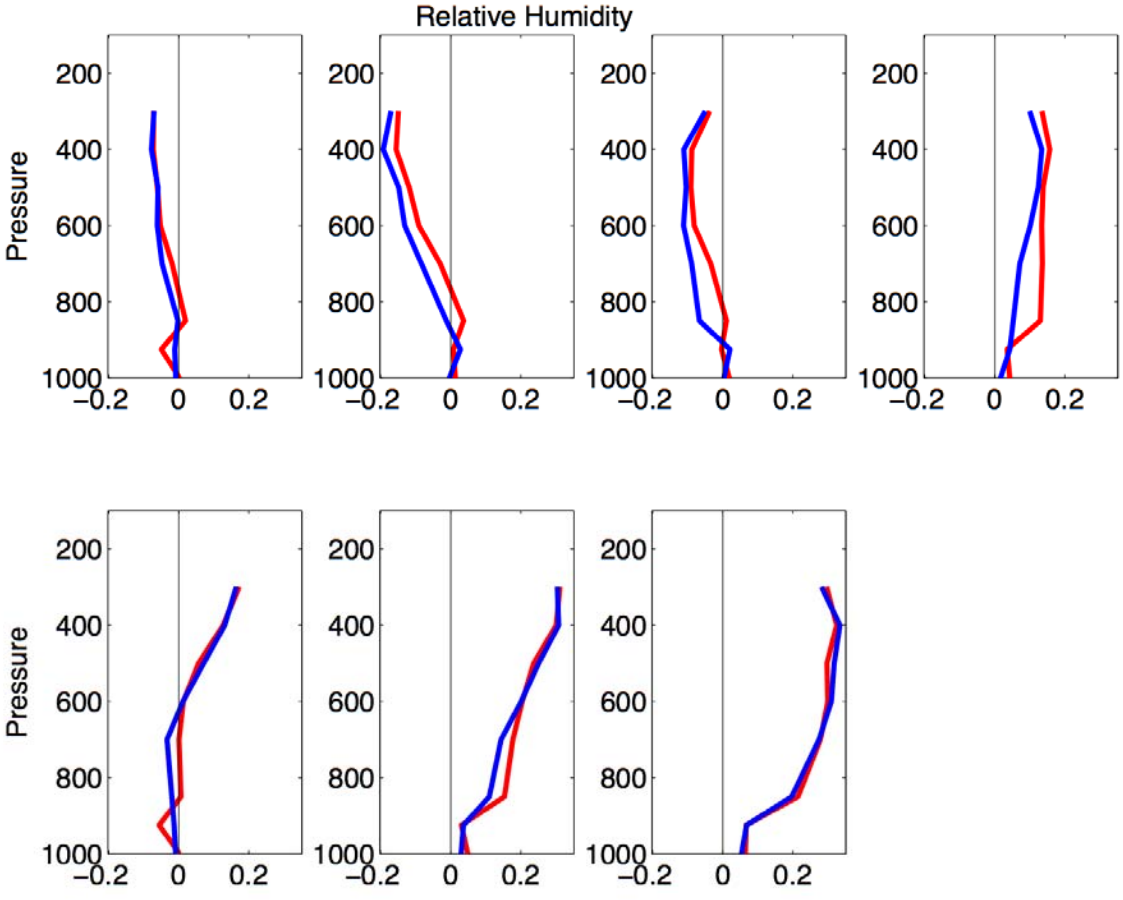


Figure 4.4 - Mean profile of anomaly relative humidity for warm (red) and cold (blue) subgroups of each cluster.

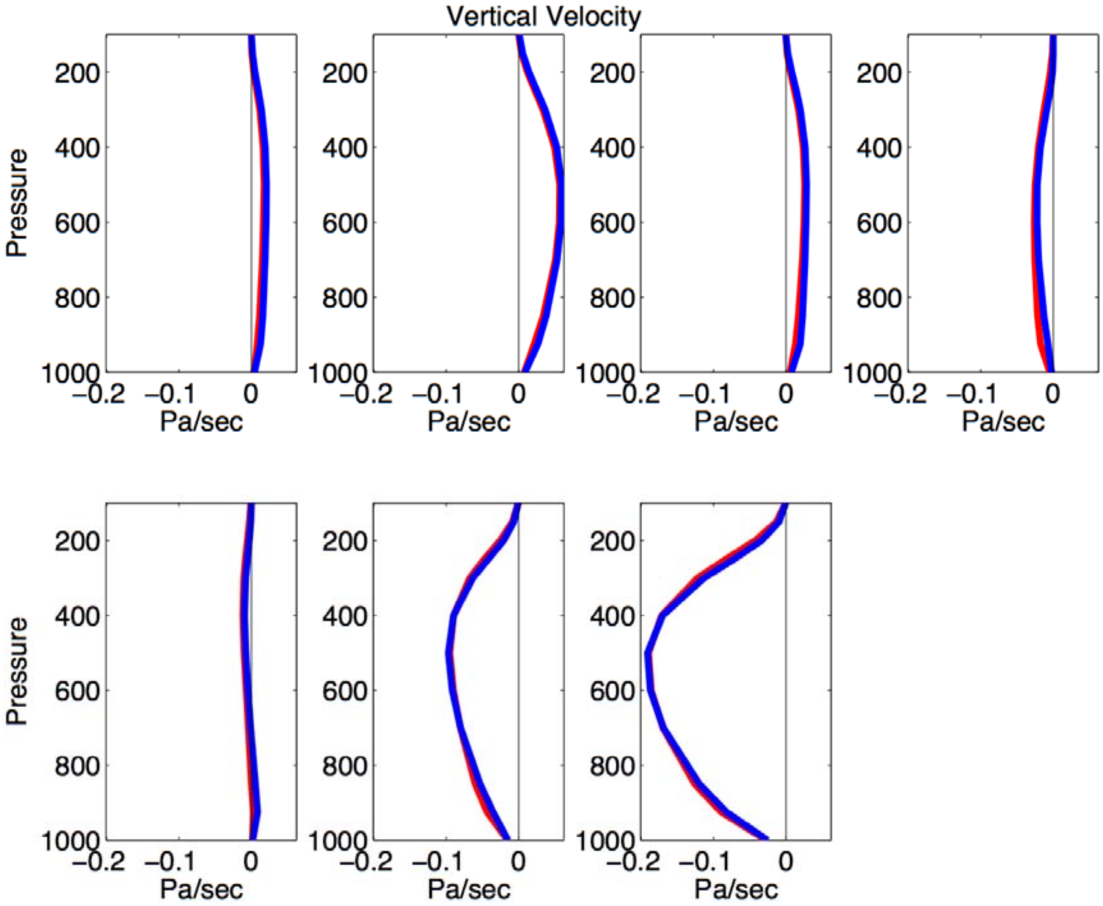


Figure 4.5 - Mean profile of anomaly pressure vertical velocity for warm (red) and cold (blue) subgroups of each cluster.

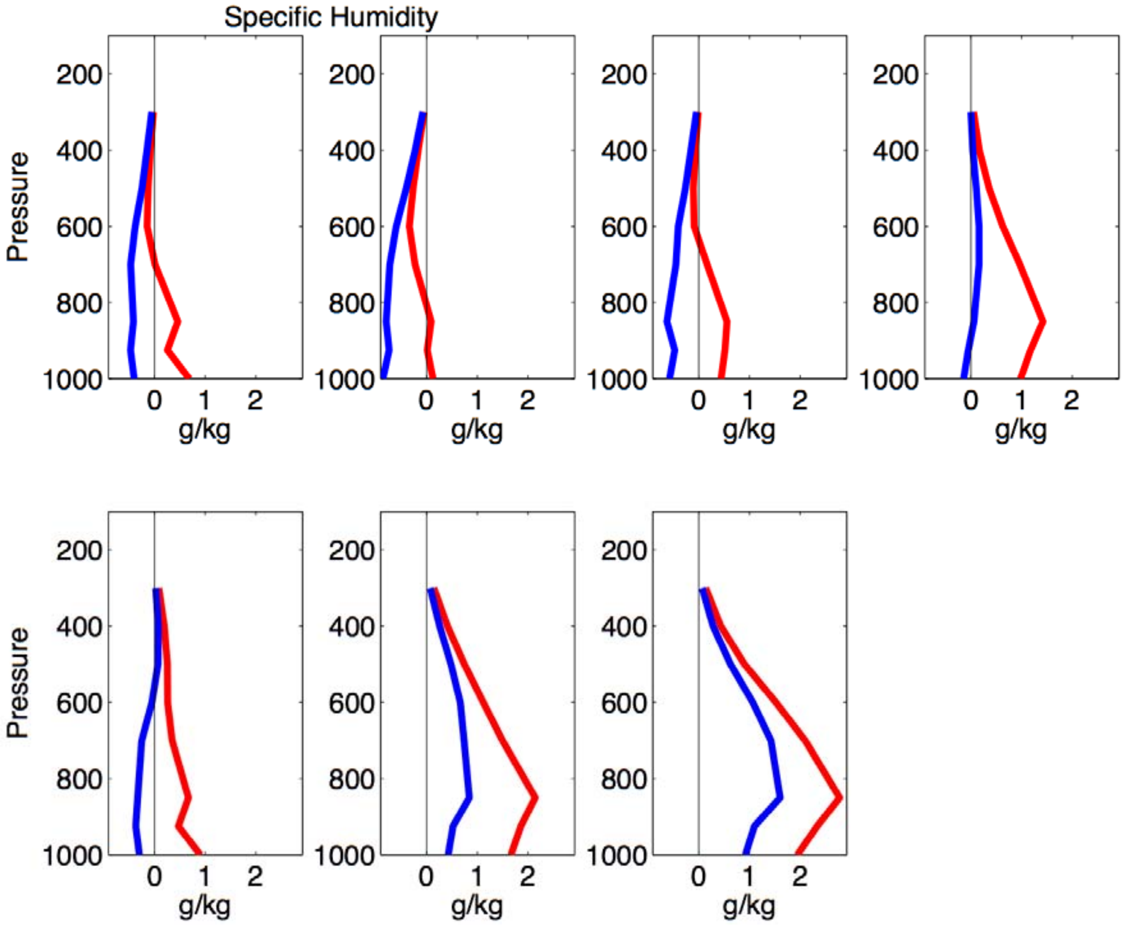


Figure 4.6 - Mean profile of anomaly specific humidity for warm (red) and cold (blue) subgroups of each cluster.

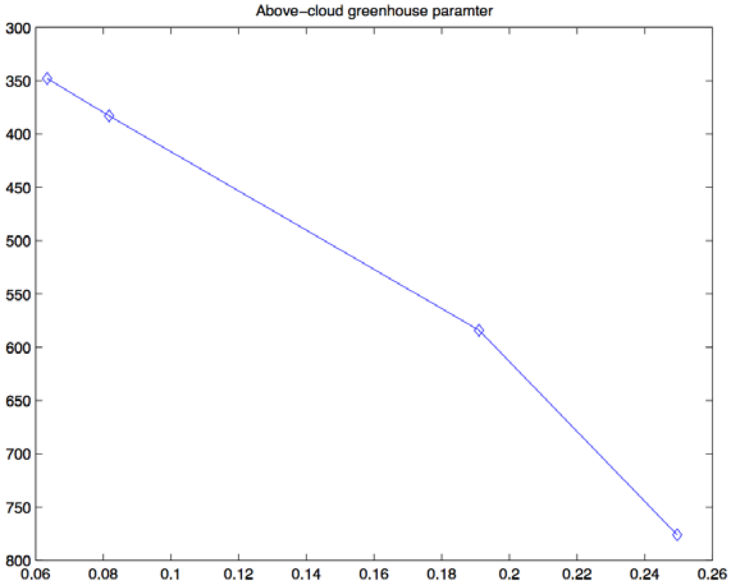


Figure 4.7 – Relationship between the above-cloud greenhouse parameter and cloud-top pressure.

CONCLUSION

Clouds cool Earth’s surface by reflecting incoming solar radiation and warm the surface by absorption of longwave radiation, a subtle balance that makes clouds a crucial component of Earth’s climate. Yet the representation of clouds in climate models continues to be the largest source of uncertainty in simulations of future climate [*IPCC,* 2007]. In order to accurately predict changes to our climate system, it is critical that we understand the character and substance of clouds and the exact nature of the role they play in the climate system.

This study has built a method whereby a k-means clustering algorithm is applied to satellite-derived cloud properties. This statistical routine builds groups of cloud scenes that all have similar properties. By looking at how these cloud properties vary within relatively homogeneous groups, we can better understand the sensitivity of cloud regimes to changes in the atmosphere.

The first section demonstrates how satellite observations of midlatitude cool-season continental cloudiness can be grouped into distinct cloud regimes by application of a k-means clustering algorithm to gridbox mean cloud fraction, cloud reflectivity, and cloud-top pressure. The cloud regimes correspond to typical cloud types associated with various synoptic conditions over land during winter: extensive cirrus, patchy cirrus, frontal/nimbostratus, stratus/stratocumulus, cumulus/cirrus, and a mixture of clouds at a variety of levels. Averages of ground-based retrievals of cloud fraction, cloud height, and cloud thickness are consistent with satellite cloud distributions and provide additional insight into the vertical structure of the cloud regimes. Consistency is also found between cloud properties of each regime and vertical profiles of meteorological parameters averaged over a domain approximately the size of a GCM gridbox and constrained by a dense network of observations to conserve column-integrated mass, water, energy, and momentum. In particular, a close relationship is found between mean vertical profiles of water vapor advection, relative humidity, and cloudiness for each regime. We investigated cloud properties over the ARM SGP site, since it was one of the few locations on Earth with accurate observations of water-vapor advection, but our general results should be applicable to many midlatitude land regions around the globe.

The second section of our study builds on this clustering algorithm and extends it to all midlatitude ocean basins. The data used for this study were cloud fraction in three bins of cloud-top pressure and three bins of cloud optical thickness. Atmospheric dynamics, derived from NCEP-NCAR Reanalysis, allowed us to consider these cloud clusters as part of the synoptic environment in which they reside. Our findings were also consistent with the spatial distribution of each of the clouds, allowing us to label these cloud clusters as forced by the synoptic meteorology. By analyzing large areas, as opposed to a single observation site, we are better able to compare the relationships between clusters. The midlatitude ocean clusters are very similar to those found over the ARM SGP site, with the marked exception that low clouds are much more frequent over the oceans.

Clustering provides a tool for examining large amounts of data and extracting information about patterns within the data. The clustering of cloud properties can group observations with similar large-scale dynamics, an effective method of diagnosing the ability with which GCMs will accurately predict cloud properties. Instead of requiring that models reproduce zonal or seasonal means, we can learn whether a model can recreate the observed cloud clusters with the same frequency and spatial distribution as is observed in nature. Also, if certain clusters are not simulated well, we can compare the model-simulated dynamics for a given cluster to the observed dynamics to better determine the source of a model's deficiency.

Additionally, since within each cluster there are relatively uniform dynamics, we can determine the thermodynamic change in cloud properties due to surface warming. These clusters are then used as the initial kernel for producing a set of observations whose dynamics are relatively homogeneous. In addition to cluster membership, other dynamics were factored in so as to minimize variability in dynamics that are important for cloud formation. We can then isolate changes in cloud properties resulting from thermodynamic changes in the atmosphere.

Across clusters, there is a consistent reduction in cloud fraction and an increase in optical depth for warm clouds. The changes in cloud-top pressure are less consistent. The inclusion of the ISCCP flux data allows us to examine how these changes in cloud properties affect the radiative flux in the atmosphere. If our study is to be applicable to the climate change scenario, where an external forcing results in a near-uniform warming of the troposphere, we must take caution in the analysis of our results. Instead of considering the difference in LWCRF between the warm and cold subgroups, we instead examine how the LWCRF is sensitive to individual cloud properties. The sum of these changes can then provide information about the change in LWCRF in the climate change scenario.

To understand changes in the shortwave flux, it is important to carefully normalize the TOA fluxes by the insolation, so as to consider mean radiative effects and so as not to introduce excess influence from summertime and equatorward points. The changes in the SWCRF could then be divided into those resulting from changes in cloud fraction or cloud albedo. For all clusters except Sc/St, the two effects were competing, with a reduction in cloud fraction leading to a warming and an increase in cloud albedo leading to a cooling. The partial cancellation of shortwave effects, along with strong changes in cloud-top temperature resulting from a lowering of cloud-top, leads to an average cloud forcing of -0.47 W/m^2 per degree K, which suggests that clouds could act as a negative feedback on the climate system.

REFERENCES

Ackerman, T.P. and G.M. Stokes (2003), The Atmospheric Radiation Measurement Program. Phys. Today, 56, 38-44.

Ardanuy, P.E., L.L. Stowe, A. Gruber, M. Weiss, and C.S. Long, 1989: Longwave Cloud Radiative Forcing as Determined from Nimbus-7 Observations. J. Clim., 2, 766–799.

Bony, S., K.-M. Lau, and Y.C. Sud (1997), Sea surface temperature and large-scale circulation influences on tropical greenhouse effect and cloud radiative forcing. J. Clim., 10, 2055-2077.

Bony, S., J.-L. Dufresne, H. Le Treut, J.-J. Morcrette, and C. Senior (2004), On dynamic and thermodynamic components of cloud changes. Clim. Dyn., 22, 71-86.

Clothiaux, E.E., T.P. Ackerman, G.G. Mace, K.P. Moran, R.T. Marchand, M.A. Miller, and B.E. Martner (2000), Objective Determination of Cloud Heights and Radar Reflectivities Using a Combination of Active Remote Sensors at the ARM CART Sites. J. Appl. Meteor., 39, 645-655.

GFDL Global Atmospheric Model Development Team (2004), The new GFDL global atmosphere and land model AM2/LM2: evaluation with prescribed SST simulations. J. Climate, 17 (24), 4641-4673.

Ghan, S., D. Randall, K.-M. Xu, R. Cederwall, D. Cripe, J. Hack, S. Iacobellis, S. Klein, S. Krueger, U. Lohmann, J. Pedretti, A. Robock, L. Rotstayn, R. Somerville, G. Stenchikov, Y. Sud, G. Walker, S. Xie, J. Yio, and M. Zhang (2000), A comparison of single column model simulations of summertime midlatitude continental convection, J. Geophys. Res., 105, 2091-2124.

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

Hahn, C.J. and S.G. Warren (1999), Extended Edited Cloud Reports from Ships and Land Stations over the Globe, 1952-1996. Numerical Data package NDP-026C, Carbon Dioxide Information Analysis Center (CDIAC), Department of Energy, Oak Ridge, Tennessee (79 pp).

Hartigan, J.A., 1975: Clustering Algorithms. Wiley Press, New York, NY, 351pp.

IPCC, 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, C.A. Johnson (eds). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881pp.

IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.

Jakob, C. (2003), An improved strategy for the evaluation of cloud parameterizations in GCMs. Bull. Amer. Meteor. Soc., 84, 1387-1401.

Jakob, C. and G. Tselioudis (2003), Objective Identification of Cloud Regimes in the Tropical West Pacific. Geophys. Res. Lett., 30 (21), Art. No. 2082.

Jakob, C., G. Tselioudis, and T. Hume (2005), The radiative, cloud and thermodynamic properties of the major Tropical Western Pacific cloud regimes. J. Climate, 18, 1203-1215.

Kalnay et al. (1996), The NCEP/NCAR 40-year reanalysis project. Bull. Amer. Meteor. Soc., 77, 437-470.

Katzfey, J.J. and B.F. Ryan (2000), Midlatitude frontal clouds: GCM-scale modeling implications. J. Climate, 13, 2729-2745.

Kim, B.-G., S. Klein, and J. Norris (2004), Continental liquid-water cloud variability and its parameterization using ARM data. J. Geophys. Res., 110, D15S08, doi:10.1029/2004JD005122..

Klein, S.A. and C. Jakob (1999), Validation and Sensitivities of Frontal Clouds Simulated by the ECMWF Model. Mon. Wea. Rev., 127, 2514-2531.

Lin, W.Y. and M.H. Zhang (2004), Evaluation of clouds and their radiative effects simulated by the NCAR Community Atmospheric Model against satellite observations. J. Climate, 17, 3302-3318.

Lock, A.P., A.R. Brown, M.R. Bush, G.M. Martin, and R.N.B. Smith (2000), A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. Mon. Wea. Rev., 128, 3187-3199.

Moorthi, S. and M.J. Suarez (1992), Relaxed Arakawa-Schubert: a parameterization of moist convection for general circulation models. Mon. Wea. Rev., 120, 978-1002.

Norris, J.R. and C.P. Weaver (2001), Improved Techniques for Evaluating GCM Cloudiness Applied to the NCAR CCM3. J. Climate, 14, 2540-2550.

Norris, J.R. and S.F. Iacobellis (2005), North Pacific cloud feedbacks inferred from synoptic-scale dynamic and thermodynamic relationships. J. Climate, 18, 4862-4878.

Ramanathan, V., B.R. Barkstrom, and E.F. Harrison (1989), Climate and the Earth's Radiation Budget. Physics Today, 42(5): 22-33.

Rossow, W.B., A.W. Walker, D.E. Beuschel, and M.D. Roiter (1996), International Satellite Cloud Climatology Project (ISCCP) Documentation of New Cloud Datasets. WMO/TD-No. 737, World Meteorological Organization, 115 pp.

Rossow, W.B. and R.A. Schiffer (1999), Advances in Understanding ISCCP. Bull. Amer. Meteor. Soc., 80, 2261-2287.

Rotstayn, L.D. (1997), A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. I: Description and evaluation of microphysical processes. Quart. J. Roy. Meteor. Soc., 123, 1227-1282.

Rotstayn, L.D., B.F. Ryan, and J. Katzfey (2000), A scheme for calculation of the liquid fraction in mixed-phase clouds in large-scale models. Mon. Wea. Rev., 128, 1070-1088.

Ryan, B.F., J.J. Katzfey, D.J. Abbs, C. Jakob, U. Lohmann, B. Rockel, L.D. Rotstayn, R.E. Stewart, K.K. Szeto, G. Tselioudis, and M.K. Yau (2000), Simulations of a cold front by cloud-resolving, limited-area, and large-scale models, and a model evaluation using in situ and satellite observations. Mon. Wea. Rev., 128, 3218-3235.

Tiedtke, M. (1993), Representation of clouds in large-scale models. Mon. Wea. Rev., 121, 3040-3061.

Tselioudis, G. and C. Jakob (2002), Evaluation of midlatitude cloud properties in a weather and a climate model: dependence on dynamic regime and spatial resolution. J. Geophys. Res., 107, 4781, doi:10.1029/2002JD002259.

Weare, B.C. and AMIP Modeling Groups (1996), Evaluation of the Vertical Structure of Zonally Averaged Cloudiness and Its Variability in the Atmospheric Model Intercomparison Project. J. Climate, 9, 3419-3431.

Weaver, C.P. and V. Ramanathan (1996), The Link Between Summertime Cloud Radiative Forcing and Extratropical Cyclones in the North Pacific. J. Climate, 9: 2093-2109.

Weaver, C.P., J.R. Norris, N.D. Gordon, and S.A. Klein (2004), Dynamical controls on mesoscale cloud variability for ARM Case 4. J. Geophys. Res., 110, D15S05, doi:10.1029/2004JD005022.

Webb, M., C. Senior, S. Bony, and J.-J. Morcrette (2001), Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF, and LMD atmospheric climate models. Clim. Dyn., 17, 905-922.

Williams K.D. and G. Tselioudis (2007), GCM intercomparison of global cloud regimes: Present-day evaluation and climate change response. Clim. Dyn. doi:10.1007/s00382-007-0232-2.

Wylie, D.P. and H.M. Woolf (2002), The diurnal cycle of upper-tropospheric clouds measured by GOES-VAS and the ISCCP. Mon. Wea. Rev., 130, 171-179.

Xie, S., R.T. Cederwall, M. Zhang, and J.J. Yio (2003), Comparison of SCM and CSRM forcing data derived from the ECMWF model and from objective analysis at the ARM SGP site. J. Geophys. Res., 108, 4499/2003JD003541.

Xie, S., R.T. Cederwall, and M. Zhang (2004), Developing long-term single-column model/cloud system resolving model forcing data using numerical weather prediction products constrained by surface and top of atmosphere observations. J. Geophys. Res., 109, D01104, doi:10.1029/2003JD004045.

Xie, S., M. Zhang, M. Branson, R.T. Cederwall, A.D. Del Genio, Z.A. Eitzen, S.J. Ghan, S.F. Iacobellis, K.L. Johnson, M. Khairoutdinov, S.A. Klein, S.K. Krueger, W. Lin, U. Lohmann, M.A. Miller, D.A. Randall, R.C.J. Somerville, Y.C. Sud, G.K. Walker, A. Wolf, X. Wu, K.-M. Xu, J.J. Yio, G. Zhang, J. Zhang (2004), Simulations of midlatitude frontal clouds by SCMs and CRMs during the ARM March 2000 Cloud IOP. J. Geophys. Res., 109 (D1): Art. No. D01104 JAN 9 2004.

Xu, K.-M., M. Zhang, Z.A. Eitzen, S.J. Ghan, S.A. Klein, X. Wu, S. Xie, M. Branson, A.D. Del Genio, S.F. Iacobellis, M. Khairoutdinov, W. Lin, U. Lohmann, D.A. Randall, R.C.J. Somerville, Y.C. Sud, G.K. Walker, A. Wolf, J.J. Yio, J. Zhang (2004), Modeling springtime shallow frontal clouds with cloud-resolving and single-column models. J. Geophys. Res., 110 (D15): Art. No. D15S04 APR 21 2005.

Zhang, M.H., J.L. Lin, R.T. Cederwall, J.J. Yio, and S.C. Xie (2001), Objective Analysis of ARM IOP Data: Method and Sensitivity. Mon. Wea. Rev., 129, 295-311.

Zhang, Y., W.B. Rossow, A.A. Lacis, V. Oinas, and M.I. Mishchenko (2004), Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. J. Geophys. Res., 109, D19105, doi:10.1029/2003JD004457.

Zhang, M.H., W.Y. Lin, S.A. Klein, J.T. Bacmeister, S. Bony, R.T, Cederwall, A.D. Del Genio, J.J. Hack, N.G. Loeb, U. Lohmann, P. Minnis, I. Musat, R. Pincus, P. Stier, M.J. Suarez, M.J. Webb, J.B. Wu, S.C. Xie, M.-S. Yao, J.H. Zhang (2004), Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. J. Geophys. Res., 110 (D15): Art. No. D15S02 MAY 3 2005.