10/7-21-97 (1)

SANDIA REPORT

SAND97-8252 • UC–402
Unlimited Release
Printed April 1997

M97053780

Data and Image Fusion for Geometrical Cloud Characterization

Lawrence R. Thorne, Kenneth A. Buch, Chen-Hui Sun, and Carl Diegert

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94551
for the United States Department of Energy
under Contract DE-AC04-94AL85000

Approved for public release; distribution is unlimited.

MASTER

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

SAND97-8252 • UC-402
Unlimited Release
Printed April 1997

Data and Image Fusion for Geometrical Cloud Characterization

by

Lawrence R. Thorne, Kenneth A. Buch, Chen-Hui Sun
Sandia National Laboratories
Livermore, California

and

Carl Diegert
Sandia National Laboratories
Albuquerque, New Mexico

Abstract

Clouds have a strong influence on the Earth's climate and therefore on climate change. An important step in improving the accuracy of models that predict global climate change, general circulation models, is improving the parameterization of clouds and cloud-radiation interactions. Improvements in the next generation models will likely include the effect of cloud geometry on the cloud-radiation parameterizations. We have developed and report here methods for characterizing the geometrical features and three-dimensional properties of clouds that could be of significant value in developing these new parameterizations. We developed and report here a means of generating and imaging synthetic clouds which we used to test our characterization algorithms; a method for using Taylor's hypotheses to infer spatial averages from temporal averages of cloud properties; a computer method for automatically classifying cloud types in an image; and a method for producing numerical three-dimensional renderings of cloud fields based on the fusion of ground-based and satellite images together with meteorological data.

ACKNOWLEDGMENTS

We wish to acknowledge and thank W. P. Kegelmeyer, Jr. and M. C. Allmen of Sandia National Laboratories for many helpful discussions regarding image processing issues associated with this work; Janet Shields, Richard Johnson and T. L. Koehler of the Marine Physical Laboratory of the Scripps Institution of Oceanography for discussions related to the Whole-Sky Imager operation, data products and for providing the imagers and data reduction for our two-imager studies; and Robert Endlick of the White Sands, New Mexico, Testing Range for providing the facilities for the same imager studies.

This work was supported by Sandia's, Laboratory-Directed Research and Development program (LDRD).

CONTENTS

FIGURES

TABLES

1. Introduction

A major Department of Energy (DOE) goal is to improve the accuracy of general circulation models (GCM's) used to analyze and predict the timing and magnitude of "greenhouse" gas-induced global warming. The primary concern comes from the global climate change that may be triggered by the huge quantities of CO_2 that are emitted annually from burning fossil fuels to produce energy. CO_2 is considered a radiatively active gas or "greenhouse gas" because it transmits incoming solar radiation (short-wave radiation) but it absorbs and reradiates terrestrial radiation (long-wave radiation) in all directions (including back to the earth's surface). The over all effect of atmospheric CO_2 is to increase the absorption of solar radiation by the earth-atmosphere system (i.e. the surface and troposphere) [1]. This increased absorption may be viewed as a radiative climate forcing that would not exist if CO_2 were not present. For the present day CO_2 concentrations of ~365 ppm, the radiative forcing amounts to about 50 W m^{-2}. This compares with about 100 W for water vapor and clouds together, and ~4.5 W m^{-2} for all other radiatively active gases combined. The total forcing of ~155 W m^{-2} due to radiatively active atmospheric gases causes the earth's surface to be about 33 °C warmer than if clouds and the radiatively active gases were absent [1].

What concerns DOE and others is the effect of increasing amounts of CO_2 on the climate. The half-life of CO_2 in the atmosphere is ~150 years and the levels of CO_2 prior to the industrial revolution were ~265-290 ppm, as estimated from air bubbles trapped in ice cores. So the current levels of 365 ppm represent an increase of ~85 ppm and indicate that the earth-atmosphere system cannot absorb the additional CO_2 as rapidly as it is being emitted by human activities. Because the levels of CO_2 seem to be constantly increasing, those who study the climate effect of added CO_2 often use the conditions resulting from doubling the pre-Industrial Revolution levels as a point of comparison, i.e., an atmosphere with 560 ppm CO_2.

The additional radiative forcing resulting from doubling the atmospheric CO_2 concentration from pre-Industrial Revolution levels can be calculated accurately; it is ~4 W m^{-2}. Unfortunately, there is no easy way to determine the response of important climate variables, such as surface temperatures or the vigor of the hydrological cycle, from this increased radiative forcing. The earth-atmosphere system is just too nonlinear and the coupling between individual processes too complex. Thus, DOE's efforts to predict the effects of increasing CO_2 concentrations centers on developing and improving computer models that attempt to account for the major processes and couplings in the earth-atmosphere system, specifically, general circulation models, GCM's.

To reveal and quantify the model processes that need further refinement, DOE recently sponsored a comparison between eleven GCM's. One of the requirements of the GCM's was to predicted the surface temperature change resulting from doubled CO_2. The predicted surface temperature increase ranged from 0 to 9 °C. The models had the greatest agreement for cloud-free skies and for totally overcast skies. Model results for partially overcast skies showed the greatest disagreement. The source of the disagreement was traced to cloud-radiation interactions--not the radiative forcing due to CO_2 [2].

These results highlight the highly non-linear nature of climate couplings. In fact, other research has shown that cloud radiative feedback is the single most important effect determining the magnitude of possible climate responses to human activity [1-10]. Yet clouds are not well parameterized in GCM's and are, in fact, presently the greatest factor in limiting the accuracy of GCM's [2]. Thus, clouds exert the largest influence on radiative

forcing and climate while at the same time present the largest uncertainties in GCM models used to predict climate.

1.1 Need for 3-D Characterization of Clouds

One approach DOE is taking to better model cloud-radiation effects in GCM's is to improve the parameterizations for cloud formation and cloud properties. This effort is supported by a closely-coupled field measurements program [3]. Specifically, DOE has established the Atmospheric Radiation Measurement program, ARM, in which the Cloud and Radiation Testbed (CART) sites play a major role in providing the needed ground measurements. The CART sites are being used as an experimental testbed for the study of models of the terrestrial radiation field, properties of clouds, the full life cycle of clouds and the incorporation of these process-level models into climate models. As such, "an important feature of the ARM Program Plan is to establish a surface-based cloud imaging system at each of the research sites that will gather data that will aid in parameterizing solar flux over an entire [GCM atmospheric] grid cell [approximately 200 x 200-km ground footprint] " [3]. A further requirement is to provide some form of cloud "visualization" system for mapping cloud extent and cloud typing [3], i.e., for characterizing the geometrical properties of clouds.

For the ARM program, measuring macroscopic cloud properties is part of an over all three-dimensional (3-D) mapping strategy for atmospheric water in all its meteorological significant phases (vapor, liquid, ice); this is required to understand the spatial effects of cloud processes. The need for mapping is driven by the fact that water is not well mixed in the troposphere over the length scales of DOE's CART study sites, so mapping is required to properly treat the variations in thermodynamic processes and radiation transport over the region. The principal requirements are to answer questions related to cloud life cycle: where do clouds form? under what local conditions? what governs their persistence and breakup? and how can answers to these questions be parameterized for inclusion into GCM's? For cloud-radiation interactions: how can the assumption of plane-parallel cloud shape be modified to capture the interactions of real, 3-D cloud shapes on the radiation field? (e.g., by using an "effective" cloud cover fraction which is greater than the actual geometrical cloud cover fraction). (Currently, plane-parallel (i.e., slab-like) is the only cloud shape used in GCM's. It is known to be a poor approximation for many cloud fields [10-12]).

Early studies of effective cloud cover fraction have shown that it is not only greater than the geometrical cloud cover fraction, but that the ratio of the effective cloud cover to the geometrical cloud cover increases as the cloud aspect ratio (horizontal length / vertical height) increases. The effect can be quite dramatic as demonstrated in a study in which clouds were modeled as cylinders [10]--a geometrical cloud cover fraction of 0.3 for clouds having an aspect ratio of 2 gives an effective cloud cover fraction in terms of radiation interactions equivalent to 0.7 for plane parallel clouds. Clouds modeled as cubes [11] and spheres [13] gave similar results. In terms of the radiation budget, ignoring cloud aspect ratio can lead to calculated tropospheric heating rates that are in error by 10-15% [10]. These studies indicate that most important cloud characteristics that govern cloud-radiation interactions are the cloud cover fraction, cloud aspect ratio (both geometrical properties of clouds) and the cloud optical depth [10].

While 3-D characterization of clouds is a high priority for CART measurements, the only demonstrated method for obtaining explicit 3-D measurements of clouds over wide areas is the volume-imaging lidar developed by Eloranta and coworkers [12]. This technique has revealed stunning detail about the 3-D structure of thin cirrus clouds showing

quite inhomogeneous and hole-riddled shapes. Unfortunately, the penetration depth of the lidar probe laser light and return signal is limited by the strong light scattering caused by cloud particles. Thus, adequate volume lidar return signals are only possible for optically thin clouds (of optical depths of 1 or less; but optical depths of most clouds of interest range from 1-20). This limits the volume-imaging lidar technique to 3-D visualization of thin cirrus clouds and aerosols.

Thus at the present time, no single imaging technique is capable of 3-D cloud characterization for a wide range of cloud types and optical depths. So, by necessity, 3-D cloud characterization will require fusing images from several sources (e.g., ground-based and satellite imagers) as well as fusing data from a variety of meteorological instruments. This is a formidable but interesting task and it is the goal of our LDRD-sponsored work that we report here.

More specifically, the primary objective of our LDRD-sponsored research is to develop methods for spatial and 3-D characterization of clouds that are applicable to a wide range of cloud types and coverages and that are based on field measurements and images. While the end goal of the DOE's ARM program is better parameterization of GCM's through field measurements, we have limited our LDRD efforts to a related essential subtask--that of improving geometrical cloud characterizations. These geometrical characterizations will be useful in developing the next generation of GCM cloud parameterizations in which cloud shape will be a consideration in the cloud-radiation interactions. Even though we have developed these methods for naturally occurring clouds, they are equally applicable to characterizing other types of clouds that may be artificially produced in the atmosphere (e.g., from military field operations involving obscurants and the plumes from large forest or oil well fires etc.).

The kinds of spatial cloud characterizations that we have developed and report here range from statistically-derived cloud cover fraction to the high-resolution 3-D numerical reconstruction of a cumulus cloud and stratus cloud field.

We have placed emphasis in our LDRD program on cumulus clouds and cumulus cloud fields. Cumulus clouds are often present when there is strong vertical motion of a moist air mass. These conditions are climatologically significant because they serve to pump moisture and energy into the lower atmosphere (Troposphere) and even into the lower Stratosphere when there is strong upward convection (e.g., the tropical regions near the equator). In addition, from an imaging stand point, cumulus clouds are more easily defined spatially because they have more distinct boundaries than most other type of clouds. Thus, cumulus clouds are ideal for our image and data fusion studies, because they are important in climate change studies and they are more easily imaged than other types of clouds.

1.2 Relationship of 3-D cloud reconstruction to GCM's

Present day GCM's cannot possibly handle the massive amount of data that would be required to map clouds over the entire surface of the Earth at some meteorologically significant scale, such as, the scale over which cloud radiative properties can vary (~1 km). Thus, clouds and cloud effects must be parameterized in GCM's for them to be computationally feasible. So, how can GCM's benefit from the highly detailed information embodied in a 3-D rendering of clouds or of any detailed local cloud field description?

This problem is currently being addressed in the ARM program through the use of cloud ensemble models. These models produce ensemble-averaged properties using detailed cloud information as input [9]. Such models are needed for averaging cloud-radiation interactions, because the nonlinearities of the various atmospheric process invalidates simple linear averaging. Thus, the principle utility of our LDRD work for climate change studies is to provide real cloud field geometries as input to cloud ensemble models which then produce appropriately averaged properties that can subsequently be used for GCM parameter improvement.

2. Imaging the Sky Dome

In our LDRD work, we have used extensively the images produced by the Whole-Sky Imager. This is an imager developed by the Marine Physical Laboratory (MPL) at the Scripps Institute of Oceanography [14]. Several models have been developed. The one we have used most extensively is the EO-5. This instrument uses a fish-eye lens and charge injection device (CID) imaging array detector. An example of a WSI image is shown below.

In order to image the whole sky, the scene must be projected on to a flat image plane. Several types of projections are possible, but perhaps the most straightforward for an electronic imaging system is the equal-angle projection. In the MPL WSI, this projection is produced by a fish-eye lens. The equal-angle projection preserves the azimuthal angle (compass angle). But it maps the angle between the overhead direction to the object in the scene (i.e., the zenith angle) as a distance from the center of the image. Thus, objects directly overhead appear in the center of the image; objects near the horizon appear near the outer edge of the image. The imaging specification of the imager are given in Section 6 below. For zenith angles in the range of 0-70°, the distorted WSI image may be "made flat" using a Cartesian transformation developed by MPL [15] and Sandia [16]. The entire range of angles from 0-90° cannot be flattened, because the transformation produces extreme distortion at large zenith angles. This is an intrinsic problem because the points at the horizon must map into infinite distance in the flattened image. Nevertheless, flattened images have proved to be very useful in this study be cause they may be compared directly to satellite images.

Interpolation and smoothing are needed in producing the flattened images, because the linear spatial resolution of a WSI pixel decreases as the radial distance from the imager to the object increases (as opposed to the angular resolution which remains constant). When the spatial resolution of the WSI is finer than that of the 3D cloud field, linear interpolation is used to provide a unique value for the WSI image pixel. When the spatial resolution of WSI is more coarse than that of the 3D cloud field, the point on the cloud surface that projects closest to the WSI point is used.

We have also used satellite images in this study to provide additional cloud image data for wide area coverage. Satellite-based imagers produce down-looking, planar images of the cloud fields. Brief descriptions of the satellite imaging systems that we used in this study are given in Section 6 below. Section 6 also shows a WSI image and a satellite of the same cloud field. A very useful result from comparing these images is that the cloud profile shapes within the cumulus cloud field are very similar whether viewed from above or from beneath. This observation facilitated our 3D reconstruction because it made navigating the images form the satellite much easier.

3.0 Synthetic Cloud Images for Testing Image Processing Algorithms

In order to develop cloud base height and 3D cloud reconstruction algorithms, we have developed a numerical testbed. This is a suite of computer programs that produce synthetic clouds and then create synthetic images of these clouds given the properties of the imaging system. This testbed has proven to be extremely useful in that we could "produce" clouds at a specified height and cloud cover fraction, then produce a sequence of images of these clouds as they would be imaged by the WSI. We then used the algorithms that we were developing to "recover" the cloud base height and other cloud geometrical properties from these synthetic WSI images. In this way we were able to judge the accuracy of the extracted data because we knew the "correct" answer before hand. This also allowed us to produce and evaluate our algorithms on many more cloud field types than we would have been able to collect easily form field images.

As the first step in developing the testbed, we developed the equations and algorithm for projecting a scene onto the WSI image. Figure 1 shows a simulation of square array of cylinders with aspect ratio of 1 projected through the Whole-Sky Imager electro-optical system. The image is not circular because the imager CCD uses a 480 X 360 imaging array. The bottoms of the cylinders are shaded darker than the sides to make them more distinguishable. This image shows clearly how the sides of clouds can be visualized with the WSI for clouds that are not directly overhead. This effect can be seen in the WSI image of a real cloud scene in Figure 2, a WSI image of a real cumulus cloud field taken at White Sands New Mexico in the summer of 1992. The cloud bottoms are darker in the image and to the human observer, because they are shaded by the mass of cloud above them which scatters the suns rays. The black, rectangular- shaped portion of the image is the portion of the image that is blocked by the imager's sun occultor. The occultor blocks out the direct solar rays. This is needed because the direct rays would cause the pixels around the sun's image to "bloom" and obscure much of the image. The direct rays would also cause specular reflection off of particles and scratches on the lens and WSI dome that would produce undesired white flecks in the image.

To produce numerical synthetic clouds, we adapted to our needs a fractal cloud model developed at TASC by Cianciolo [17]. The cloud scene simulation model is a fractal-based model which generates a time series of cloud fields. The sky is represented as a cube of data with each cell in the cube having some amount of liquid water. If the liquid water content is sufficiently high in a cell, then that cell is considered to be part of a cloud. By varying the distribution and amount of water in the cube volume, different types of clouds are produced. To generate a cloud field, the cloud cover, cloud type, cloud base and top, horizontal domain size, grid resolution and the atmospheric temperature, wind and humidity profiles, etc. are given as inputs. By tuning the parameters in the model, the morphology and texture of different cloud types including cumulus, stratus and cirrus can be simulated. Also, the input wind profile can be used to generate a time series of cloud fields. Figure 3 shows an example of a synthetic cloud field generated in this manner.

We then used these simulated -D cloud fields to form 2D WSI images. This is done by using a simple cloud illumination algorithm to generate the illumination effect of the cloud surface. The following assumptions are made.

The incoming light is parallel. Illumination of the cloud surface is inversely proportional to the sum of the water content along the light path which is in the same direction as the incoming light. Light is scattered uniformly in all directions in the clear air and there is no attenuation effect in the clear air. There are no reflection effects on the cloud surface.

Light rays enter the cloud in parallel. The cloud liquid water content, LWC, is summed along each ray and each summation is used as the illumination of the point where the ray exits the cloud. The illumination is the same regardless of viewing position underneath the cloud.

The third is that light is scattered uniformly in all directions, therefore, the gray level of a point in an image is the same regardless of the WSI's position relative to that point. With the illumination of the cloud surface points determined, the final phase of producing synthetic cloud images is to generate the WSI images of 3-D cloud scene in a manner that simulates the real WSI cameras. For a WSI with a specified angular resolution, each non-occluded cloud surface point visible to the WSI can be projected on to the WSI image plane given zenith and azimuthal angles of that point relative to the position of the WSI.

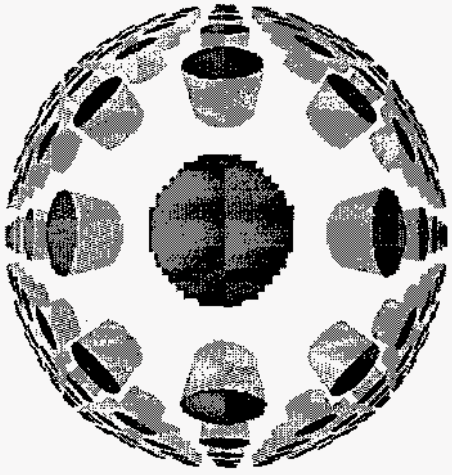


Figure 1. WSI image of a square array of cylinders.

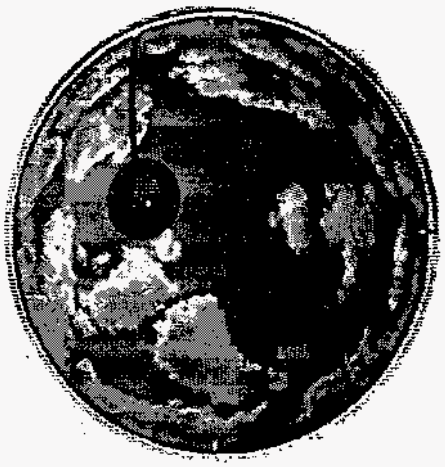


Figure 2. WSI image of a cumulus cloud field.

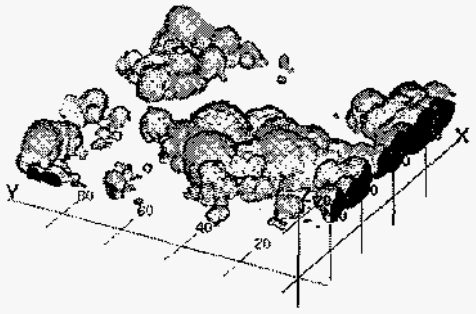


Figure 3. Cumulus cloud field simulated by the adapted TASC cloud simulation model

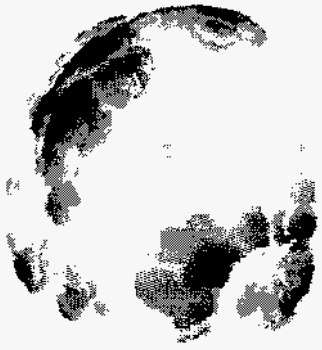


Figure 4. The simulated cloud field from Figure 3 as would be imaged by the WSI

Because of the complicated optical properties of clouds in the atmosphere, these assumptions are a simplification of the real-world process. Nevertheless, the 2D images look very much like real clouds and provided good objects with which to test our algorithms.

Figure 4 shows the WSI image of the synthetic cloud scene shown in Figure 3.

4. Low-Order Cloud Geometrical Characterizations: Inferring Spatial Cloud Statistics from Limited Field-of-View Zenith Observations

4.1 Introduction

Many of the measurements needed to improve the parameterizations of clouds and radiation used in GCM's are desired over a large fraction of the sky dome. In practice, however, many of the observations are limited to narrow fields of view around the zenith direction. Thus, significant benefit would be gained by a method that would permit the spatial properties of clouds to be inferred from sequences of narrow field-of-view zenith observations. We have developed one such procedure based on Taylor's hypothesis for turbulent flow [18].

Taylor's hypothesis provides a method of relating time-averages to spatial averages for statistical properties, provided certain conditions are satisfied. In this section we examine the applicability of Taylor's "frozen flow" hypothesis to various types of cloud fields recorded using the whole sky imager, WSI, mentioned above[14].

Taylor's hypothesis states that when the turbulence velocities are small compared to the mean flow velocity, the flow field is "frozen" and the spatial velocity correlation (taken in the mean flow direction) can be inferred from the temporal velocity correlation. This relation between spatial and temporal statistics has also been applied to an underlying scalar field in a turbulent flow, where the spatial statistics are obtained from measurements in the direction of the mean flow [18].

For cloud fields, we wish to apply Taylor's hypothesis to such parameters as cloud cover fraction, cloud radiance and cloud liquid water content. Thus, if Taylor's hypothesis holds, $R(dt) = R(dx)$ and $dx = U\,dt$ along the direction of the mean wind; where R(dt) is the correlation of the cloud parameter in time, $R(dx)$ is the correlation of the cloud parameter in space, and U is the mean wind speed.

The objective of this portion of our study is to determine under what conditions Taylor's hypothesis holds or does not hold for cloud fields. The WSI images provide a very useful data set for evaluating Taylor's hypothesis because it provides the full two-dimensional spatial result from which we can calculate the true spatial correlation and compare it with the correlation that we derive from the temporal data. In addition, the high temporal resolution of the two-dimensional WSI images provides the opportunity to extract one-dimensional time series and compute the mean flow speed and direction. We obtain the mean flow needed to test Taylor's hypothesis from WSI images using optical flow analysis [16] as discussed below.

We illustrate the application of Taylor's hypothesis by extracting a time series of narrow field-of-view, zenith cloud cover measurements from the full WSI image data.

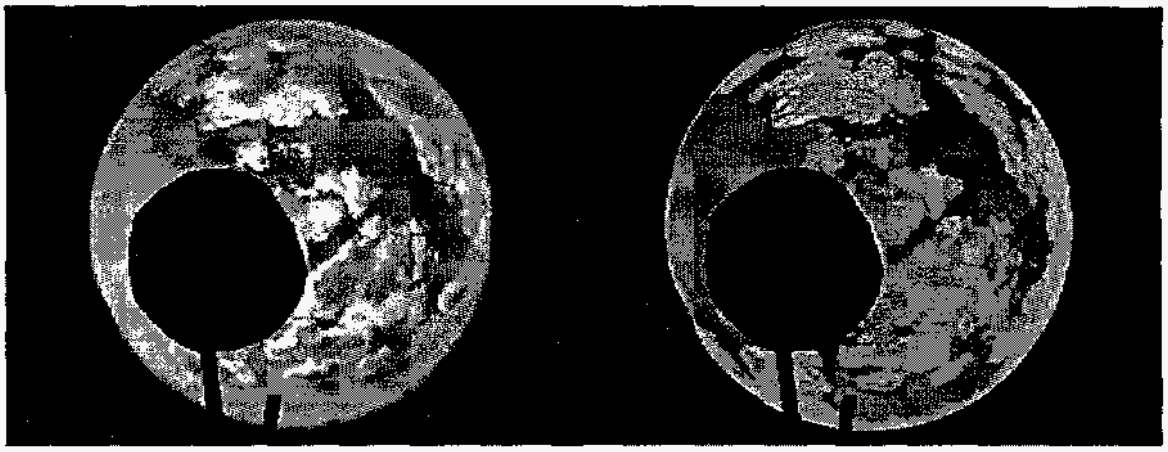


Figure 5. Left, blue-filtered image of cumulus cloud field. Right, cloud decision image determined from red and blue images of same scene. Black portion of the image corresponds to the sun shade (occultor).

Using these spatially-limited measurements, we use Taylor's hypothesis to estimate the cloud cover fraction over a 126° field of view that existed during the period of the time series data acquisition. This is analogous to using a time series of ceilometer measurements (which have a narrow field of view and indicate cloud/no cloud) to determine the cloud cover fraction. In this study we identify the cloud conditions under which the time series can yield reliable estimates of the cloud cover fraction. We also discuss the accuracy of the estimation.

4.2 Preparing the data set

The Whole Sky Imager (WSI) data set used in this analysis was collected at White Sands, New Mexico, during nine days in May 1992. A set of four optically filtered, digitized images were acquired every minute by each of the two Scripps, E/O 5 WSI's [14]. The WSI's were placed about 5 km. Each of the images in a set were acquired with a different filter; a blue filter, a red filter, a blue filter with a neutral density filter, and a red filter with a neutral density filter. From each set of images, a cloud decision image, which defines the cloudy and non-cloudy pixels in the image, was generated from the ratio of the blue and red filtered images and an estimate of the red-blue ratio for the clear sky as a function of azmuthal and zenith angles [19]. This data set has also been used to develop algorithms for determining the cloud base height from paired WSI's [16].

The WSI has a fish-eye lens that provides wide field-of-view images (~130°) of the sky dome with an angular resolution of 1/3°. The images are equal-angle projections with the distance from the center of the image being proportional to the zenith angle in the scene. Depending on the cloud height, horizontal cloud distributions from 10 km to 50 km in horizontal extent can be recorded. Figure 5 shows the blue image and the cloud decision image of a cumulus cloud field. Due to the character of the WSI, the images have to be processed (as described in the next section) before being analyzed. As described below, the cloud movement can be estimated by calculating the optical flow of the image using the methods developed by Allmen and Kegelmeyer [16].

4.2.1 WSI data processing

A few comments on the character of the WSI images as related to processing the images are appropriate. Because the WSI views clouds from beneath, primarily bottoms of clouds are seen near the center of the WSI image. Progressively more of the sides of the clouds are seen as the view approaches the horizon, i.e., edge of the WSI image. In addition, because of the projection properties of the fish-eye lens, the clouds appear to be more strongly distorted from their apparent shape in the scene the closer they are to the horizon. Because a common focus is difficult to achieve for a wide field of view, the clouds scenes are slightly blurred at the edge of the image.

Also, because of the changing scattering angle, the image is darker in a direction 90° away from the sun [19]. These effects will affect the accuracy of the cloud statistical calculations, because we assume that the apparent intensity (pixel grey level) is independent of viewing angle. We apply several corrections to the data in order to reduce these effects that would otherwise bias the statistical calculation. First, the dependence of the gray level on the scattering angle (*i.e.* the angular distance from the sun in the original WSI image) is calculated from an overcast stratus cloud image. The accuracy of this correction depends on the uniformity of the cloud layer thickness and the homogeneity of the cloud internal structure and composition, etc., but we believe the stratus covers to be adequate.

Then for cumulus cloud fields, a range of thresholds based on the relationship of gray level to scattering angle are determined and applied to the image to detect and filter out the bright cloud sides. We do this by assuming that the radiance measured from the cloud bottoms is similar to the radiance measured for the overcast stratus conditions that we used to correct for the angular dependence of the pixel gray level. Any pixels with radiances significantly brighter than the bottom radiances are assumed to be cloud sides and are removed from the image. Next, all cloud images are normalized by the overcast stratus correction image to make the gray level more uniform.

Then, the image is flattened by using a Pseudo-Cartesian coordinate transformation [16] to reduce the projection distortion of the fish-eye lens. This transformation estimates the Cartesian coordinates of each pixel in the WSI image assuming that the cloud base height is uniform. This is a reasonable assumption for many cumulus cloud fields because the bottom heights are close to the lifting condensation level which is usually quite uniform over a region of 100 km in extent. We found that the Pseudo-Cartesian coordinate transformation was needed in order to obtain usable spatial correlations. Without the transformation, the distortion introduced by the fish-eye lens gave correlations that were too low and inaccurate.

4.2.2 Optical flow calculation

WSI data sets provide a sequence of cloud images at one minute intervals. This temporal sampling rate is high enough to smoothly resolve the movement of the clouds. We use an optical flow algorithm which is based on a hierarchical correlation method to calculate the movement between two consecutive WSI images recorded by one of the WSI cameras [16]. This algorithm generates a vector indicating the direction and amount of motion at each pixel. In the hierarchical correlation method, each of the two images is blurred and subsampled so that the new image is half the size in each direction as the original. This process continues until the image is only a few pixels on each side. Each of these two sets of images is called an image pyramid. Starting at the top of the pyramids (i.e., with the images having the fewest pixels), the motion at each pixel is calculated by correlating the two images. Then, at the next level, the motion at each pixel is determined by correlating

points only in the areas in which the motion occurred at the level above. This process continues down to the bottom level (images whit the largest number of pixels) where the motion at each pixel in the original image is calculated. By using this hierarchical method, the motion at several different scales are considered, reducing the effect of random noise on the optical flow. This gives a smoother optical flow pattern which is more realistic than when the optical flow is computed from the highest resolution images alone. Figure 6 shows the optical flow vectors overlaid on top of the processed cumulus cloud base image.

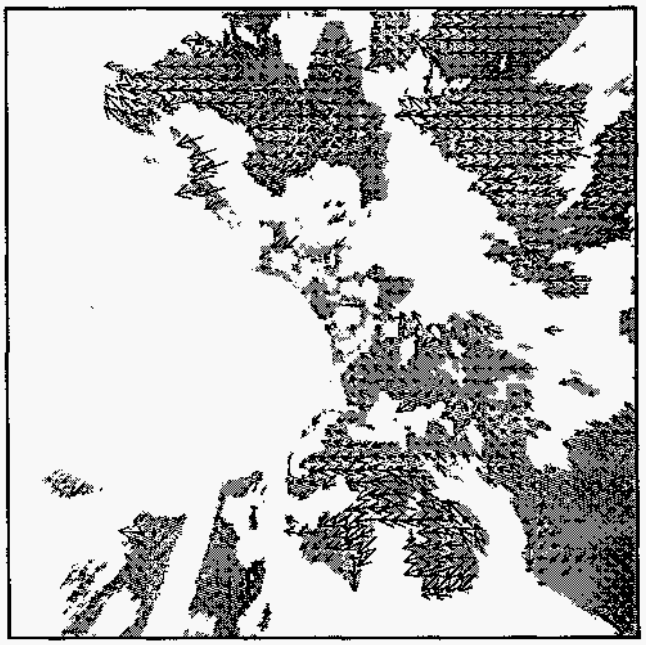


Figure 6. Optical flow overlaid on the processed cumulus cloud base image

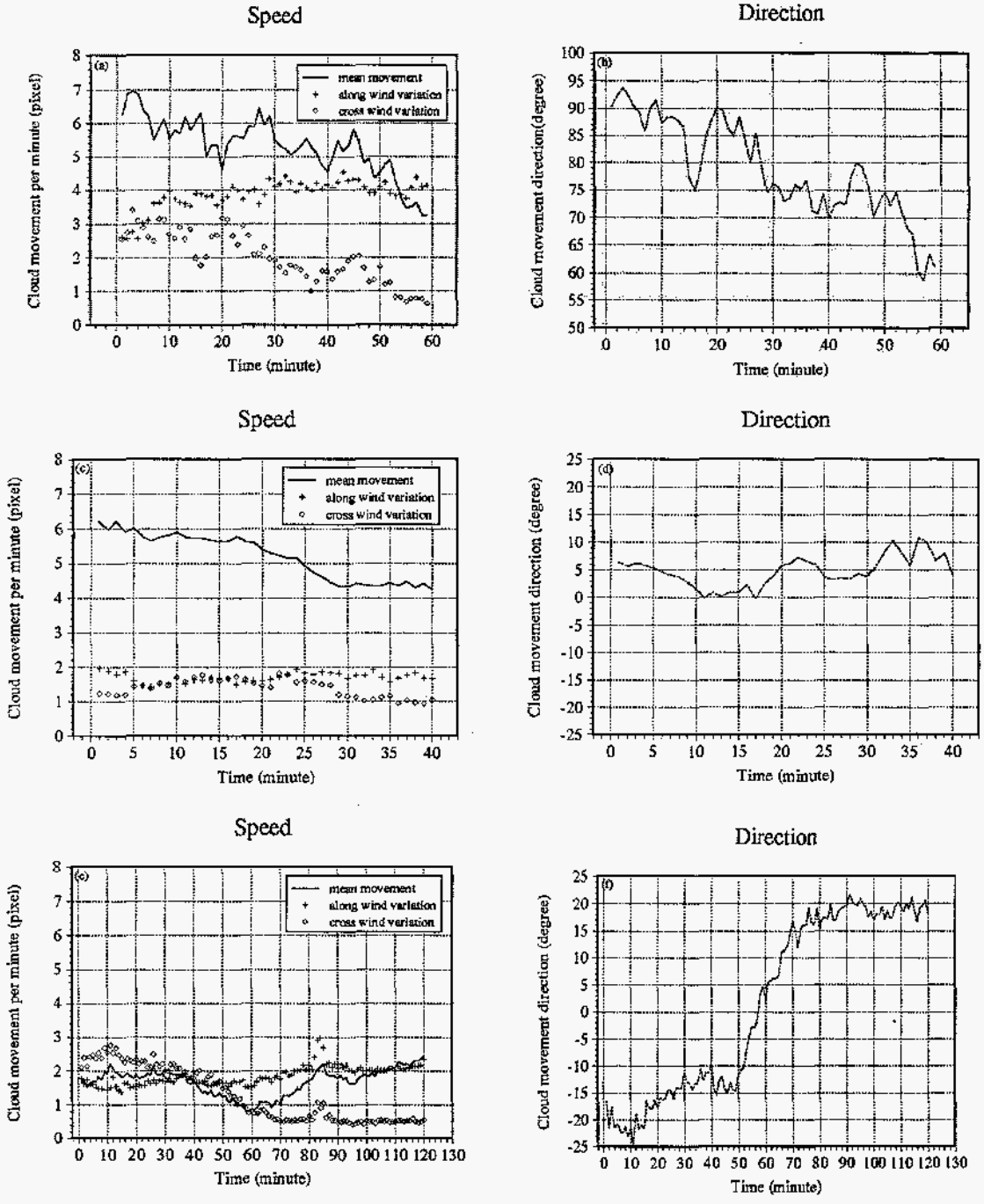


Figure 7. Mean and fluctuating cloud motion results for three different cloud types.

4.3 *Testing Taylor's hypothesis*

Taylor's hypothesis is often used in field experiments to save the expense and effort of data sampling over a wide area [20]. Taylor's hypothesis states that if the turbulence velocities are small compared to the mean velocity, then the temporal velocity correlation at a fixed location and the spatial velocity correlation along the direction of the mean velocity are the same [18]. For cloud fields, Taylor's hypothesis may also be applied to scalar correlations of cloud properties such as liquid water content, cloud backscattering and cloud

base height. Measurements of such properties are often restricted to zenith views of the atmosphere when full spatial information is desired.

To test how well narrow-field-of-view temporal observations can be used to infer the spatial cloud correlation scale, Taylor's hypothesis was tested on the WSI pixel brightness for a range of cases.

The mean wind vector needed to test Taylor's hypothesis is approximated by the mean cloud motion which is obtained by averaging the optical flow result (described in the previous section). The fluctuations in the cloud motion are estimated by taking the standard deviation of the optical flow velocities along and normal to the mean cloud motion direction. Figure 8 shows the mean and fluctuating cloud motion results for three different cases. For the cumulus (Cu), stratocumulus (Sc), and cirrus cases (Ci), respectively. The mean wind speed for the three cases is 2.8, 2 and 1.2 pixels/min.; and the wind direction is 30, 10 and 45 degrees respectively.

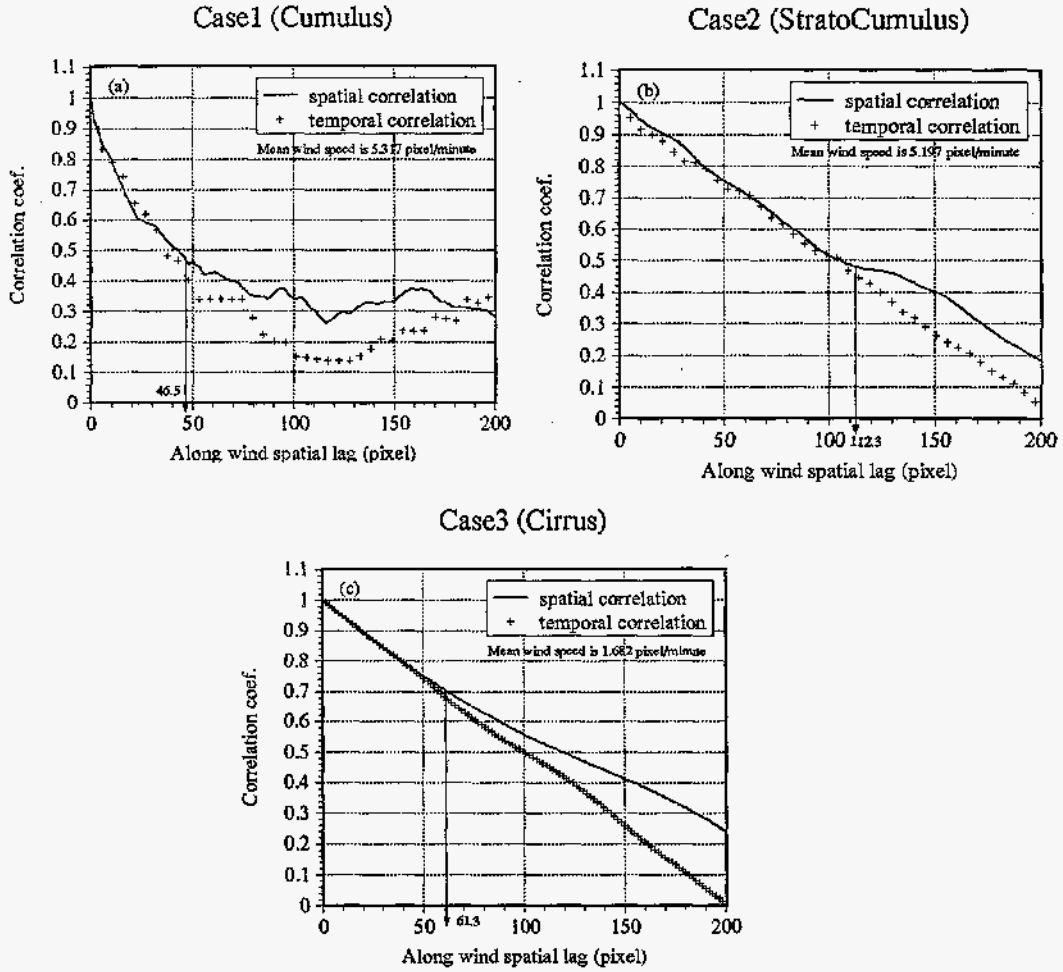


Figure 8. Temporal pixel brightness correlations at a single point for three different cloud cases -- Case 1: Cumulus; Case 2: Stratocumulus; Case 3: Cirrus.

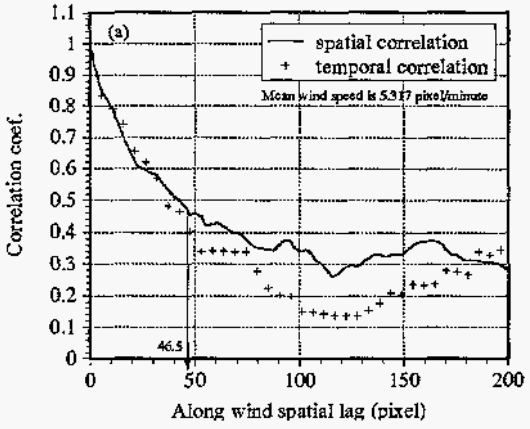
4.3.1 Results

Figure 9 compares the temporal pixel brightness correlations at one point -- *R(Udt)* -- with the spatial pixel brightness correlation along a line in the mean wind direction -- *R(dx)* -- for the different cloud cases. Taylor's hypothesis holds when $R(dx) \sim R(Udt)$, where R is the correlation function coefficient of the cloud brightness; dx is the spatial lag; dt is time lag; and U is the mean speed of cloud movement. For Case 1 (Figure 9a), the images used were acquired during a period of one hour. The mean cloud base height was about 2.5 km [16]. During this period of time, fair weather cumulus clouds were distributed throughout the 10-km field of view and the mean wind speed was about 2.22 m/s (*i.e.* a movement of 5.3 pixels from image to image) as determined from our optical flow analysis. The correlation results indicate that the spatial and temporal correlations are in good agreement for about 46.5 pixels (which for this case is equivalent to 1.162 km or 8.75 min.). Therefore, Taylor's hypothesis holds for this period of time and over this distance.

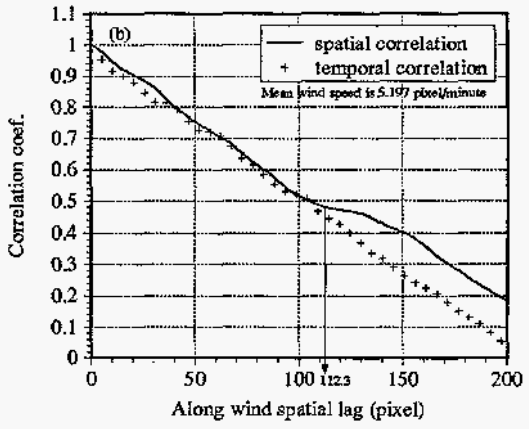
For Case 2 (Figure 9b), the images we used were acquired during a period of 40 minutes and the mean cloud base height was about 4.0 km. During this period of time, stratocumulus clouds were distributed over the 16-km field of view and the mean wind speed was about 3.5 m/s (i.e. 5.2 pixel from image to image). Note that while the wind speed was higher for Case 2 than for Case 1, the cloud movement in pixels was about the same due to the difference in cloud base height. That is to say, the angular movement in both cases is about the same, but because the clouds are higher in Case 2, the clouds are actually moving at a higher speed. The results for Case 2 indicate that Taylor's hypothesis holds over a distance of about 112.3 pixels (which is equivalent to 4.49 km or 21.6 min.).

For case 3 (Figure 9c), the images used for analysis were acquired during a period of 120 minutes. The mean cloud base height was about 10 km. For this case, cirrus clouds were distributed over the 40-km field of view and the mean wind speed was about 2.8 m/s (i.e. 1.68 pixel per minute). This is lower than the typical wind speed at this height, and this apparently low speed may not be real. The low speed could result from the optical flow algorithm making the texture of the cirrus clouds too smooth. This lack of texture in the WSI image makes the correlation difficult. Nevertheless, the results for Case 3 indicate that Taylor's hypothesis holds over about 61.3 pixels (or, equivalently, 6.13 km or 36.4 min.). If the actual wind speed is higher than our estimated speed, the coherent spatial scale would be larger and the coherent temporal scale would be longer.

Case1 (Cumulus)



Case2 (StratoCumulus)



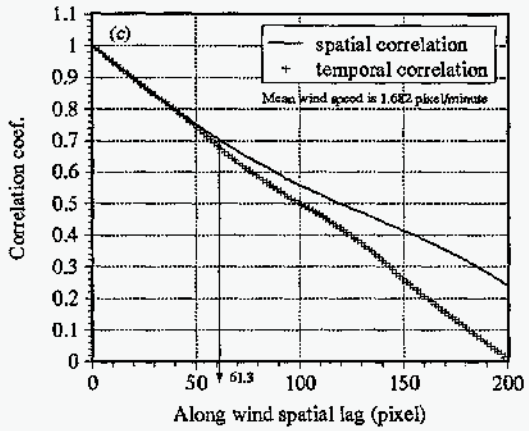


Figure 9. Temporal pixel brightness correlations for the different cloud cases; (a) Case 1 (Cumulus), (b) Case 2 (Stratocumulus); (c) Case 3 (Cirrus)

Table 1 gives a summary of the relationship between cloud spatial and temporal correlation scales. The ratio of the velocity fluctuations to the mean flow velocity is often used to indicate when Taylor's hypothesis holds. Willis et al. [21] suggests a value of 0.5 for this ratio from diffusion experiments. In this study, the ratios are 0.73 and 0.36 for the cumulus and stratocumulus cases, respectively. But for the cirrus case, Case 3, the ratio is over 1. However, due to the smooth texture of the cirrus clouds, the derived cloud motion is less than typical for clouds at that height, and so the actual ratio is probably lower than 1.

Table 1. Relationship between spatial and temporal correlation scales

Parameter	Case 1 Cumulus (Cu)	Case 2 Stratocumulus (Sc)	Case 3 Cirrus (Ci)
U *(pixel/min)*	5.3	5.2	1.7
θ (°)	79.0	4.7	1.4
σ_U *(pixel/min)*	3.8	1.7	1.8
σ_V *(pixel/min)*	2.0	1.4	1.3
σ_U/U	0.72	0.36	1.28
σ_V/U	0.38	0.30	0.89
Spatial Scale (pixel)	46	112	61
Temporal Scale (min)	8.8	21.6	36.4

4.4 2-D spatial correlation function

We have shown that 1-D temporal and spatial correlation functions along the mean cloud motion direction agree well for certain time and spatial scales for certain cloud fields (see Figure 9). We now compute the 2-D spatial auto-correlation functions for the cloud fields to show that the 2-D cloud morphology and isotropy determine whether we can infer the spatial scale from the temporal distribution; not just in the mean wind direction but in the full two-dimensions.

Figure 10 shows the autocorrelation functions (ACF) for cumulus (Cu), stratocumulus (Sc), and cirrus (Ci) cloud fields. The inherent modulated structure of the cumulus and stratocumulus cloud fields can be seen in their ACFs. The cirrus cloud field ACF shows an elongated shape along the mean wind direction. Inspection of the contour lines reveals that, as would be expected, the inner lines are more isotropic than the outer lines. If we look at the shape of the ACFs for each cloud type over the coherent length scale indicated in Table 1 (i.e., 46.5 pixels for Cu, 112.3 pixels for Sc, and 61.3 pixels for Ci), we see that the contour lines are relatively isotropic for the Cu and Sc cases but relatively non-isotropic for the Ci case. This might be expected from the elongation of the cirrus clouds along the mean wind direction.

4.5 Deriving cloud cover fraction from time series measurements

If Taylor's hypothesis holds and the 2-D distribution is sufficiently isotropic, 2-D spatial averages can be determined from temporal averages. To illustrate this for cloud fields, we estimate the cloud cover fraction from the WSI data with a narrow 15 pixel by 15 pixel (8.6 degree) field of view. The estimation is made by using varying window sizes for the average at each time step. At each time step, average cloud cover fractions are computed with all feasible window sizes and compared with the wide field of view (126 degree) cloud cover fraction. The window size giving a value closest the actual cloud cover fraction (as determined from the full image) is used. Figure 11 shows a comparison of the best moving-window average for images taken May 4 & 5, 1992. The labels, 1-11, signify the various cloud types and conditions that predominate in the time period indicated. They are as follows:

1. Stratus cloud overcast
2. Broken and decaying stratus
3. Decaying stratus and some cumulus
4. Individual cumulus
5. Clustering and vertically developing cumulus
6. Half of the image is clear with strong horizontal wind shear
7. Big clustering cumulus moving in
8. Almost clear with little cirrus
9. Cirrus moving in, mean wind changing from SW to S
10. Cirrus above cumulus
11. Stratus

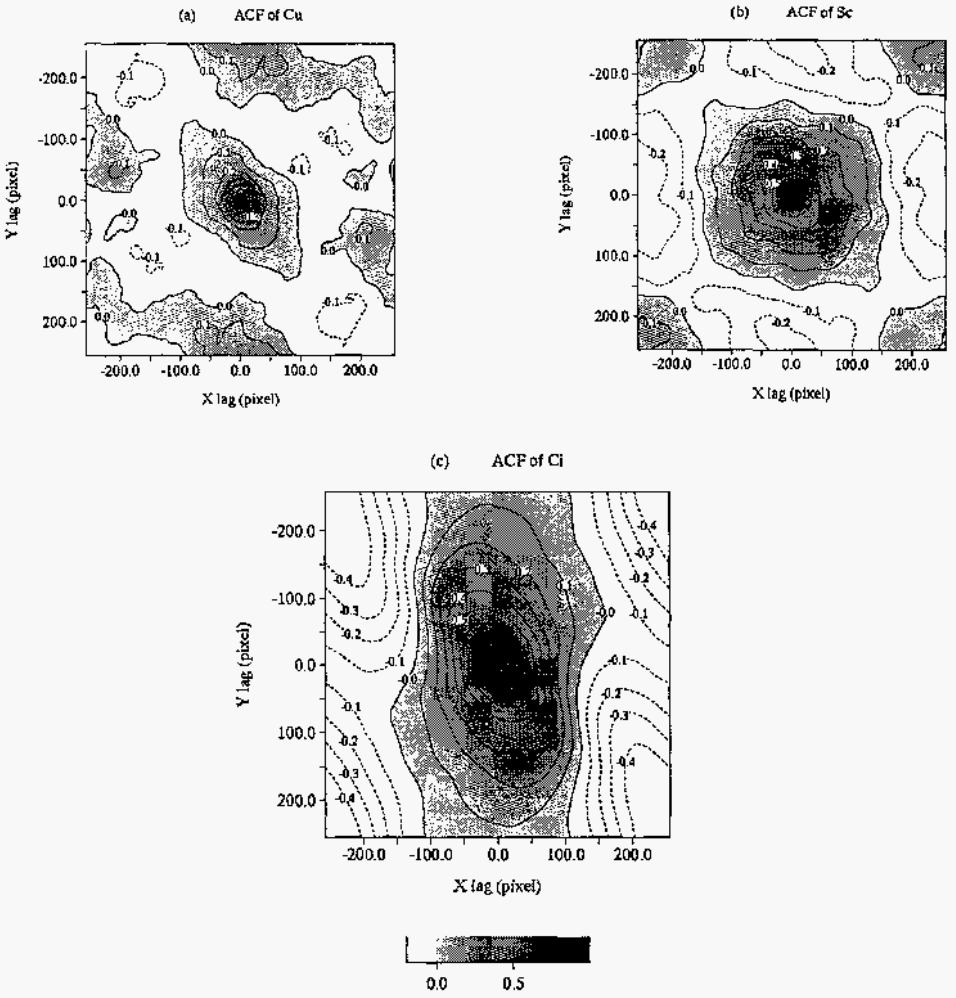


Figure 10. two-dimensional spatial autocorrelation functions for three different cloud classes.

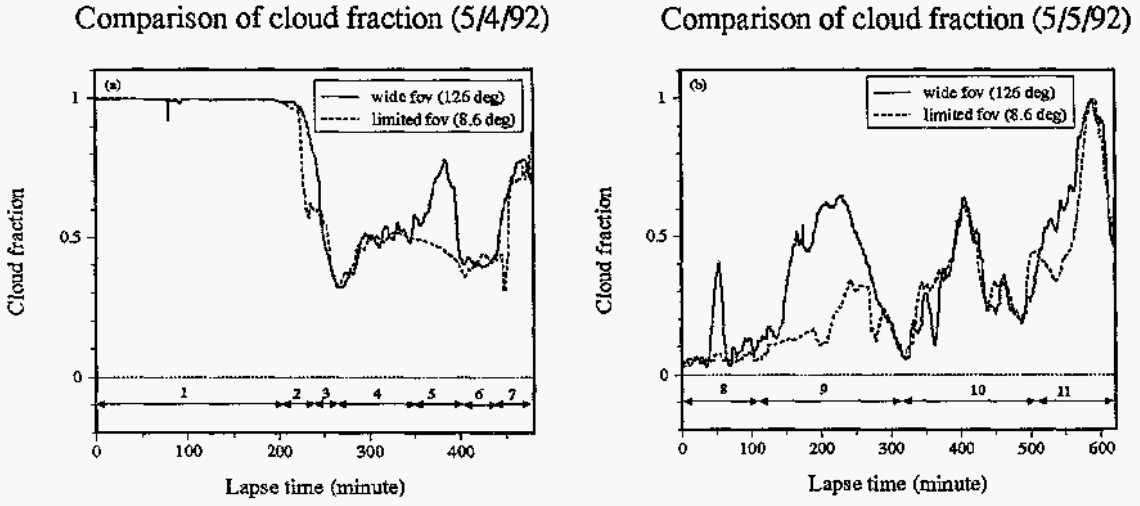


Figure 11. Comparison of limited field-of-view measurements of the cloud cover fraction with the actual cloud cover fraction (i.e., 126° field of view).

When Taylor's hypothesis holds and the cloud field is isotropic, the time series cloud cover approximation provides a good estimate of the real situation, *i.e.*, the actual cloud cover fraction spatially averaged over the 126-degree field of view of the WSI. We have analyzed different cases for various cloud types and conditions for May 4 & 5, 1992. We summarize in Table 2, the conditions for which Taylor's hypothesis applies to 2-D cloud fields.

Table 2. The Conditions for which Taylor's hypothesis holds for estimating cloud cover fraction.

Holds for these conditions	• Wind has uniform speed and direction over time • Statistically isotropic over horizontal region greater than 10 km for cumulus • Statistically isotropic over horizontal region greater than 15 km for stratus
Generally holds for these conditions	• Mostly overcast stratus • Mostly overcast cirrus • Individual cumulus • Advection of large clustered cumulus • When cirrus occur above cumulus • Isotropic broken stratus
Generally does not hold for these conditions	• Clustering cumulus with rapid vertical development • Aniostropic broken stratus • Mostly broken cirrus
Does not hold for these conditions	• Wind changing speed and direction over time • When clouds are developing rapidly • When clouds are dissipating rapidly • When the strong horizontal wind shear exists • When there is a transition from one cloud field type to another When the cloud field is anisotropic

In addition to the May 4 and 5 data, we analyzed 9 days of data to quantify the overall accuracy of the cloud cover fraction estimation using limited field-of-view data and to optimize some of the sampling parameters used in the method.

First, we determined the best moving-window size for sampling the spatial distribution. We found that the optimum size is dependent on the wind speed and cloud distribution. We chose that window size which produced the smallest absolute error in cloud cover fraction for each day. Table 3 shows the best window sizes and the associated error for each of the 9 days. As is evident from the table, the best window sizes change from day to day as the conditions and cloud types vary. However, a fixed window size of 100 pixels produces good results for all days as shown in Table 4; the error does not increase drastically from that associated with the optimal window size.

The overall error of the cloud cover estimations are shown in Figure 12. Figure 12a shows the estimations are scattered around the true values with a maximum error of about 0.4. Also, the estimates tend to under-estimate the true value of cloud cover fraction. Figure 12b shows the estimated cloud cover fraction versus the average absolute error. The overall error is about 0.15 with a maximum of less than 0.3. The error is greatest for

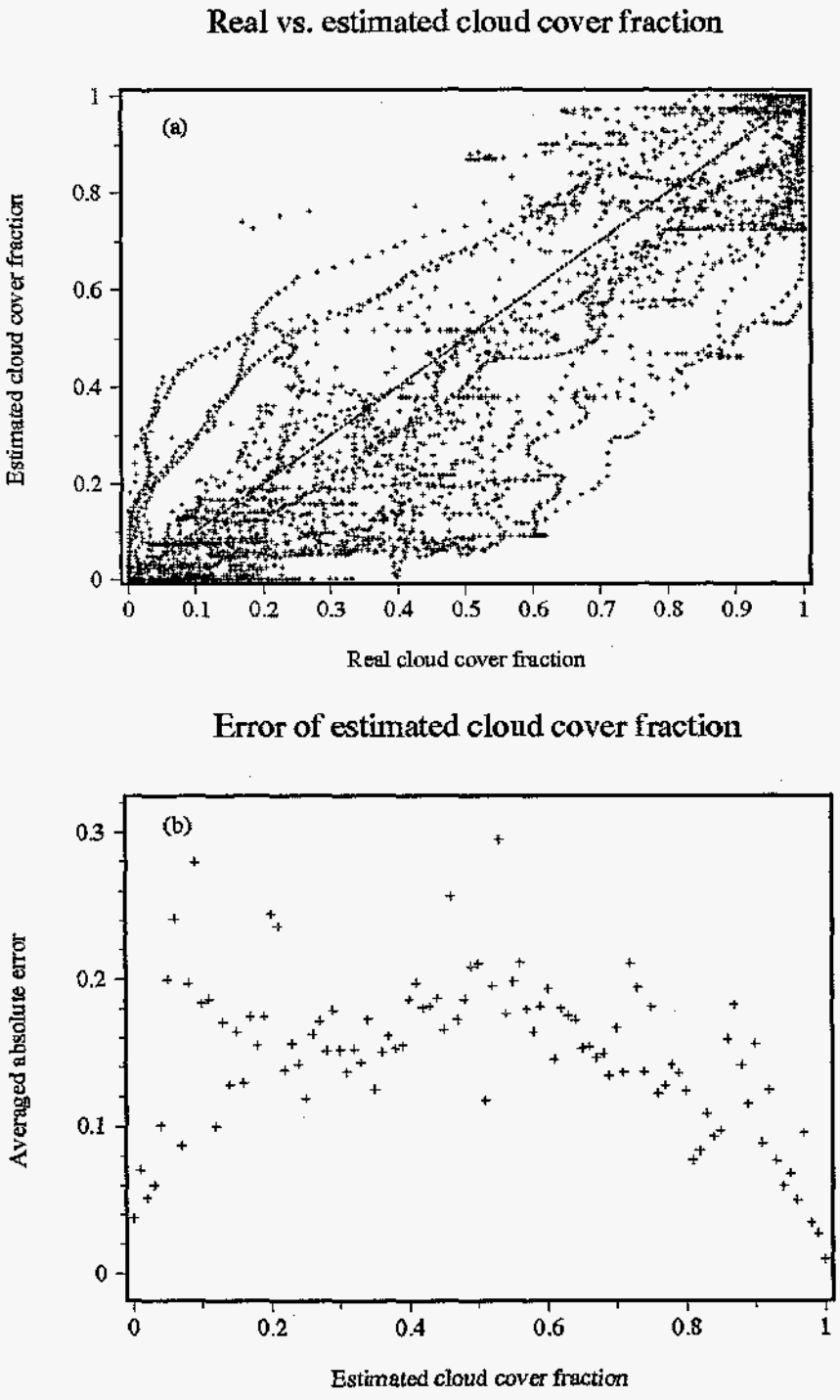


Figure 12. Error in the estimated cloud cover fraction determined by the moving-window average technique.

partially cloudy conditions and smallest for clear and totally overcast skies as might be expected.

Table 3. Error in cloud cover fraction using optimal window sizes.

Date	*Window Size (pixel)*	*Average Absolute Error*	*Maximum Absolute Error*
5/04	41	0.140	0.457
5/05	101	0.181	0.504
5/13	69	0.132	0.584
5/14	105	0.111	0.359
5/15	85	0.0.96	0.442
5/16	47	0.046	0.413
5/18	55	0.046	0.214
5/19	56	0.131	0.397
5/27	26	0.077	0.371

Table 4. Error in cloud cover fraction using fixed window sizes.

Date	*Window Size (pixel)*	*Average Absolute Error*	*Maximum Absolute Error*
5/04	100	0.144	0.573
5/05	100	0.182	0.506
5/13	100	0.137	0.500
5/14	100	0.111	0.375
5/15	100	0.0.97	0.373
5/16	100	0.051	0.304
5/18	100	0.056	0.378
5/19	100	0.140	0.408
5/27	100	0.153	0.413

4.6 Conclusions

It is often desirable to infer spatial properties of cloud fields from zenith, time series observations. This can be done using Taylor's hypothesis when it applies. In this study, we have estimated cloud cover fraction from limited-field-of-view observation using Taylor's hypothesis. We have identified some for broken cloud field conditions for which it holds and some for which it does not hold. These are summarized in Table 2.

In brief we have found that:

1. 1-D Spatial statistics of the cloud field brightness (i.e., cloud cover fraction) can be inferred from temporal statistics when Taylor's hypothesis holds.

2. 2-D spatial averages can be inferred from the time averages when Taylor's hypothesis holds; generally when the cloud field is more isotropic. Taylor's hypothesis holds for a variety of broken-cloud conditions--generally for cumulus and for mostly for overcast stratus or isotropic stratus cloud fields.

3. Reasonably accurate mean cloud velocities can be estimated by using the optical flow algorithm for Cu and Sc clouds. But this approach does not work well for cloud fields with smooth textures, such as, Ci.

4. The ratio of turbulence and the mean wind speed can not be always used as a condition to test for the applicability of Taylor's hypothesis; primarily because the cloud texture does not always reflect the turbulent field, especially for smooth texture cloud like Ci.

4.7 Acknowledgments

This research is sponsored by the Laboratory Directed Research and Development Program (LDRD), Sandia National Laboratories. The Marine Physical Laboratory (MPL) of the Scripps Institution of Oceanography, provided the whole-sky imagers and the processed cloud decision images used in this study. We thank Janet Shields of MPL who set up and operated the imagers and Bob Endlick of the U. S. Army Research Laboratory, White Sands Missile Range, who provided the sitting of the WSI's at White Sands, NM.

5. Automatic Cloud Classification

5.1 Motivation

The 3D cloud reconstruction is dependent on the type of clouds in the cloud scene. For example, convective clouds would utilize different models, methods, and assumptions than would stratiform clouds. Also, it is desired that the reconstruction be as autonomous as possible, so that arbitrary cloud scenes can be reconstructed without excessive preliminary user intervention. Therefore, automatically determining and classifying the types of clouds present in cloud scenes is desired. Furthermore, the dependence of the cloud reconstruction on the local cloud field, the desired resolution of the reconstruction, and the cost and low temporal resolution of satellite imagery make it very attractive to be able to perform the cloud classification from the ground based WSI data

5.2 Problem Statement

Much work has previously been done on cloud classification using multispectral satellite data [22-28]. Most of this work has utilized textural features of the large scale cloud systems as inputs to a wide variety of classification methods (*e.g.* neural networks, clustering, and thresholding algorithms). The WSI, on the other hand has a much more local view of the cloud scene (10s of kilometers as opposed to satellite images that view 100s of kilometers). Because of this limited view, many clouds extend beyond the edges of the image and are not captured in their entirety. Another difference between the classification using satellite data and that using WSI data is the projection of the image. The satellite images are essentially flat and all the clouds are seen from virtually the same perspective. The WSI, on the other hand, being a ground-based observation instrument, produces a distorted image. The image is essentially equi-angular, that is, the azimuth angle is preserved and the zenith angle is proportional to the distance from the center of the image. This yields an image where both the cloud bases and sides are visible and the are projected towards the edges of the image while the cloud sides are projected towards the center.

The classification technique we have developed, based on the ground-based WSI data, utilizes binary decision trees to distinguish between the various types of clouds. Binary decision trees were chosen for several reasons. First, there is a multitude of features that describe the cloud scene and that could be used for the classification. It is impossible to know *a priori* which are the most important and useful features to use. Binary decision trees offer the advantage of allowing a comparison of the effect of different features on the classification result thereby making it possible to choose the most relevant features. Second, binary decision tree algorithms can give an estimate of the misclassification probability. This allows the user to perform further processing or visual inspection on only the points with a high probability of being misclassified instead of on the entire data set. This reduces processing time and user interaction.

5.3 Classification Method

Binary decision tree algorithms use a set of training data, for which the cloud class is known, and generate a hierarchy of thresholds based on a series of measured features. For each point in the training sample, n different features are measured. These features are numbers describing the cloud scene at that point. Possible features include texture meas-

ures and position within the image. These measurements are placed in an n-dimensional space and successive splits (which are equivalent to thresholds on individual feature values) are used to separate the points into different regions or nodes. At every split, each region is individually analyzed to find the optimal split, *i.e.* the split which most reduces some measure of the node impurity.

The ideal end result is a set of pure leaf nodes (terminal nodes that contain points belonging to only one class). This is an ideal tree structure because, in general, it is obtained by too much splitting and the trees are much larger than the data warrant. This results in a tree that is susceptible to noisy data and one that has a higher true misclassification rate than the optimally-sized tree. On the other hand, if the training data is not split far enough, important information in the data may be ignored. The resulting tree will be too small and will also have a higher true misclassification rate than the right sized tree.

Since 'growing' the binary decision tree out until all leaf nodes are pure usually results in a tree that does not represent the data well, there are two methods for 'pruning' or reducing the size of the tree. For the first method, the tree is grown out until every node is pure but then 'pruned' back to allow impure leaf nodes. The method used in this work is cross-validation. Basically, subsets of the full training data set are used to grow several trees. These trees are compared to find that portion of the main body that is constant from one tree to the next. The leaf nodes not belonging to this constant portion are then removed from the tree. This method also gives misclassification rates closer to the actual rates for the training data.

The features utilized in this work can be grouped into three categories -- texture, position, and pixel brightness. As noted before, while cloud shape, boundary, size, *etc.* could also be used, and would probably be relatively important, the WSI data precludes these from being used due to its limited field of view. The first texture measure is the standard deviation of the image brightness (which is the normalized radiance in a narrow band centered at either 450 nm or 650 nm) computed over a small neighborhood about each point. The remaining texture measures are computed using the Laws kernels Laws [29]. These 5×5 matrix kernels are used to quantify the response of the image to 25 different types and orientations of texture. Each of the 25 matrix kernels is derived by taking the outer product of two out of five vectors. Each vector is designed to correspond to a different basic texture, namely, Level (L5), Edge (E5), Spot(S5), Wave(W5), and Ripple(R5). The matrix kernels, then, correspond to different combinations of these textures (*e.g.* L5L5 or L5S5) and the orientation information comes from taking the transpose of each (*e.g.* L5S5 and S5L5). The feature derived from the Laws kernels is actually a measure of the texture energy in a small neighborhood about each point. The image is convolved with each of the 25 kernels and the absolute value of the result is averaged over the neighborhood -- the result is a measure of the texture energy at each point.

The two position features arise from the nature of the WSI projection. The first is the pixel distance from the center of the image which corresponds to the zenith angle. Since the image is nearly equi-angular, the spatial resolution of each pixel is degraded with distance from the center of the image (*i.e.* since the field of view of each pixel is a constant solid angle, the horizontal extent viewed by each pixel increases with the zenith angle) and, therefore, the texture will also be degraded with increasing distance from the center. The second position feature is the pixel distance from the location of the sun. The observed radiance of the cloud field is directly related to the scattering angle from the sun to the observation point.

The third feature, the normalized pixel brightness, captures some indication of the optical thickness of the cloud. Since the image is a view from the ground, the more optically thick clouds will appear darker. Also, this feature can indicate whether a particular point is viewing the base or side of a cloud, since the sides (especially of fair weather cumulus) are much brighter than the base.

The texture features mentioned above only capture information about the small-scale texture, that is, the intra-cloud texture. We anticipated that the larger-scale texture would also be important, especially the inter-cloud texture. Therefore, we employed a pyramid-type scheme Burt [30] to capture this larger-scale information. The basic idea of the pyramid scheme is to reduce the resolution and sample density (and therefore the physical size of the image) so that the neighborhood used for the texture features (which is kept a constant size) is effectively made larger compared to the image size. For this work, four pyramid levels were used, the first level (level 0) being the full image and each successive level being half as big as the previous.

5.4 First Iteration

The first step was to use four cloud classes (altocumulus, cirrus, cumulus, and stratus), the clear sky class, and the features mentioned above. Also, since the classification is being performed at every pixel, the result can be quite noisy. For the first step, we reduced the noise through simple k-nearest neighbor filtering (the value of the pixel at the center of the filter window is replaced with the mode of the values within the window).

5.4.1 First Iteration Results

The training data are outlined in Table 5 showing the number of images and the total number of points labeled as each class. The features mentioned above (which total 107 when the four pyramid levels are counted) were reduced to the nine most important features. These nine features (in order of decreasing importance) were: the Laws feature L5L5 for the second pyramid level; the distance from the sun location; the distance from the center of the image; the standard deviation for levels 0 and 1; the pixel brightness; and Laws features L5E5, L5R5, and E5L5, all for level 3. The importance of these particular features demonstrate that both large-scale and small-scale features are important for classifying cloud types. The level 3 Laws features L5E5, L5R5, and E5L5 are most likely *inter*-cloud features while the standard deviation for levels 0 and 1 and the pixel brightness are most likely *intra*-cloud features. Furthermore, the L5E5 Laws feature and its transpose E5L5 are both important. This demonstrates that this particular texture, which is an edge in one direction and level in the other, is important at more than one orientation. The two distances relevant to the WSI image geometry were also very important, as would be expected, since the cloud texture varies greatly from the center to the edges of the image and the cloud appearance is such a strong function of the scattering angle.

Table 5. The composition and size of the training data and test data sets.

Cloud Class	Number of Training Images	Number of Training Points	Number of Test Images	Number of Test Points
altocumulus	3	80197	1	32264
cirrus	5	153760	1	52347
cumulus	5	114955	2	75235
stratus	5	187394	2	152595
clear sky	16	469152	4	125606

Once the important features are identified from the training data, these features and the hand-classified data are passed to the binary decision tree algorithm which determines the appropriate thresholds to apply to each feature to effect the optimal separation. The best way to test this separation is to then apply the same thresholds to a set of test data; data not used in the training step. Table 5 also outlines the number of images and the total number of points labeled as each class for the test data. These data are also hand-classified and used as ground truth, but the algorithm is not given this information. After the predetermined thresholds are applied to the test data and the k-nearest neighbor filtering is performed, the results are compared to ground truth. Table 6 is the confusion matrix generated from this comparison. The labels along the left are the class types as determined by the ground truth. The labels along the top are the class types as determined by the binary decision tree algorithm and filtering. The numbers are the probabilities of a point that is class *i* being classified as class *j*. For a perfect result, the confusion matrix would be diagonal. Reading across the top row, for example, the confusion matrix tells you that 43% of the altocumulus cloud pixels were correctly classified as altocumulus, 19% of the altocumulus pixels were misclassified as cirrus, 12% of the altocumulus cloud pixels were misclassified as cumulus, and so on. From these numbers, the overall misclassification rate can be determined. For the current set of test data, the overall misclassification rate was 39%.

Table 6. Confusion matrix for the test data after filtering with the k-nearest neighbor filter.

True Class	altocumulus	cirrus	cumulus	stratus	clear sky
altocumulus	0.43	0.19	0.12	0.10	0.17
cirrus	0.07	0.35	0.21	0.24	0.12
cumulus	0.08	0.14	0.54	0.14	0.11
stratus	0.04	0.12	0.12	0.46	0.26
clear sky	0.01	0	0.01	0	0.98
Overall misclassification rate = 0.39					

sponding diagonal element of the confusion matrix. This product allows a less represented but more accurately known class to be dominant over a more represented but less accurately known class. Then, beginning with the least represented class and working up, the co-existence of each class with the dominant class is checked based on the rules. If the class is not allowed, all pixels with that class designation are changed to the next most likely class. The new class is also determined from the confusion matrix. By normalizing each column of the matrix to 100 (rather than each row), the probability that a class labeled j is really i is obtained. After one class change is made, the dominant class is again determined with the change in classes being accounted for. For example, if the pixels initially labeled cumulus were changed to stratus, during the next dominant class determination, the probability that stratus clouds were misclassified as cumulus would be used for these stratus pixels and not the probability that stratus clouds were classified as stratus. When no more changes are made, the cloud neighbor rules are invoked. Each cloud group's neighbors are found and it is determined, according to the neighbor rules, whether these two clouds can share a border in the image. If they cannot, then one of the cloud group's pixels are changed to the type of cloud most common on the border.

5.5.1 Second Iteration Results

Using the eight classes outlined above, the seven most important features were extracted from the original set of 107 were: the standard deviation at levels 1, 2, and 3; the pixel brightness; the scattering angle; and Laws features L5L5 and W5W5, both at level 3. Preliminary results with test data indicate that this classification is not as good as the first iteration result. Most likely this is due to trying to spit up the data too much by using 8 classes. The classes are not different enough to warrant all 8 classes, and so the resulting classification is very data specific -- that is, it works well on the training data but not on the testing data.

The rule-based filtering technique was tested on the results of the first iteration. For this test, the clear sky case was ignored. The co-existence rules for the four cloud classes (altocumulus, cirrus, cumulus, and stratus) are as follows: 1) cirrus and stratus can exist with an altocumulus as dominant, 2) all four classes can exist with cirrus as dominant, 3) only cirrus can exist with cumulus as dominant, and 4) altocumulus and cirrus can exist with stratus as dominant. These rules basically reflect the idea that cumulus clouds cannot exist with stratiform clouds and that cirrus clouds can exist with any cloud type. The probabilities used in the filtering are given in Table 7, where the values have been renormalized to exclude the clear sky case.

Table 7. The probabilities that a pixel originally labeled class i is actually class j.

	ORIGINAL CLASS LABEL			
ACTUAL CLASS	altocumulus	cirrus	cumulus	stratus
altocumulus	0.47	0.12	0.05	0.03
cirrus	0.12	0.34	0.15	0.13
cumulus	0.20	0.20	0.55	0.11
stratus	0.21	0.34	0.25	0.73

5.4.2 Conclusions and extensions for the next iteration

While the thresholds determined from the training data accurately described the training data itself (the misclassification rate for the training data was only 0.06), they did not describe the test data as well (misclassification rate = 0.39). This can be attributed to several possibilities. First, the training data may not be representative of the WSI data in general. This shortcoming of the training data can be overcome by including more samples of each cloud class. This would supply the binary decision tree algorithms with information about a larger range of clouds in each cloud class at the expense of longer computational times. With the immense variability of cloud structure, shape, *etc.*, this is almost certainly a contributing factor to the high misclassification rate. It is also one of the hardest to overcome due to computational issues.

Second, there are several common errors in the classification by the binary decision trees that can be recognized by inspection of the classification results. The two most obvious errors are in determining the correct cloud class near the edges of the image and near the occultor. Near the edges of the image, the spatial resolution is degraded and much of the detail in the texture is lost - all cloud classes tend to look more alike than they do closer to the center of the image. The most promising way to deal with this is to place less importance on the classification results near the edge and, instead, infer the classification from the more accurate center results. Also, the measured radiance is the highest near the occultor and this tends to reduce the dynamic range of the measurement, thereby de-emphasizing the texture in that region. This issue is better handled with data from the latest WSI systems which contain a 16-bit CCD chip and provide a much larger dynamic range on the measurements.

One change that would most likely reduce the high misclassification rate would be to utilize a different filtering technique. The k-nearest neighbor filter used at this stage does not take into account any physical aspects of the problem. While using the k-nearest neighbor filter does reduce the misclassification rate (for the test data described here, the misclassification rate was 0.45 prior to the filtering step), a more physically motivated filtering technique would probably produce better results; specifically, a filter based on a set of co-existence rules for the various cloud classes.

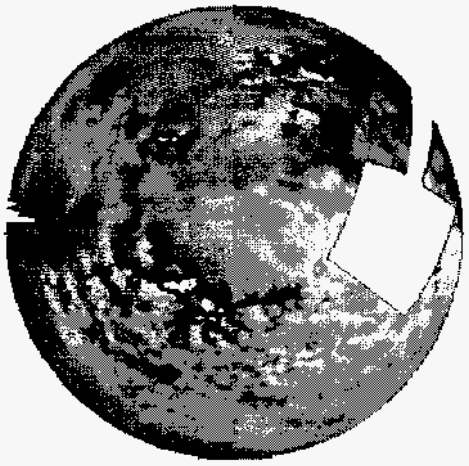
5.5 Second Iteration

Based on the results presented above, the cloud classes were expanded in an attempt to more accurately address the variability of clouds. The original four classes were expanded to eight: altocumulus (stratiform clouds with small, individual, modulated components), cirrus I (spreading cirrus), cirrus II (modulated cirrus), cumulus, stratocumulus I (stratiform clouds with large modulated components), stratocumulus II (stratiform clouds with small modulated components but not broken), stratocumulus III (stratiform clouds with varying sized modulation), and stratus (smooth and spreading stratiform clouds).

The second change was the filtering technique where we adopted a two-step approach. The classification tree output was still filtered with the k-nearest neighbor filter to remove much of the noise and a second, physically motivated step was introduced. The second filtering step was a rule-based approach based on the definition of a dominant cloud class, a set of cloud co-existence rules, a set of cloud neighbor rules, and the confusion matrix generated from the test data results. Each cloud class in the raw classified image (*i.e.* the direct output of the binary decision tree algorithm) was assigned a dominance factor equal to the product of the number of pixels labeled as that class and the corre-

The cloud neighbor rules were restricted to cirrus clouds. We assume that there are no holes in a cloud layer and since cirrus clouds are the highest clouds, if cirrus are visible they must have clear sky bordering them. Any cirrus cloud groups that do not border clear sky are changed to the cloud type that is most common along its border.

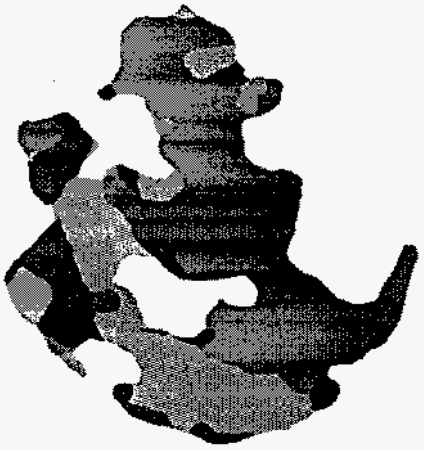
Figure 13 shows the results of this rule-based filtering. Part a shows the original WSI cloud image. Note that the cloud types are predominantly stratus and altocumulus. Part b shows the output of the classification tree. Since each pixel is classified independently, this images quite noisy. Part c shows the result of the k-nearest neighbor filtering step. Note that the general cloud class distinction is maintained, but the image is much cleaner. Part d shows the result of the rule-based filtering. Note that the cumulus clouds have been changed to stratus, since 1) stratus was the dominant cloud class, and 2) according to Table 7, stratus clouds were most often mislabeled as cumulus. Also note, that the cirrus clouds that were totally surrounded by stratus have been relabeled as stratus since we assumed no holes in a cloud layer and therefore these cirrus clouds would not be visible through the stratus layer.



13a. Whole-sky image

13b. Direct output of classification tree

13c. Result of k-nearest neighbor filtering



13d. Result of rule-based filtering

Figure 13. Cloud classification images showing each step in the classification process for rule based filtering. For images b, c, and d, the four grey levels denote cloud class; white = cumulus, light grey = altocumulus, medium grey = stratus, dark grey = cirru

This two-step filtering process, which accounts for the physical aspects of the classified cloud scene had a strong impact on the classification results for this image. The percentage of stratus cloud pixels correctly labeled as stratus went from 50% with no filtering to 71% with just the k-nearest neighbor filtering to 87% with the rule-based filtering. Also, the percentage of correctly labeled altocumulus cloud pixels went from 44% to 58% with the k-nearest neighbor filtering and remained the same with the rule-based filtering.

5.6 Conclusions

We have begun developing a cloud classification method for WSI data. This is a challenging problem because of the limited field of view, the ground observation perspective, and the varying spatial resolution of the measurements. The first classification step is a statistical-based decision step based largely on the texture observed in the images. Texture was used as the basis because cloud size, boundary shape, etc., are not easily extracted from the data. Based on a 4-cloud class and the preliminary 8-cloud class results, these texture features are not sufficient for an adequate classification. The misclassification rates were high and it was obvious where the classification was failing. Therefore, a second classification step was added -- a rule-based decision step. This step used physically motivated rules to reduce the misclassification rate. Cloud co-existence and one cloud neighbor rule (namely, cirrus must border clear sky) were included in the current work, but this rule base can and should be expanded with a very probably increase in effectiveness. Further cloud neighbor rules and cloud size restrictions are two examples of rules that can be added. The results from the rule-based decision step were very promising. The decrease in misclassification rate was significant and it is computationally very fast. Also, the idea of physically assessing the viability of a classified cloud scene and then altering the scene to make it more viable based on the most likely errors is very attractive.

6. Data Fusion for 3D Cloud Field Reconstruction

6.1 Specific needs and Applications for Numerical, 3-D, Cloud Renderings

It is significant to note that modeling 3-D cloud effects has been done only for cloud fields composed of clouds having regular geometric shapes; no numerical cloud shapes representative of real clouds were available prior to our LDRD-sponsored work. In a broader context, the numerical, 3-D renderings resulting from our work will provide detailed realistic cloud fields against which reduced or simplified geometrical representations of clouds can be compared. We believe that this will serve as an invaluable guide in parameterizing the geometrical effects of clouds for radiation transfer and clouds and radiation interactions.

Numerical, 3-D cloud renderings can also play an important role in the study of cloud life cycle by helping to establish, for example, the volume fraction of available water vapor that is converted to cloud condensates as the cloud evolves and ultimately dissipates. For convective clouds, a major focus of the ARM models [3,9] is understanding why clouds form where they do and the causes of variable growth rates of individual clouds. In addition, since convective clouds are initiated by radiative heating of the surface, and since these clouds strongly modulate their local radiation fields, 3-D cloud renderings would be a valuable tool in diagnosing the surface-cloud-radiation interactions that drive cloud evolution. This is especially important because of the generally large aspect ratio of convective clouds (2-5) which, as indicated above, has a strong influence on the effective cloud cover fraction.

6.2 Producing Numerical, 3-D, Cloud Renderings, Scope of the Problem

One of our goals in the LDRD-sponsored work is to develop methods for generating 3-D numerical cloud renderings based on field observations; i.e. numerical reconstruction of real cloud fields. Given a spot on the earth's surface, clouds that are over head have the greatest effect on modulating the local radiation field. As the distance between the surface point and the cloud increases, the radiative influence exerted by the cloud decreases; inversely proportional to the distance squared. In previous studies, clouds 15 km distant from the surface point were considered to have little effect on the locally measured radiation field. Thus, we define a vertical atmospheric column with a base radius 15 km as an empirically defined, radiatively significant volume over which to reconstruct 3-D cloud renderings. The atmospheric column with 200 x 200-km base is also important to study because of it is roughly the size of one grid cell of a GCM. This volume may be characterized with less detail than the 30-km column by extrapolation of the 30-km data using the methods developed above in Section 4, for conditions in which Taylor's Hypothesis holds.

The required spatial resolution needed for cloud evolution and radiation studies are comparable at about 100-500 m. Radiation studies require temporal resolution of one rendering per 15 minutes [9], which is adequate for cloud life cycle studies except for conditions of strong convection in which case renderings at one per 5-10 minutes are needed [14]. If the atmospheric column is divided into individual volume elements 100-m on a side in the altitude range of from 0.5 to 12-km, the major domain for clouds, 8×10^6 vol-

ume elements are produced for the 30-km diameter study area and 4.6 x 10^9 for the 200 x 200 -km study area (or 7.2 x10^7 for 500-m volume elements). We are primarily interested in the boundaries of the clouds in this study, so the total number of volume elements (or cloud surface points) may be reduced significantly, depending on the mean cloud size (surface to volume ratio) and their total number. A reduction by 50 to 500 is not unreasonable for very cloudy conditions. And of course, the number of surface cloud volume elements goes to zero for clear conditions, representing an infinite reduction factor. Thus, it is reasonable to expect that for the 30-km diameter study area, we need to resolve ~ 10^5 100-m volume elements for the cloud surface renderings and 10^7 500-m volume elements for the 200 x 200-km study area.

As discussed more in detail below, the primary image input for the fusion is whole-sky imager (WSI) and satellite images. Considering a typical example for a cloud cover fraction of 0.5 at 7-km altitude, the WSI would produce 1.3 x 10 ^{5} cloud surface pixels with linear resolution ranging from 40 m at the zenith to 2 km at a zenith viewing angle of 82° for clouds 50 km distant from the WSI. Of these pixels, about 1/2 are at resolutions higher than required for specifying the 100-m cloud surface elements, so 6 x 10^4 pixels are available to define over 10^5 cloud surface elements.

The problem of numerical reconstruction of clouds form field images is highly underconstrained--many more volume elements or surface elements exist in the real scene than are recorded in the sum total images. Moreover, the WSI will image primarily cloud bottoms at small zenith viewing angles and mostly sides at large zenith viewing angles, and it cannot see occluded surfaces., i.e., no tops and no sides opposite the viewing side and no surfaces occluded by cloud multiple layers or clouds in the foreground.

This problem is aided enormously by satellite images, but only the tops are imaged. No direct information is available on the bottoms and little or none for sides; the problem remains data deficient and cannot be solved uniquely by using images alone. Thus, some additional information is needed to lead us to the solution(s) that are most "reasonable" or to aid in rejecting the many potential solutions that would be consistent with the limited information available in the observed images.

6.3 Model-Based Approach for Fusion Algorithm

To address the problem of insufficient input data, we developed model-based fusion. Using this approach we exploit the fact that the objects being imaged are clouds, and cloud physics models along with knowledge of typical cloud types, shapes and their optical properties can be used to supplement image and meteorological data to the extent that a representative, numerical, 3-D rendering of the CART cloud field can be computed. This was our primary motivation for developing automated cloud typing algorithms discussed above. Specifically, cloud top heights are routinely obtained from satellite data and methods are now being developed to infer base heights as well [15, 16]. Cloud base heights can be determined from pairs of WSI images [18] such as those that will be available from the Southern Great Plains CART site in the near future and tops may be inferred in a manner similar to extracting base heights from satellite data. These primary image input data can be further supplemented with radar measurements of cloud base and top heights and heights multiple cloud layers if they exist, and ceilometer cloud base height measurements at specific CART locations. And, identifying cloud types in the cloud field permits us to infer the boundaries of clouds that are not imaged.

(In principle, the whole process could then be iterated to self-consistency using a radiation transport code to relate the observed image radiance with the presence, height, location and geometrical thickness of the clouds. The essential idea is to obtain initial estimates of the 3-D cloud field using satellite data alone and WSI data alone, then combine the two into a self-consistent 3-D cloud field using common radiation transport and cloud properties models in an optimization code. Though we did not attempt in our LDRD work, such a "grand" data fusion it is feasible and would be a logical and useful extenuation of our LDRD work).

6.5 Ancillary Problems

6.5.1 Cloud detection.

Determining which pixel elements are cloudy and which are clear is a conceptually simple but operationally difficult problem. For ground-based whole-sky imagers, Koehler and coworkers [17] have developed a variable threshold technique for determining the presence of clouds from the ratio of red-filtered and blue filtered WSI images. The clear sky red-blue ratio is calculated for various sun angles and compared with the observed red-blue ratio. Each pixel is then classified as thick cloud, thin cloud or indeterminate. The input needed for the clear-sky reference file is obtained from WSI data accumulated at the site for clear days or mostly clear portions of the sky. In our own work on WSI images, we have identified clouds through optical flow methods from a time series of WSI images [18]. The essence of the discrimination is that the clear-sky pixels show little motion on a minute-by-minute basis, while the opposite is true for clouds.

The satellite cloud discrimination problem is well-studied and a number of successful techniques have been reported [19-21]. Among the more reliable ones are those that use detection thresholds based on bispectral detection, spatial coherency criteria and those that use radiative transport codes to compute the expected clear-sky radiance as observed by the satellite sensor. Sorting pixels according to the radiance expected for clear sky is similar to the method used by Koehler for WSI images. Nevertheless, we anticipate that some refinements in the current satellite cloud retrieval schemes may expedite their fusion with WSI data. So, we have initiated a collaboration with Patrick Minnis at NASA Langley Research Center to optimize the utility of retrieved satellite cloud data for the 3-D fusion process. Minnis is familiar with our WSI work and has extensive experience in retrieving cloud properties from satellite images.

Zenith-viewing radar planned for the Central Great Plains CART site and the nearby scanning NEXRAD radars of the National Weather service will also aid in detecting clouds over the CART site and will be used as input in to the optimization/rendering process.

6.5.2 Cloud geometrical thickness

We have developed an image segmenting technique for cumulus cloud fields that permits the automatic detection of cloud bottoms. For a cloud mass that is not directly over head, those pixels that do not belong to the bottom belong to the sides. Thus, if the approximate cloud base height is known, the cloud geometrical thickness can be determined from the WSI image by simple trigonometry.

6.5.3 Computer requirements.

We have used Sun Sparc 20 machines to perform most of the calculations in our LDRD work. Image fusion, cross correlation, and producing a synthetic cloud field usually took about 10-100 minutes of CPU time. Building binary decision trees for automatic cloud classification required 10-100 hours of CPU time. Determining optical flow and cloud base heights from widely spaced WSI's required 1-10 hours of CPU time. To develop a reconstruction in which the cloud base height was determined from many sequential images from two or more WSI's, the CPU time would be thousands of CPU hours. One of us, Carl Diegert (SNL/NM), successfully converted the WSI cloud base height algorithm to parallel code and ran it on Sandia's Intel Paragon computer. The run time to determine a cloud base height was about 10 CPU seconds. We anticipated needing this greatly accelerated speed when we advanced to an optimized iterative cloud field reconstruction involving a radiation transport code in a potential follow-on study. Since there are many picture elements (~10^6) in each image and many images (~10), many calculations per element (~10^4) and we estimated a moderate number of iterations (10-100) to achieve self-consistency, the total number of mathematical operations would be substantial (~10^{13}). We anticipate processing times to be 10 min to 5 hr, depending on how much of the code could be made parallel.

6.6 Input data

At the present time, no single data stream can provide enough information to reconstruct the three-dimensional cloud field as mentioned above. Therefore, several data streams must be fused in order to get a complement of information sufficiently dense to allow the reconstruction. Of course, even using all the data streams mentioned here, we would not have enough information to reconstruct the cloud field without assumptions. The problem is too under-constrained as mentioned in the introduction. The ultimate goal is the methodology for reconstructing convective cloud fields, and so we define the cloud field structure and shape with the following parameters appropriate for convective clouds: cloud base height, cloud top height, cloud base shape, and vertical cloud boundary profile shape. Some of these parameters can be measured either directly or indirectly and some can be calculated from models. In the following sections, we discuss the data streams and models used to measure and estimate the various cloud geometry parameters. Table 8 lists details of the data streams we used in this work. The results we obtained using these data streams are 1) an initial reconstruction of a convective cloud, and 2) the reconstruction of a stratiform cloud case. These two cases will be presented and discussed below.

Table 8. Details of the data streams utilized in this study.

Whole Sky Imager (WSI)		
Description: Ground-based, spatially resolved measurements of entire sky dome		
Bands:	2 visible	0.45 μm, bandpass 0.07 μm 0.65 μm, bandpass 0.07 μm
Data collection frequency:	6 Images/hr	

Table 8. Continued.

Ceilometer	
Description:	Ground-based point measurement
Location:	Within several hundred meters of the WSI
Data collection frequency:	Every 30 sec

Radiosonde Sounding	
Description:	Vertical measurements made with drifting balloon
Location:	Within 5 km of the WSI
Data collection frequency:	2-3/day

LandSat Satellite (Satellite Sensor)		
Sensor: Thematic Mapper (TM)		
Orbit: Polar		
Bands:	3 visible	0.45 - 0.52 μm
		0.52 - 0.60 μm
		0.63 - 0.69 μm
	IR band	0.75 - 2.35 μm
Data collection frequency:		Same location captured once every 16 days
Spatial resolution:		25 m
Overpass time:		27 May 1992, 17:02:38.53 - 17:03:05.398 UT
Area covered:		White Sands, NM
Navigation:		Provided by LandSat company using Universal Transverse Mercator (UTM) projection to get the 25-m equally spaced scene centered on the WSI location.

GOES (Satellite Sensor)		
Sensor: VAS		
Orbit: Geostationary (NOAA)		
Bands:	1 visible 12 IR	broadband
Data collection frequency:	same location captured once every 30 min	
Spatial resolution:	Visible: IR:	1-km 4-km
Overpass time:	First image: Second image:	04 May 1992, 21:20 UT 27 May 1992, 17:01 UT
Area covered:	White Sands, NM	
Navigation:	Satellite image is in GOES line and element coordinates which is approximately equally spaced in the mid latitudes. GOES supplies a navigation code which provides the latitude and longitude for each pixel.	

DEM	
Description:	Digital elevation map
Spatial resolution:	3 arc second
Navigation:	Using *arc2utm*, we converted the arc units to UTM units which allowed navigation and overlapping with the satellite images.

6.6.1 Cloud Horizontal Distribution. Images form the Whole-Sky Imager (WSI)

One data stream that has proved very useful in this work is the Whole Sky Imager (WSI). The WSI is a ground-based imager that acquires images of the entire sky-dome in two narrow spectral ranges centered at 450 nm and 650 nm. A detailed description of the instrument and the data which it produces can be found at the Atmospheric Radiation Measurement (ARM) World Wide Web site at *www.arm.gov*. For our work, the important features of the WSI are the following. The data from the WSI are images of the sky radiance. The images are very close to equi-angular projections due to the ground perspective and the lens used for the imaging. That is, the azimuth angle is preserved and the zenith angle is proportional to the distance from the center of the image. The range from the WSI to the object is not known. Thus, to locate all three Cartesian coordinates of a point in the

image, one length must be known. For example, the distance between the WSI and the cloud or the distance from the ground to the cloud base.

Due to the projection of the cloud scene into the WSI image, the image must be "flattened" in order to be compatible with the satellite images. This flattening is essentially a mapping of the pixels in the WSI image coordinates to Cartesian coordinates. As mentioned above, one length must be known to do this mapping and, we typically use the estimated cloud base height. This mapping is not exact in that all the cloudy points are remapped based on the same height. The result is a flat image that retains ground perspective. That is, the image appears flat, but the sides of clouds away from center are still visible.

6.6.2 Cloud Base Height

The cloud base height can be estimated by using two WSIs, a ceilometer, or by estimating the height of the lifting condensation level (LCL) and assuming that the cloud base height is at the LCL. The lifting condensation level is the level (height) at which water condenses from rising, moist air. A technique that has been developed in previous work is to correlate the images from a separated pair of WSIs and locate corresponding areas in each image. Triangulation gives the altitude if the location (*i.e.* separation) of the two imagers is known. This technique is documented in Allmen (1993) and will not be repeated in detail here. Basically, small regions in one image are correlated with the second image. A maximum in the correlation function signifies that the two image regions are stereo views of the same region of the sky. The height of this region can be then be determined from triangulation. The correlation is done in a computationally efficient way. Rather than try to correlate each region with the entire second image, each region is correlated only with those parts of the second image where a match could physically exist. That is, each point in the first image traces out a curve in the second image based on the height of the point. It is only along this curve that the correlation is performed. We found that this technique, due to the correlation not being a relatively simple function but having many high values, sometimes returned cloud base values that were either very different from surrounding values or physically unreasonable (such as a cumulus cloud base height being 14 km). To automatically eliminate some of these stray points, we conditioned the cloud base height values returned by the algorithms to lie in a pre-set range based on the type of cloud present in the image. In other words, we effectively shortened the curve along which the correlation was performed by putting a constraint on the height.

The ceilometer is a zenith viewing instrument that measures the cloud base height from the round-trip time for laser light between the ceilometer and the cloud base. For fair weather cumulus cloud fields, the cloud base height is relatively constant and is generally determined from the altitude at which water condenses (*i.e.* the LCL). Thus, the time series of ceilometer measurements gives some average cloud base height for the cloud field.

We used a fair weather cumulus cloud model for the lifting condensation level (LCL) to estimate the cloud base heights. This model assumes a plane parallel atmosphere with a well mixed lower region. Solar heating of the air near the ground creates thermals of hot, moist air parcels that rise due to a positive buoyancy force. As these air parcels rise, their pressure adiabatically decreases causing their temperature to drop. The LCL is the height where the temperature decreases to the condensation temperature of the moist air in the parcel. We assume this height is the same as the cloud base height. In general, this is a valid assumption if the clouds are not being affected by frontal passage. The mixing ratios of the air parcels are determined from radiosonde sounding data. The computed LCL is

actually the mean height obtained from many air parcels that are "released" from different levels in the mixed atmosphere region.

6.6.3 Cloud Top Height

The cloud top can be estimated by several methods. From IR satellite imagery, from models, and from the WSI. For our purpose, which requires high spatial resolution to match the resolution of the WSI, the LandSat thematic mapper (TM) images are the best choice of satellite images, having 25-m resolution. This allows determination of the cloud top height at the local scale from infrared (IR) images. However, the TM has several limitations. First, the effect of ground contamination cannot be totally eliminated (*i.e.* contamination of the cloud top infrared radiation by the infrared radiation emitted from the ground surface). This can give a cloud top height estimate that is sometimes below the actual cloud top height because ground contamination yields an effectively warmer top, therefore a lower cloud top height. The second limitation is the sparseness of data since the TM is a polar orbiting satellite and covers a particular swath only once every 16 days. This makes it impossible to match the several minute temporal resolution of the WSI and other data streams.

The second method for obtaining the cloud top height for cumulus clouds is by modeling the convective motions in the atmosphere. We have done this by extending the model we used for estimating the cloud base height, the positive buoyancy force continues to carry the air parcels upwards. To make the calculation tractable, we added the following three assumptions to the model: (1) no ambient air is entrained into the rising air parcel, (2) condensed water is not accounted for in the heat or momentum budgets, and (3) the air parcels do not overshoot their final height. The air parcels continue to rise and the water vapor condenses as the temperature continues to drop pseudo-adiabatically (the latent heat of condensation is accounted for, but the cooling of the liquid water is not). The convection continues until the buoyancy of the air parcel decreases to a minimum. Figure 14 shows how the normalized buoyancy force changes with height for data from White Sands on July 11, 1994. At 3036 m, the buoyancy has a deep drop and we define this as the cloud top height.

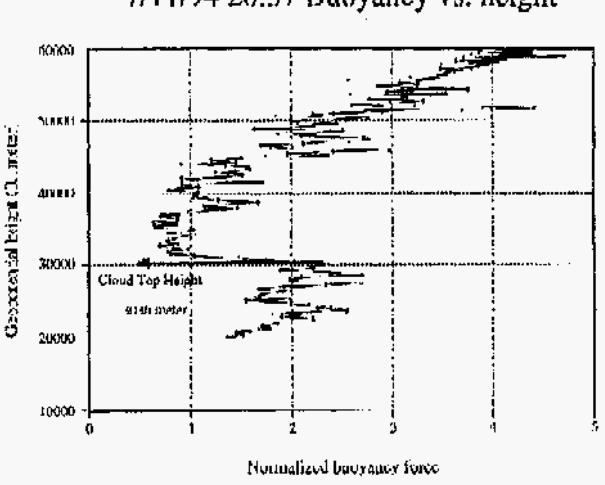


Figure 14. Normalized buoyancy as a function of height. Cloud bottom height occurs at 3036 m as indicated by the rapid decrease in buoyancy.

The third method for cloud top height estimation is the WSI data. Since the WSI is a ground-based instrument, the resulting image will contain the sides of clouds viewed off

zenith. Once the physical scale is set, the height of the highest visible part of the cloud can be calculated. For cumulus clouds that are not of the typical "mushroom" shape, the highest visible part may be lower than the actual highest part of the cloud. However, for clouds nearer the edge of the WSI image, this becomes less likely. For now, we use the assumption that the highest visible part corresponds to the actual cloud top height.

6.6.4 Cloud Shape Parameters

The WSI gives a high spatial resolution image of the local cloud field. This image can be used to extract the other parameters needed for the reconstruction, namely, the cloud base shape and vertical profile shape. For the ground-based WSI and a convective cloud field, the individual clouds can readily be discerned by the general shape and shading. The flat bottoms are darker, appear smooth, and appear towards the outer edge of the image. The sides of the clouds are brighter and are projected towards the center of the image. Once the length scale for points making up the cloud base is known (by setting the cloud base height, for example), and assuming that the entire cloud base is at the same height, the position, size, and detailed shape of the entire cloud base can be extracted by extracting the darker cloud bottoms from the image.

Furthermore, for clouds sufficiently far from the center of the image (*i.e.* for small elevation angles), the profile shape of a vertical slice through the cloud can be extracted. This profile will not necessarily contain the highest point of the cloud, nor will it necessarily be the profile of truly vertical slice through the cloud. This profile will, however, be one possible such profile around the cloud perimeter. Furthermore, for a typically-shaped cumulus cloud, this profile will be very close to truly vertical, center slice profile. These profile shapes are some of the highest resolution shape information available on convective clouds and can be used to generate statistical information about the shapes of the clouds, or compute the fractal dimensions of clouds. This information can also be used to generate physically viable cloud shapes for models.

For this method, any error in locating the true cloud top in the WSI image will result in an error in the extracted cloud top height. Furthermore, for clouds at lower elevation angles (*i.e.* nearer the edges of the image), the same error in locating the position of the cloud top in the image will result in a corresponding larger error in the cloud top height. Figure 15 shows the corresponding error in the cloud top height for a given error (in pixels) in locating the cloud top in the image. The different curves represent different elevation angles and cloud base heights.

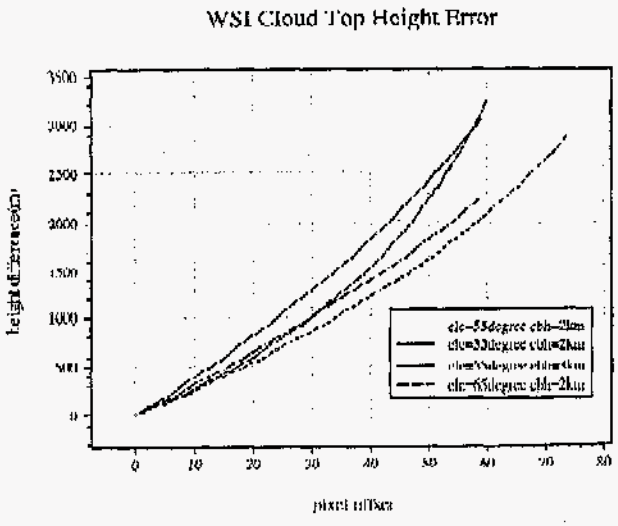


Figure 15. Error in derived cloud top height as a function of pixel error (pixel offset). Error increases as the elevation angle decreases.

6.7 Fusion Processes

The fusion process proceeds by combining the various data mentioned above. First, is the comparison of similar results from different data sources. The cloud base height can be estimated from the LCL model and measured by the ceilometer. Figure 16 shows a comparison of the cloud base height from these three sources. The cloud top height can be estimated from the fair weather cumulus model and from the WSI data. The error bar on the WSI-derived cloud top height value reflects the uncertainty due to mislocating the actual cloud top in the WSI image (as detailed in the previous section). For the data on July 11, 1994.

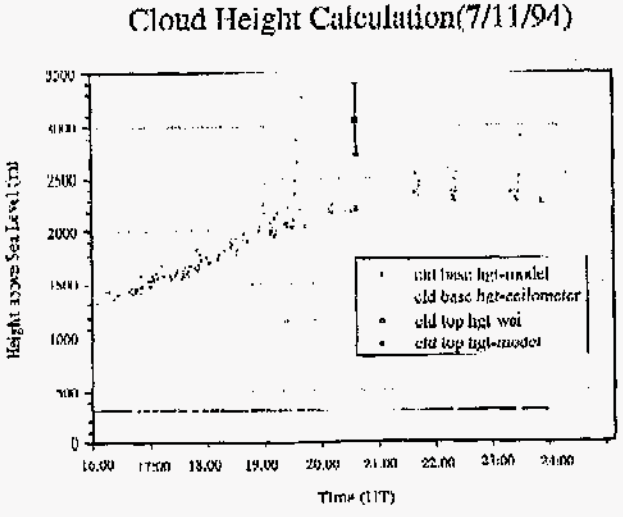
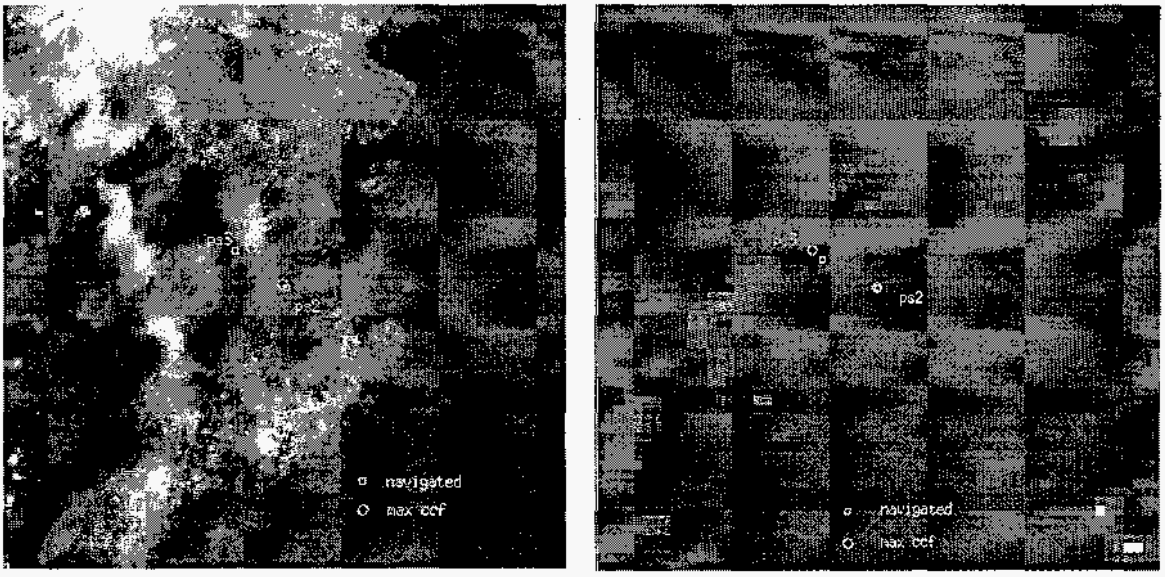


Figure 16. Results of cloud top height determination by three different methods.

Next is an assessment of the registration of the satellite and WSI data. This registration is important for a comparison of the WSI-derived cloud base height and the satellite-derived cloud top height. As previously mentioned, due to ground contamination of the

TM data, the satellite-derived cloud top height can be lower than the actual cloud base height as in the example we present below. This is physically impossible. For the registration of the WSI and TM images, correlation of two images were performed. A correlation of the radiance images was attempted, but was unsuccessful due to the large radiance difference between the upper and lower views of the clouds. Therefore, for the correlation, the images were converted to binary cloud/clear images by thresholding. This yielded better agreement but there were still multiple maxima in the correlation result. Figure 17a shows the best correlation result. The locations of the two WSIs are labeled with ❑ and the locations of the WSI based on the correlation are marked with ○. The distances between the actual and correlation-derived locations are 354 m and 1521 m. For this result, the TM data was degraded to 250-m resolution to allow the correlation to be computed in a reasonable amount of time. Figure 17b shows the correlation result between the GOES data and the WSI, again at 1-km resolution. This result is better than the TM result and may be due to the cloud type present in the images. It appears that cumulus clouds are better for the registration by correlation than are altocumulus or stratus. This is physically reasonable since the cumulus cloud scene will have more well defined structure (*i.e.* the cloud boundaries) than will stratus and some types of altocumulus.



Thematic Mapper Image (TM) GOES Satellite Image

Figure 17. (a) TM and (b) GOES satellite images showing the nadir view of the WSI site. Navigation marks are indicated.

Once the TM and WSI images are properly navigated, the cloud thickness can be determined from the difference between the TM-derived cloud top and the WSI-derived cloud base heights. Figure 18 shows a contour plot of cloud thickness. For this figure, the cloud distribution (*i.e.* which pixels are cloud and which are clear sky) is determined from the TM image. The cloud pixels are then assigned a value equal to the difference between the cloud top and base heights. Because of the constraint imposed on the WSI-derived cloud base heights, some cloud pixels did not have a cloud base height associated with them. These gaps were filled in using an interpolation process based on the surrounding pixel's cloud base height values. As mentioned above, due to ground contamination in the TM image, some of the cloud top heights were below the WSI-derived cloud base heights resulting in a negative cloud thickness. For this figure, the cloud base heights were assumed more reliable than the cloud top heights and any pixels with a negative thick-

ness were removed and then replaced with an interpolated cloud top height based on surrounding values.

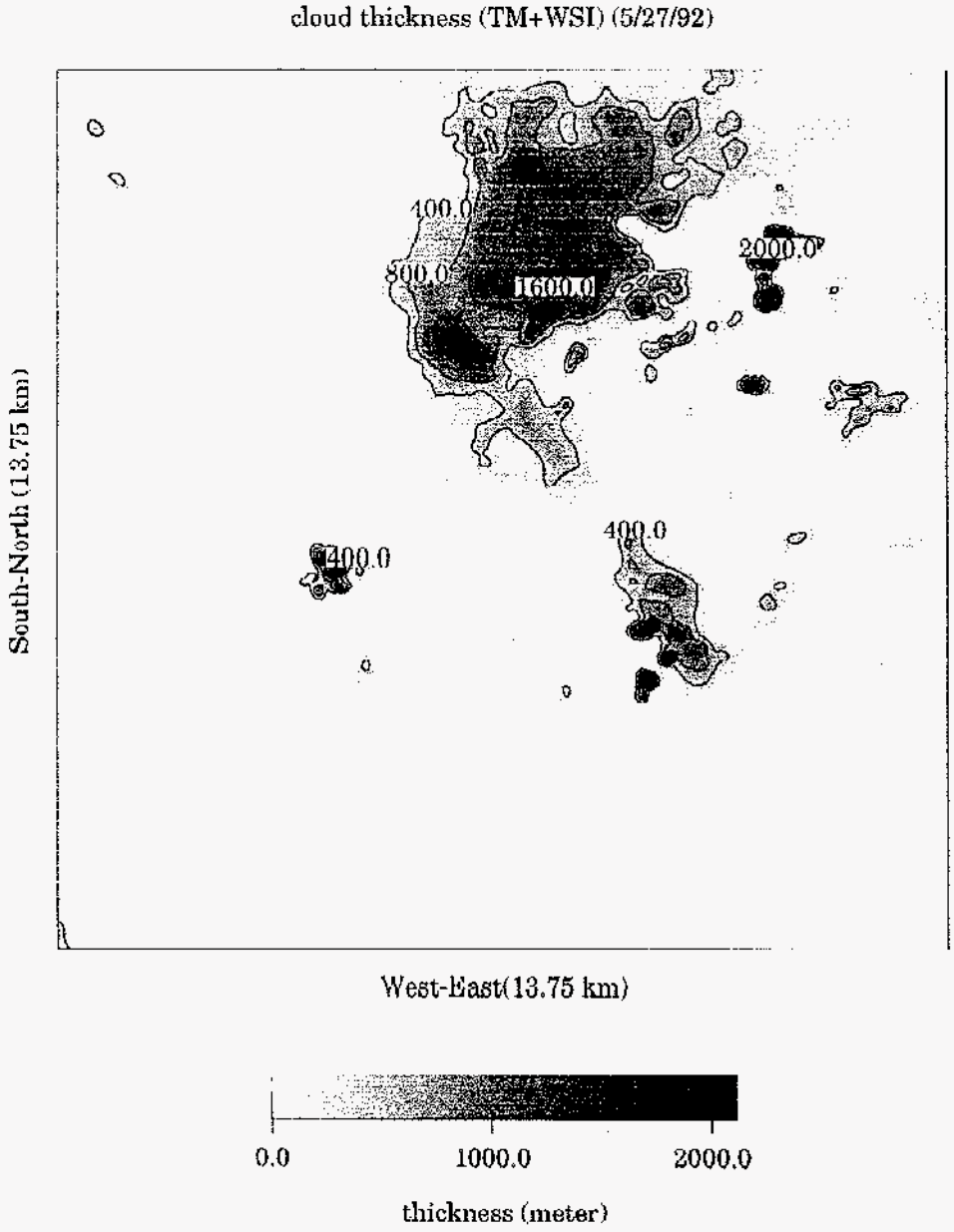


Figure 18. Contour plot of the cloud thickness.

6.8 Results

We now present two cloud reconstructions obtained from the processes and methods outlined above. One is the reconstruction of a stratiform cloud field and the other is the reconstruction of an individual cumulus cloud. The stratiform cloud case is essentially a subset of the convective cloud case in that not all the data streams and assumptions are required for the reconstruction. The stratiform case is shown in Figure 19. This scene is centered on the WSI locations at the White Sands Missile Range in New Mexico. The

WSI's are in a valley. The upper left figure shows the WSI image and the upper right shows the corresponding flattened image. The lower left image shows the TM image. Note the strong similarity between the flattened WSI image and the TM image. The lower right image is the reconstructed 3D cloud scene, here shown with the ground topography obtained from the digital elevation map (DEM).

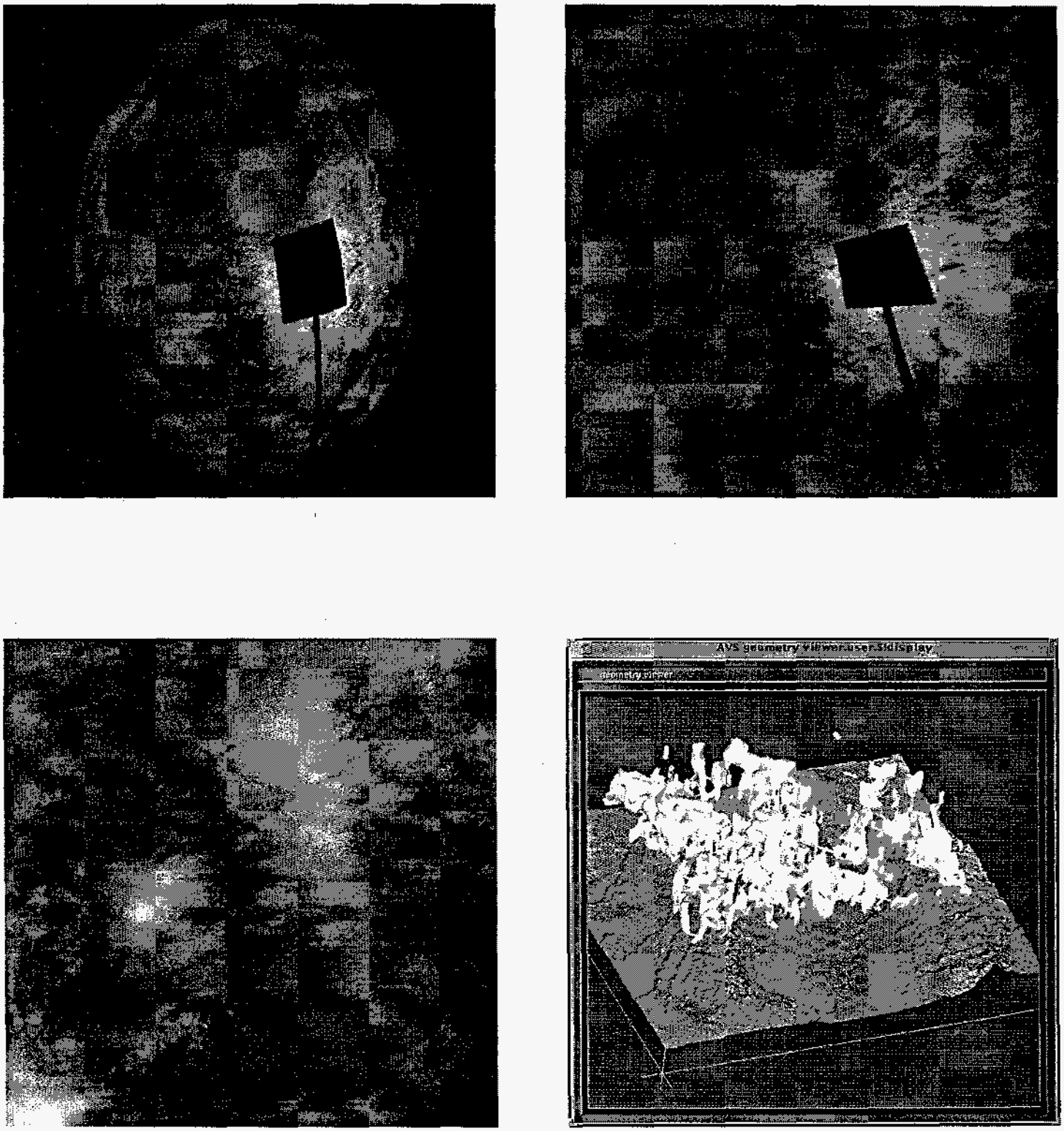


Figure 19. 3-D reconstruction of stratus cloud field.

In order to point out the relation between the cloud structure and the ground topography, Figure 20 shows a plane view of the DEM and the TM image with the ridges surrounding the valley superimposed. Figure 9 is the surface weather map for the same time. From Figure 21, we can see an occluded front near the White Sands area that has persisted for several days. An occluded front occurs when an active cold front overtakes a warm front. As the advancing cold air wedges the warm front air mass vertically upward; the new occluded front emerges. The weather in such a situation is generally complex. Occluded fronts are common east of the Rockies. Precipitation often accompanies such a front. This occluded front formed on May 26 and is moving slowly towards the east. In the GOES and TM satellite images, we can see stratus, cirrus, and cumulus clouds moving into the area, but the majority of the clouds are stratus. The DEM image shows the two high mountain ridges running north-south to the east and west of the White Sands area. The surface wind is from the south. In the cloud image, we can see the shape of the cloud cluster resembling the shape of the valley floor. It is apparent that the clouds have entered the valley from the south and they are confined to the valley by the mountain ridges. Also apparent are four bands of cloud streets aligned with the wind direction.

The second result, shown in Figure 22, is the reconstruction of an individual fair weather cumulus cloud. The cloud base height is assumed uniform and is estimated from the LCL model. The cloud top profile can be extracted from the side view of the cloud in the WSI. The horizontal cloud base shape can also be extracted from the WSI image, however, for simplification, we used a circular shape for this example. Since only one arc across the top of the cloud is measurable in the WSI data, we need to interpolate the rest of the cloud top and sides based on this one measured profile. This interpolation is based on a normalized radial cloud top profile ($r = 0$ to 1, where $r = 0$ is the center and $r = 1$ is on the cloud base edge) and smoothly varies the profile using the two measured radial profiles as boundary conditions. The cloud top profile used in this reconstruction is shown in Figure 23, where, for presentation purposes, one side of the profile is shown going from 0 to +1 and the opposite side is shown from 0 to -1. Each radial profile to be interpolated is generated by a weighted average of the two measured radial profiles. The weighting is calculated from the angle between the profile to be interpolated and the two measured profiles and is given by $p_1 \cdot (\theta/\pi) + p_2 \cdot [1-(\theta/\pi)]$ where profile p_1 is at an angle of π and profile p_2 is at an angle of 0. Each side of the cloud is treated independently.

6.9 Discussion

This section will discuss some of the difficulties encountered in this work and identify the areas that require further work. First, reconstruction of the cloud field is a very difficult problem due to the limited information we have available. We have essentially two views of the exterior of a highly complex object and this is not enough information to perfectly reconstruct the cloud scene. The reconstruction of ideal fair weather cumulus clouds is beginning to be realizable, but handling more realistic cases with multiple cloud types and cloud layers is still very far off.

The cloud top height is probably the most difficult parameter to extract from the data. The fair weather cumulus model we used contains many assumptions and can only give some average value for the top height. The TM data, while having superior spatial resolution, is contaminated by the ground signal and the existing algorithms for extracting cloud top height can not deal with this contamination for the thinner clouds. The cloud top height estimated form the WSI is also an estimate and will not, in general, be the true height.

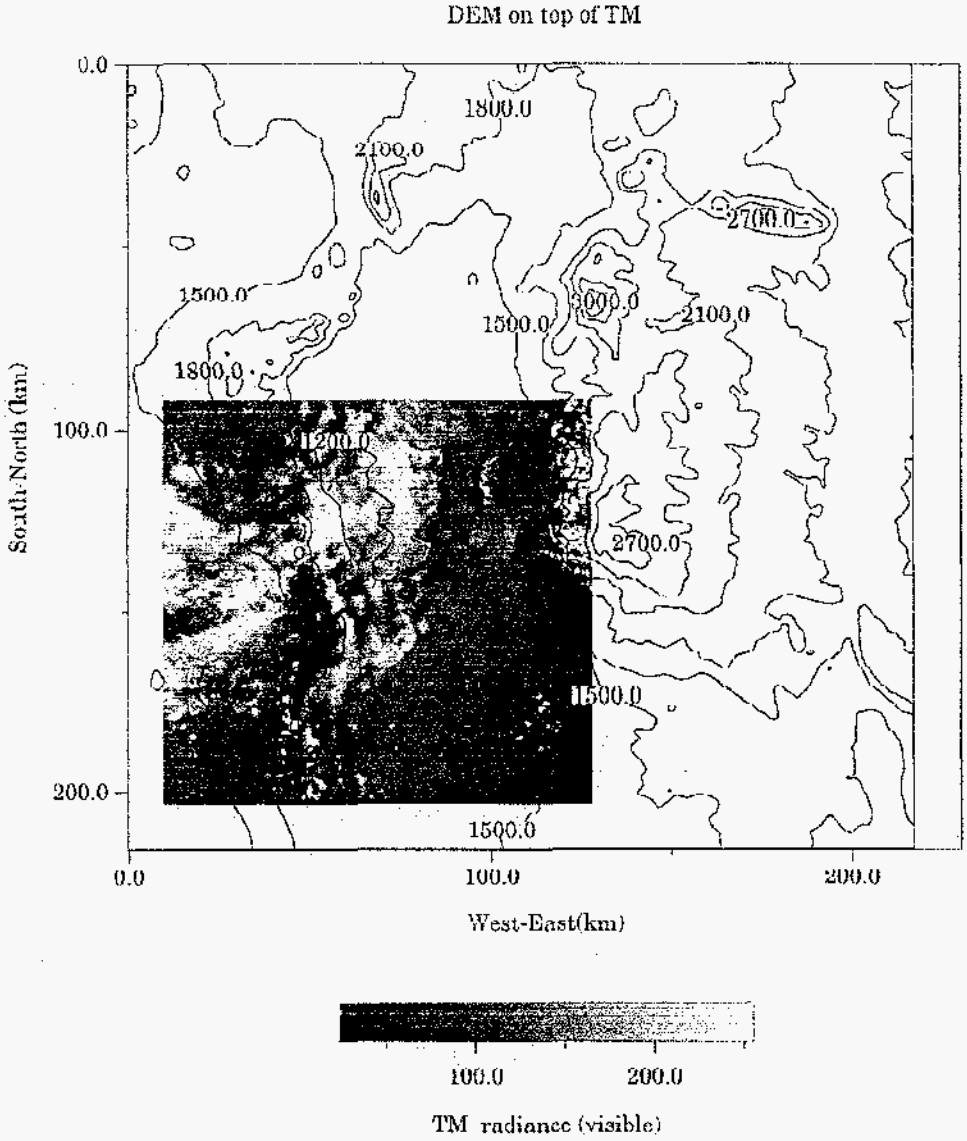


Figure 20. 2-D map of cloud distribution showing surface topography.

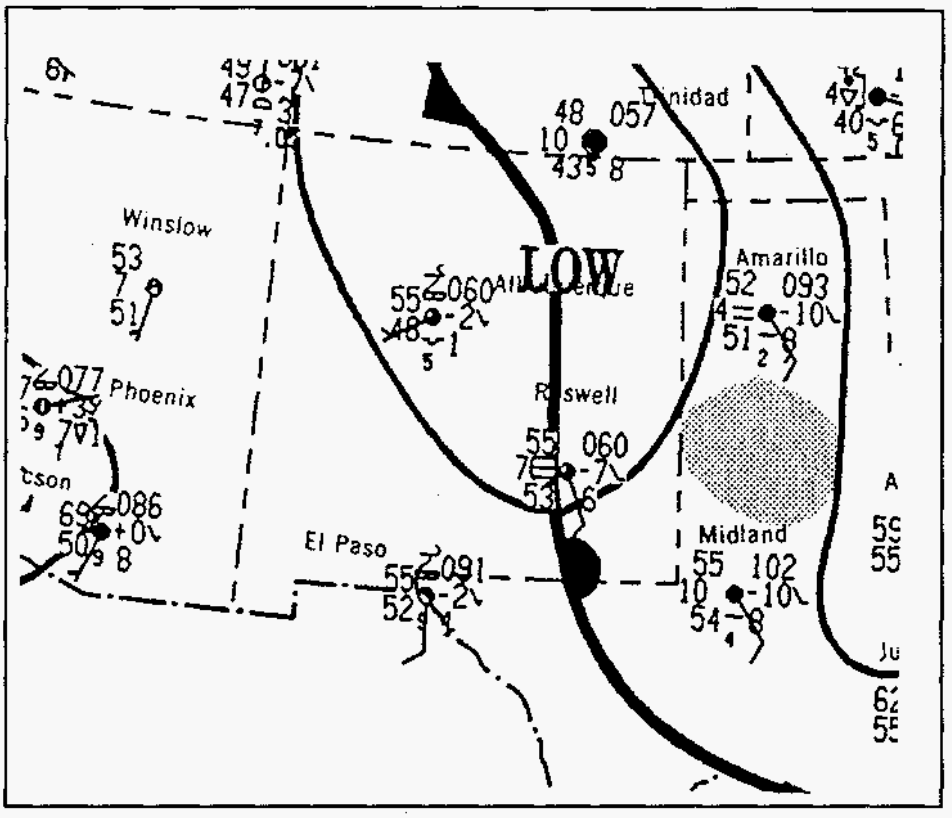


Figure 21. Weather map showing occluded front. The WSI's are located at White Sands, New Mexico, about 90 miles north of El Paso, Texas.

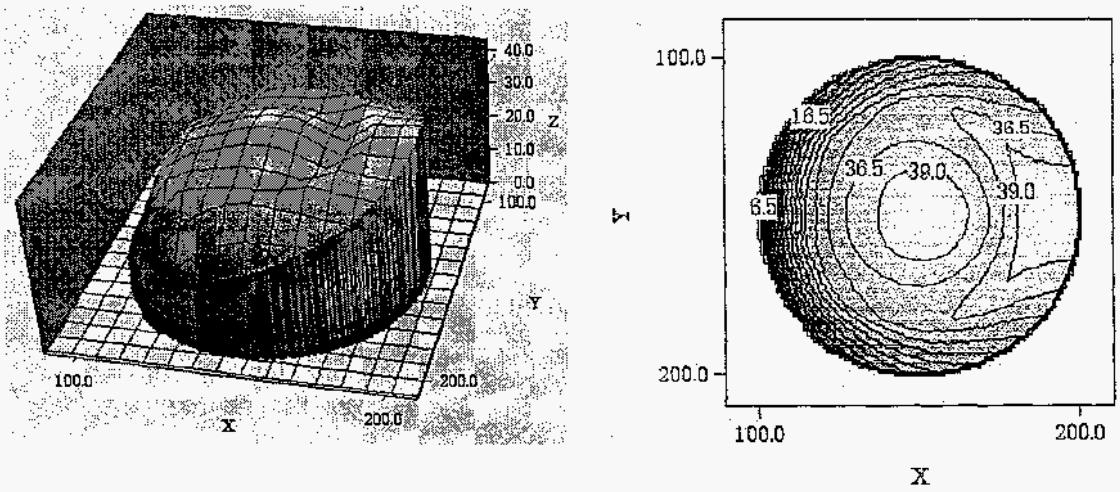


Figure 22. Left: 3-D reconstruction of a cylindrical section of a fair weather cumulus cloud; Right: 2-D topographic map of the top of the cloud surface.

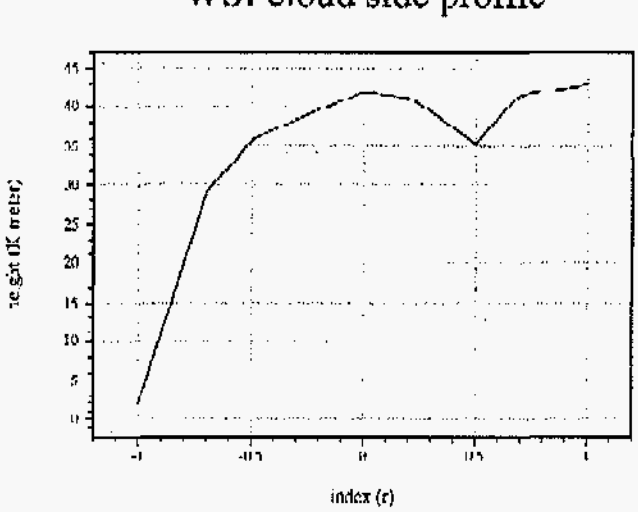


Figure 23. Cloud profile as derived from the WSI image.

The cloud base height is somewhat easier to determine. The convective cloud case is very amenable to the uniform cloud base height assumption which is measured by the ceilometer and estimated by the LCL model.

Navigating the satellite and WSI imagers is also not trivial. The two instruments have very different views of the cloud field and there is only a general correlation in the appearance of the clouds to these two instruments.

Furthermore, we have made no attempt to discern the internal structure of the clouds, which is very important to the part that clouds play in the radiation budget. Nevertheless, we have succeeded in producing a reconstructed cloud field, albeit with many assumptions.

7 Summary

Clouds have a strong influence on the Earth's climate and therefore on climate change. An important step in improving the accuracy of models that predict global climate change is improving the parameterization of clouds and cloud-radiation interactions. An important improvement in the next generation models will likely be the inclusion of cloud geometries in the cloud-radiation parameterizations. We have developed and reported here methods for characterizing the geometrical features and three-dimensional properties of clouds that could be of significant value in developing these new parameterizations. We developed and report here a means of generating and imaging synthetic clouds which we used to test our characterization algorithms; a method for using Taylor's hypotheses to infer spatial averages from temporal averages of cloud properties; a computer method for automatically classifying cloud types in an image; and a method for producing numerical three-dimensional renderings of cloud fields based on the fusion of ground-based and satellite images together with meteorological data.

References

1. "The 'Greenhouse' Effect and Climate Change," J. B. F. Mitchell; Rev. Geophysics 27, 115-139 (1989).

2. "Interpretation of Cloud-Climate Feedback as produced by 14 Atmospheric General Circulation Models," R. D. Cess, G. L. Potter, J. P. Blanchet, G. J. Boer, S. J. Ghan, J. T. Kiehl, H. Le Treut, Z.-X. Li, X.-Z. Liang, J. F. B. Mitchell, J.-J. Morcrette, D. A. Randall, M. R. Riches, E. Roeckner, U. Schlese, A. Slingo, K. E. Taylor, W. M. Washington, R. T. Wetherald, I. Yagai; Science 245, 513-516 (1989).

3. "Atmospheric Radiation Measurement Program Plan". U. S. Department of Energy Publication DOE/ER-0441, February 1990. Office of Energy research, Office of Health and Environmental Research, Atmospheric and Climate Research Division.

4. "The Hydrological Cycle and its Influence on Climate," Moustafa Chahine; Nature 359, 373-380 (1992).

5. "The Observational Determination of the Greenhouse Effect," A. Raval and V. Ramanathan; Nature 342, 758-761 (1989).

6. "Sensitivity of the Earth's Radiation Budget of Changes in Low Clouds," Nature, 343, 49-51 (1990).

7. "ISCCP Cloud Algorithm Intercomparison," W. B. Rossow, et al.; Clim. Appl. Meteorol. 24, 877-903 (1985).

8. "Our Changing Planet, the FY93 U. S. Global Change Research Program," A Report by the Committee on Earth and Environmental Sciences , Washington, DC (1992).

9. "Report of the ARM Hierarchical Diagnosis Workshop," Richland, Washington, December 16-17, 1992, M. Bradley and J. Kiehl.

10. "On the Effects of Cumulus Dimensions on Longwave Irradiance and Heating Rate Calculations," R. G. Ellingson, J. Atmos. Sci., 39, 886-896 (1982).

11. "Infrared Radiative Transfer Through a Regular Array of Cuboidal Clouds," J. A. Harshvardhan and J. A. Weinmam, J. Atmos. Sci., 39, 431-439 (1982).

12. "The Measurement of Spatially Averaged Wind Profiles with a Volume Imaging Lidar," E. W. Eloranta and J. L. Schols, Abstracts of 15th International Laser Radar Conference, July 1990, Tomsk, URRS.

13. "Fluxes and Directional Effects of Broken Clouds," D. Cretel, M. Herman, and D. Tanre, IRS '88, Lenoble and Geleyn (Eds.) A. Deepak Publishing (1989).

14. "Automated visibility and cloud cover measurements with a solid state imaging system." Johnson, R., Hering, W., and Shields, J. Technical Note GL-TR-89-0061, Marine physical Laboratory, SCRIPPS Institution of Oceanography, San Diego, CA, 92151-6400, March 1989.

15. "Modeling the gometrical properties of the WSI instrument." T. L. Koehler, Tech Rep. AV89-070t, Marine Physical Laboratory, SCRIPPS Institution of Oceanography, San Diego, CA, 92151-6400, July 1989.

16. "The Computation of Cloud Base Height from Paired Whole-Sky Imagers," M. Allmen, W. P. Kegelmeyer, Fourth Symposium on Global Climate Change Studies, American Meteorological Society Meetings, Anaheim, California, January 1993.

17. "Cloud scene simulation modeling." M. E. Cianciolo, and R. G. Rasmussen, In Proceedings of the 1991 Cloud Impacts on DoD Systems Conference (1992).

18. "The Spectrum of Turbulence, ". G. I. Taylor, Proc. Roy. Soc. London, A165, 476-484 (1938).

19. "The User-Interactive Cloud Decision Program, DECISION C," T. L. Koehler, Marine Physical Laboratory, Scripps Institution of Oceanography, Report AV92-043t.

20. "An Investigation of the Application of Taylor's Hypothesis to Atmospheric Boundary Layer Turbulence," David C. Powell, C.E. Elderkin, J. Atmos. Sci., Vol. 31, 990-1002 (1974).

21. "On the use of Taylor's translation hypothesis for diffusion in the mixed layer,". G. E. Willis, G.E. and J.W. Deardorff, Quart. J. Roy. Meteor. Soc., 102, 817-822 (1976)

22. "Pattern recognition analysis of polar clouds during summer and winter.", E. E. Ebert, 1992:. *Int. J. Remote Sensing*, **13**, 97-109.

23. "Automated recognition of oceanic cloud patterns and its application to remote sensing of meteorological parameters" L. Garand, *Ph.D. Thesis*. University of Wisconsin, Madison, WI. (1986)

24. "A cloud type classification with NOAA 7 split-window measurements." T. Inoue, *J. Geophys. Res.*, **92**, 3991-4000 (1987).

25. "Cloud type discrimination via multispectral textural analysis." N. M. Lamei. and Crawford, K. D. Hutchison, and N. Khazenie, 1993 *Proc. SPIE*, **1934**, 49-61 (1993).

26. "A neural network approach to cloud classification," J. Lee, R. C. Weger, S. K. Sengupta, and R. M. Welch, *IEEE Trans. Geosci. Remote Sens.*, **28**, 846-855 (1993).

27. "Neural network cloud classification research," Smotroff, I. G., Report No.: MTR-93B0000039. MITRE Corp., Bedford, MA.(1993).

28. Remote discrimination of clouds using a neural network" S. Yool and M. Brandley. *Proc. SPIE*, **1766**, 497-503 (1992).

29. "Textured image segmentation, Laws, K. I., *:Ph.D. Thesis*. University of Southern California, Los Angeles, CA.(1980)

30. "The pyramid as a structure for efficient computation," Burt, P. J., in *Multiresolution image processing and analysis*. A. Rosenfeld. Springer, Heidelberg, Germany. 6-35 (1984).

31. For a visualization of congestive cloud formation at 10-minute intervals, see p.269 "Atmospheric Science, An Introductory Survey," J. M. Wallace and P. V. Hobbs, Academic Press, Inc. Harcourt, Brace and Jovanovich Publishers, San Diego (1977).

32. "Three-Dimensional Imaging of Cloud Structure Using Satellite Radiances, A New Tool for Cloud Studies," E. Boer and V. Ramanathan, 1992 AGU Spring Meeting (1992).

33. "Validation of a Three-Dimensional Cloud Retrieval Scheme by Incorporating ARM/CART Measurements for Surface Fluxes and Cloud Fields," L. Shi and V. Ramanathan, Fourth Symposium on Global Climate Change Studies, American Meteorological Society Meetings, Anaheim, California, January 1993.

34. "On the Determination of Cloud Cover from Satellite Sensors: The Effect of Spatial Resolution," B. A. Wielicki and L. Parker, *J. Geophysical Res.*, **97**, 12,799-12,823 (1992).

35. "Measuring Cloud Properties from Space: A Review," W. B. Rossow, *J. Climate* **2**, 201-213 (1989).

36. "The 27-28 October 1986 FIRE IFO Cirrus Case Study: Cloud Parameter Fields Derived from Satellite Data," P. Minnis, P. W. Heck and E. F. Harrison, *Monthly Weather Review*, **118**, 2428-2446 (1990).

INITIAL DISTRIBUTION

MS0188	D. L. Chavez, 4523			
MS0188	C. E. Meyers, 4523			
MS1109	C. F. Deigert, 1424			
MS9001	T. O. Hunter, 8000			
		Attn:	MS9002	P. N. Smith, 8500
			MS9003	D. L. Crawford, 8900
			MS9005	J. B. Wright, 2200
			MS9054	W. J. McLean, 8300
			MS9141	P. E. Brewer, 8800
			MS9405	T. M. Dyer, 8700
			MS9420	L. A. West, 8200
MS9004	M. E. John, 8100			
MS9009	R. C. Wayne, 8400			
MS9041	J. S. Binkley, 4531			
MS9056	M. Lapp, 8102			
MS9056	T. Tooman, 8120			
MS9056	J. Vitko, Jr., 8102			
MS9104	W. Bolton, 8120			
MS9213	K. Buch, 8117 (5)			
MS9056	L. R. Thorne, 8120 (10)			
MS9021	Technical Communications Dept., 8815 for OSTI (10)			
MS9021	Technical Communications Dept., 8815/Technical Library, MS0899, 4414			
MS0899	Technical Library, 4414 (4)			
MS9018	Central Technical Files, 8940 (3)			