1. Introduction

1.1 Importance of clouds in the climate system

Clouds play an important role in the earth's climate system by reflecting incoming solar radiation and absorbing and re-emitting outgoing long-wave radiation. According to the most recent report by the Intergovernmental Panel on Climate Change (IPCC 2001), probably the greatest uncertainty in future projections of climate arises from clouds and their interactions with radiation. In addition, clouds and convective systems are essential components of the atmospheric energy and water cycles. Through their vertical transport of heat, moisture and momentum, and in particular the release of latent heat, convective clouds essentially force the large-scale flows. Realistic climate simulation and prediction requires that cumulus convection be adequately represented in atmospheric general circulation models (AGCM). Indeed, parameterisation of the effects of cumulus convection on the large-scale flows in which it is embedded remains among the key unresolved challenges to modelling the atmospheric general circulation (Donner 1993).

1.2 Representation of cumulus clouds in GCMs

Cumulus clouds are formed by convective processes. Moist air in surface layers becomes unstable and rises in plumes, expanding and cooling as it does so, until saturation is reached. This simple explanation of cumulus formation belies the complex and poorly understood multi-scale nature of cloud formation and microphysical processes (Yao and Del Genio 1999). Cumulus clouds are typically 1-2km in diameter (LeMone and Zipser 1980), with the largest around 10km in diameter. Current global models do not have horizontal resolutions capable of capturing individual clouds; typical large-scale models have horizontal grid spacings of around 200 km. Sub cloud processes (e.g. turbulence, entrainment) occur on much smaller scales still, however they are closely linked to the large-scale flow, via for

example the release of latent heat. It is the sub-grid scale nature of convective clouds that necessitates their parameterisation in GCMs.

1.3 Convective parameterisations

A typical convective parameterisation scheme attempts to represent the mean effect of the sub-grid scale processes on the large-scale flow. Early parameterisation schemes did this by representing the collective influence of the many clouds rather than the effect of each individual cloud. Since convection is essentially a transport process, *mass flux* schemes have frequently be used, with the aim of providing a parameterised representation of the sub-grid convective contributions to the overall transport of heat, moisture and momentum (e.g. Anthes 1977). Mass flux schemes are still widely used for cumulus parameterisation today. They approximate the overall mass flux in a single AGCM column, however a number of clouds and possibly cloud types are likely to be found in any particular column. This detail is lost in the averaging effects of parameterisation. In an effort to mitigate the problem, grid cells are often divided into cloudy and non-cloudy regions (e.g. Yao and Del Genio 1999), based on some property such as relative humidity (Figure 1). The vertical profiles of the average properties of these two areas can then be estimated (e.g. Teidtke 1989) and these estimated properties are incorporated into a small number of fixed parameters which implicitly describe the spectrum of different clouds (Nober 2003). Relatively simple schemes such as these have the advantage of being fast enough to be used for long climate model runs.

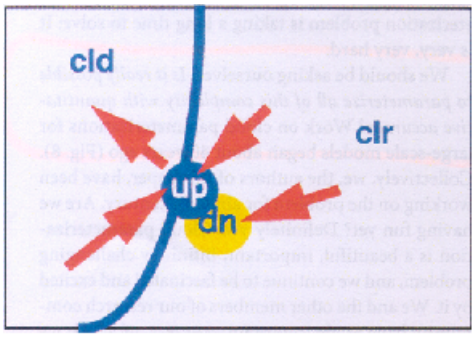


Figure 1: Schematic illustrating a cloud parameterisation with cloudy and non-cloudy regions (Randall and Fowler 1999). The blue and yellow dots represent ensemble updrafts and downdrafts respectively, and the red arrows indicate entrainment and detrainment.

More complex schemes have been proposed (e.g. Donner 1993, Nober 2003) which consider an ensemble of cumulus clouds within a grid cell rather than the collective effect of all the clouds within that cell. The ensemble of clouds can be generated using a cloud-resolving model (CRM), and the use of CRM results in conjunction with GCMs in this way is expected to lead to the improved parameterisation of cloud processes. In a recent review of the subject (Randall et al. 2003) it has been suggested that future research should focus on *superparameterisation*, i.e. the use of a CRM within each grid cell of a large-scale model (e.g. Tao et al. 1987). In particular the authors of the review advocate the use of 3D CRMs, in so called super-GCMs, as a way of tackling the complexity of the cumulus parameterisation problem. Such models would, however, require considerable computation time.

Coupling of the convective cloud system with larger scales by forcing of the GCM results in a reduction in the degrees of freedom. This has already been done with some success by Donner (1993), who used a parameterisation which calculated explicitly distributions of cumulus vertical momentum and mass flux using lognormal distribution functions (see Section 2.3), and was closed with Global Energy and Water Cycle Experiment (GEWEX) Atlantic Tropical Experiment (GATE) precipitation observations. The shift from traditional mass flux schemes and average cloud properties and effects, to schemes involving analysis of model output and the use of statistics to describe the state of a cloud ensemble, requires increased knowledge of

cloud fields structure and development. There has been much research into the physical structure of convective cloud fields, and in particular the size distribution of cumulus clouds, which it is hoped will lead to a better representation of their inhomogeneity and patchiness in GCMs. The importance of this was emphasised by Frank and Cohen (1985) who found that specifying the correct ensemble is as important as specifying an ensemble at all. As yet there is no general agreement on a universal functional form for the distribution of cumulus cloud sizes.

In the following section, the proposed functional forms will be discussed, together with some description and comparison of the methods used in the numerous studies. It is proposed that some of the variation in results might be explained by the different approaches and techniques used to analyse cloud data. A large-eddy simulation (LES) model has been used to generate convective cloud fields. Analysis of the LES output has been compared to other LES studies and recent analysis of satellite images of cloud fields. The ability of models to realistically simulate convective cloud fields is discussed in Section 3. A brief description of the model used in this study is given in Section 4. Results are given in Section 5 and discussed in Section 6. Finally, conclusions and suggestions for future work will be given in Section 7.

2. The distribution of cumulus cloud sizes

2.1. Introduction

Arakawa and Schubert (1974) were amongst the first to take the size distribution of cumulus clouds into account in their parameterisation scheme. As discussed previously, current GCMs lack an adequate representation of the spatial structure of convective cloud fields, limiting their usefulness in tackling issues as important as climate change.

Early studies involved taking measurements by flying over cloudy regions (e.g. Plank 1969). More recently the increased availability of remote sensing data has prompted research using satellite images (e.g. Sengupta et al. 1990) and radar, and studies involving LES or CRMs are now frequent, facilitated by advancements in computing technology.

The aim of many of these studies has been to find a functional relation for the cloud size density, defined as the probability density function of the number of clouds as a function of size (Neggers et al. 2003a). The total number of clouds, *N*, present in a domain can be found by integrating the corresponding cloud size density, i.e.

$$N = \int_0^\infty n(l)dl \quad (2.1)$$

where *n(l)* is the number of clouds in the domain with size *l*.

An exponential distribution, lognormal distribution, and distributions described by a power law or double power law with scale break have all been proposed.

A brief description of some common aspects of many of the studies (the use of effective diameter and threshold techniques) follows. Each proposed distribution will then be discussed in turn, namely:

- Exponential
- Lognormal
- Single power law
- Double power law (with scale break)

2.1.1 Effective Diameter

Effective or equivalent diameter has been used as a measure of cloud size in the majority of these studies. This is simply the diameter of the circle with the same area as the cloud. In studies involving remotely sensed images the effective diameter D_{eff} is calculated using the following formula:

$$D_{eff} = 2n^{1/2}\pi^{-1/2}p^{1/2} \qquad (2.2)$$

where n is the number of pixels identified as being 'cloudy' and p is the area of one pixel. Occasionally, the factor $2\pi^{-1/2}$ is omitted (such that the length of an equivalent square is calculated) and in some studies effective radius is used.

2.1.2 Cloudy pixel identification and computation of cloud sizes in satellite images

In the recent studies involving satellite images (e.g. Figure 2) a technique involving a radiation threshold is used to identify pixels as being either ‘cloudy’ or ‘non-cloudy’. The technique is essentially the same for each study and is described in detail in, for example, Machado and Rossow (1993) and Wielicki and Welch (1986).

The amount of reflected radiation that can be detected by the remote sensing instrument is divided into a number of digital counts. The number of counts associated with each pixel of the image is known. Since the clouds reflect more radiation than the background, each pixel with a count number greater than some threshold value is flagged as ‘cloudy’.

Kuo et al. (1993) give an explanation of *segmentation*, the technique used to determine the sizes of individual clouds. Clouds can be segmented according to four-connected or eight-connected methods. The four-connected method considers cloudy pixels immediately neighbouring another cloudy pixel, excluding diagonally adjacent pixels, to be part of the same cloud (Figure 3). In eight-connected segmentation, diagonally neighbouring cloudy pixels are also part of the cloud.

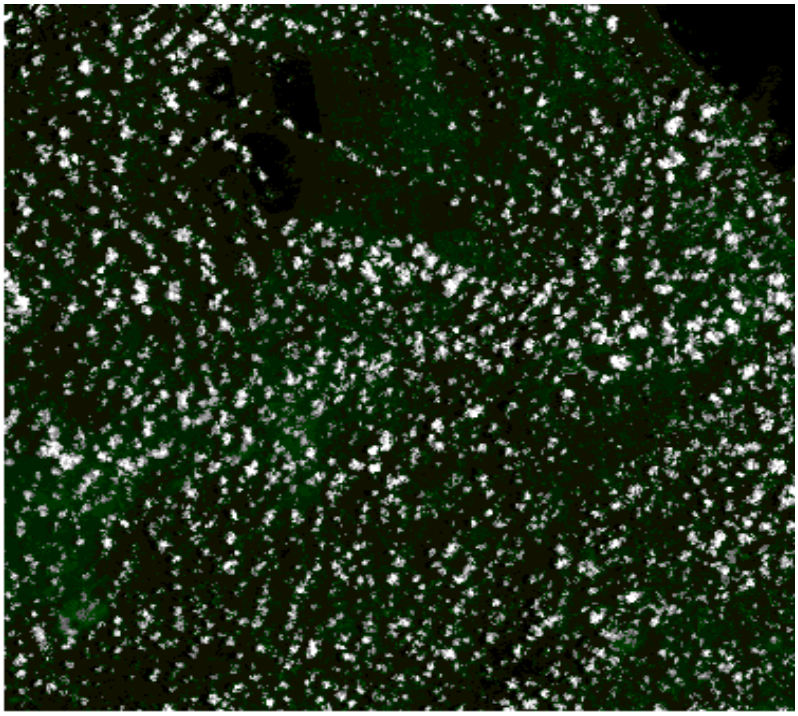


Figure 2: An example of a Landsat satellite image of a Cu field, processed using a radiation threshold and segmentation technique. The image was taken over Florida in August 1995. (Rodts et al. 2003)

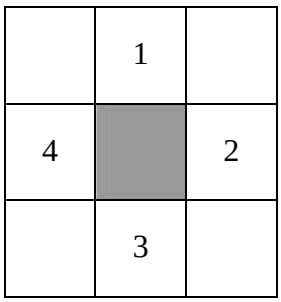


Figure 3: The four-connected segmentation technique considers four grid-squares surrounding a square identified as being cloudy

2.2 Exponential distribution

Plank (1969) was amongst the first to conduct a cumulus cloud size distribution study. His motivation, prior to the widespread use of computer models, was simply to contribute to a better understanding of the cumulus convection processes that lead to rain showers.

Plank used photographs of cumulus cloud fields taken from aircraft over Florida in the summer of 1957. Four cameras were mounted on the aircraft, arranged to obtain extensive coverage of the cumulus regions. A bounded region was marked on the aerial images, typically about 2300 km^2 (Figure 4). Cloud field images were printed so that the contrast resolution of the cloud subjects was maximised. They were placed on a light table to assist in the identification of the cloudy regions with the lowest albedo. A template with concentric circles marked on it (the diameters of which corresponded to the boundaries of the classes into which the clouds were to be sorted) was applied to each cloud in turn to establish its equivalent diameter. After the clouds had been sorted into half-mile wide classes, histograms were plotted to enable any relationship between the cloud number density *n* (i.e. the number of clouds within each class per $100km^2$) and the equivalent diameter D. These histograms revealed that for almost all sample fields, the number density of the cumuli decreased approximately exponentially with increasing effective diameter. In other words,

$$n = ae^{-bD} \qquad (2.3)$$

where $d < D < D_m$ and d and D_m are the minimum and maximum cloud sizes observed, respectively.

This relation appeared to hold well for the early morning cloud fields, however it was remarked that it failed to describe the cloud populations of the afternoon. One reason for this could be the somewhat primitive nature of the cloud size measurements. The technique is reasonably accurate for the smaller, almost circular clouds (≤ 1 mile in

diameter) however larger and more irregular clouds are clearly difficult to measure with a circular template, and their sizes were estimated using alternative methods. In addition, the measurements were limited by the resolution of the image and to a degree by the eye of the measurer when determining the horizontal extent of the clouds; each cloud was not a unique and discrete entity. These factors contributed to large uncertainties in the results of this study.

Figure 4: Example of cumulus cloud field investigated by Plank (1969). The area marked is approximately 2300 km^2

A similar study was carried out by Hozumi et al. (1982), using photographs of cloud fields taken from commercial aircraft over the East China Sea and Pacific Ocean. These photographs were taken in an arguably more crude way than Plank, using a camera attached to a window seat on the aircraft. The camera was dipped at an angle of between 8° and 10°. The height and speed of the flight was known from the aircraft instruments. Using this information an area of reasonably accurately known size could be marked on the images.

Since the horizontally projected area was required, a large dip angle was desired for increased accuracy. As such, the regions investigated were restricted to a band 20km wide of the flight path. Ideally, however, the photographs would be taken from directly above the cloud fields, to avoid problems associated with perspective. The same definition of equivalent diameter described by Plank was used. Thirty-seven distributions of clouds were analysed and corresponding histograms plotted, in a manner again similar to that of Plank. As in Plank's study, the maximum diameter of cloud observed was relatively small (around 5 km). Hozumi et al. concluded that the cloud size distribution follows the exponential function described in Equation 1. The coefficient *a* relates to the number of clouds in the cloud field and so is different to that found by Plank. The constant *b* is also different to that determined by Plank. This is likely to be at least partly because of the large uncertainties in both results, due to the relative crudeness of measuring techniques used. The physical significance of *b* will be examined later in this study (see Section 6). These results cannot be applied to larger cumulus clouds (upwards of 5 km) since these were not observed in large enough numbers.

Although there is a large amount of uncertainty associated with these results, an exponential decrease in cloud number with increasing effective diameter has been found subsequently by Wielicki and Welch (1986) using satellite images taken over the United States and Gulf of Mexico in 1979.

2.3 Lognormal distribution

A lognormal distribution[1] has been suggested for the size distribution of cumulus clouds. LeMone and Zipser (1980) investigated convection over the tropical eastern Atlantic and found that diameter, average vertical velocity and mass flux are all approximately lognormally distributed. Lopez (1977) had previously found that frequency distributions of height and maximum horizontal area attained by radar echoes of tropical disturbances were lognormally distributed. This led him to question whether lognormality was a general characteristic of cumulus convection. Using the equivalent diameter definition previously discussed and data from other studies, Lopez tested his hypothesis that cumulus cloud sizes were lognormally distributed. He found that for a number of different regions and convective situations it could not be rejected at the 5% level of significance. This does not, of course, prove that his hypothesis is correct. It was suggested that the lognormality could be because of the growth process of clouds through mixing with environmental air, or merging of smaller cloud elements. However, Lopez was looking for lognormality and did not consider other possibilities in his study.

[1] A variable is lognormally distributed when its logarithm is normally distributed.

2.4 Single power law distribution

Machado and Rossow (1993) were interested in deep convection, which is central in the exchanges of radiative energy and latent heat in the tropics. One of the main purposes of their study was to examine the size distribution of cumulus clouds in the tropics. Their data came from the International Satellite Cloud Climatology Project (ISCCP), which utilised a number of geostationary satellites (e.g. Meteosat). Cloud pixels were identified using a thresholding technique.

Since they were interested in high level convective systems a thermal infrared (TIR) threshold of 245K was employed (roughly corresponding to a height of 9km). The channel used is primarily sensitive to radiation from the earth's surface and clouds, with relatively little atmospheric absorption. A segmentation technique (eight-connected) was used to determine cloud sizes, given in terms of an equivalent radius. In fact, the study investigated clusters of individual convective cells.

It was found that *n*, the convective cluster number within a range of radius values Δr, is proportional to the radius of the lower class boundary raised to a power, such that

$$n(\Delta r) = n_0 r^{-\alpha} \qquad (2.4)$$

where n_0 is a constant and $\alpha \approx 2$.

The relation was observed to be approximately true for cloud clusters of radius up to around 50km, and only slightly affected by the choice of TIR threshold. All clusters investigated were more than 15km in radius, larger than the largest clouds studied by Plank, Hozumi et al. and Lopez and making direct comparison of results difficult. In addition, there is a large uncertainty in the areas of cloud clusters measured. Actual satellite image pixels were 5km x 5km, however uncertainties in navigation limit accuracy to about 15-30km leading to difficulties in the measurement of smaller

clusters. This does not cause many problems in this study, because of the large size of clusters being investigated.

2.5 Double power law distribution

Most recent studies have used satellite images to deduce that the distribution of cumulus cloud sizes can best be described by a double power law. This involves a change in power law exponent α (Equation 2.4) at a certain size of cloud.

One of the first of these studies was carried out by Cahalan and Joseph (1989). They were interested in the self-similarity or fractal nature of clouds. Data was sourced from the Landsat thematic mapper (TM) and multispectral scanner (MSS) instruments. A thresholding technique was used to identify cloudy and non-cloudy pixels, producing a binary image. The area of each discrete cloud was calculated and probability distributions of cloud areas were produced. For convenience, the logarithm of the square root of the area was plotted against the number of clouds within each bin of the histograms. The power law exponents found were different for various cloud types.

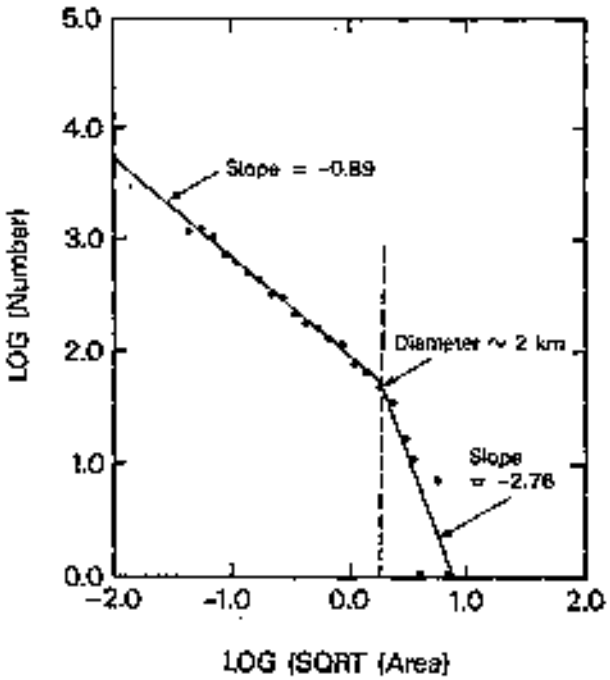


Figure 5: Histogram of cloud base areas for a fair weather cumulus cloud field off the coast of South Carolina. The logarithm of the number against the logarithm of the effective diameter has been plotted. Two best-fit lines are shown, with a distinct break clearly visible.

Figure 5 shows how two straight lines have been fit to the data, with a 'scale break' at approximately 2km. It was suggested that this scale break represents the largest individual convective cells, and that any cloud larger than this break size is actually made up of a number of smaller clouds. Neggers et al. (2003a) also suggest a possible reason for such a break, finding some connection to the height of the subcloud-layer and the intensity of the vertical wind shear. The physical explanation for the break remains the subject of ongoing research. Once again an amount of uncertainty has been introduced because of the resolution of the satellite image and the difficulties in identifying 'cloudy' and 'non-cloudy' pixels when they are in fact only partially cloudy.

Sengupta et al. (1990) conducted a study into cumulus cloud field structure because of its importance in the earth's radiation budget and association with the mesoscale environment. They noted that cloud inhomogeneities and spatial patterns are important variables when looking at radiative fluxes. Landsat MSS data was used once again, with a wavelength band that reduces Rayleigh scattering at shorter wavelengths and avoids water vapour absorption at longer wavelengths. Ten different cumulus cloud scenes were investigated. Some contained small fair-weather cumulus and some contained larger cumulus clouds. The images taken by the satellite are from between 0930 and 1000 local standard time (LST) and are in effect a snapshot of the cloud field at a particular time in its development. Clouds are grouped into cloud size classes according to effective diameter, and each cloud size boundary progressively increases by a factor of 1.5, so that as the number of larger clouds reduces, there are still a large number of clouds in each bin (for statistical reliability).

The slope of the fitted line for clouds with effective diameter less than 0.5km is markedly different from the slope for larger clouds i.e. there is double power behaviour again. In fact, it was found that the power law slope α appeared to be dependent on the maturity of the cloud field. Fair-weather cumulus cloud fields have the smallest slopes for effective diameters less than 1km and largest for greater sizes. Fields of larger clouds have slopes somewhere between. It has been suggested that

the larger clouds may be modifying their environment, and in doing so making the growth of further larger clouds easier. This modification of the environment has also been suggested as a possible explanation for the break in the power law.

The steepening of the slope at some cloud size is consistent with the results of Cahalan and Joseph (1989), Kuo et al. (1993), Nair et al. (1998) and Benner and Curry (1998).

Kuo et al. (1993) investigated whether their double power law results were dependent on whether a four-connected or eight-connected segmentation procedure was implemented, and found that they were not. They discarded clouds less than four pixels in size and as such only considered clouds greater than approximately 0.26km in diameter. During the same study the distribution of convective cells was investigated, and interestingly an exponential distribution was found to be most representative. Cells were identified by splitting individual clouds using a progressively decreasing temperature threshold. According to their definition, the difference between the maximum and minimum temperatures within a cell must be greater than 1.5°C. Cells were found to have sizes in the range 0.5km to 3km, similar to the sizes of clouds investigated by Plank (1969) and Hozumi et al. (1982).

Nair et al. (1998) were interested in the nature of regularity in cumulus cloud fields. Their data was obtained from the Landsat MSS, advanced very high-resolution radiometer (AVHRR) and Geostationary Operational Environment Satellite (GOES) and a comparison of the respective results was made. GOES and AVHRR images have a lower resolution than MSS, only capable of resolving clouds greater than 1.1km in diameter. Whilst cloud fields analysed using data from all instruments were found to be best represented by a double power law, the break in scale occurred at larger clouds for GOES and AVHRR images (1 – 1.5 km) than Landsat MSS (500 – 700m). It was observed that there appeared to be some dependence of the power law exponent on whether the clouds were organised in some way or apparently random.

Benner and Curry (1998) used Moderate Resolution Imaging Spectroradiaometer (MODIS) Airbourne Simulator images together with photographs taken from the space shuttle using a high-resolution hand-held camera. The camera was capable of capturing images comparable in resolution to that of the Landsat TM (approximately 30m). Two thresholds were applied to the satellite image pixels in this study, one for the visible channel and one for infrared (IR). Cloudy pixels were flagged when the radiance was higher than the visible threshold, and the IR radiance lower than the IR threshold. The thresholds were selected to simulate the underlying ocean surface. Clouds are more reflective and colder than this surface. The space shuttle images were digitised using a high-resolution scanner. Each image was masked with a threshold to produce a binary cloud/no-cloud image. This process involves a somewhat subjective use of graphical software.

A four-connected segmentation technique was used for all images in this study. Cloud sizes were computed and the clouds sorted into 100m wide class widths, based on effective diameter. The percentage of clouds in each size class was plotted against diameter to determine whether a power law was applicable. For completeness, an exponential distribution was also considered using the data. It was found that a double power law could best describe 17 of the 22 MODIS images, with a break at around 900m. The break for the space shuttle cloud field distributions occurred at approximately 600m. None of the distributions examined could be best represented by an exponential distribution.

Neggers et al. (2003a) have also suggested a power law function for the distribution of cumulus cloud sizes using data from large eddy simulations (LES) rather than satellites. The use of LES in studies of clouds and their properties will be discussed in more detail in Section 3. The study compares the cloud size densities produced by LES to those of observed natural cloud fields and finds that the simulated cloud populations are realistic. It is shown that the cloud size density is well represented by a power law with negative exponent. However, there appears to be a collapse in this power relation at a certain cloud size (around 1km) i.e. there is a distinct scale break.

2.6 Power laws and the fractal nature of clouds

Power law behaviour has been noted by many (e.g. Nair et al. 1998) to be evidence of the self-similarity, or fractal nature, of clouds. This self-similarity must break down towards smaller scales, since the microscopic cloud structure differs from the macroscopic structure of individual clouds. There has been much research into the scale at which this breakdown occurs (e.g. Cahalan and Joseph 1989). Recently it has been suggested that the incorporation of fractal theories into cloud parameterisations is unreliable, because of errors introduced by, for example, the measuring instruments and averaging techniques used in fractal studies (Brewer and Di Girolamo 2004).

Functional form	Paper	Data	Cloud size range	Exponent	
Exponential	Plank (1969)	Photographs from aircraft	70m - 5km	0.4 - 14	(km^{-1})
	Hozumi et al. (1982)	Photographs from aircraft	200m - 8km	0.8 - 4	(km^{-1})
	Wielicki and Welch (1986)	Satellite			
Lognormal	Lopez (1977)	Radar	1km - 50 km		
	LeMone and Zipser (1980)	Aircraft penetrations	< 4km		
Single power law	Machado and Rossow (1993)	Satellite	Clusters	2	
Double power law	Cahalan and Joseph (1989)	Satellite	< 10km	1.89	< 2km
				3.76	> 2km
	Sengupta et al. (1990)	Satellite	< 5km	1.39 - 2.35	< 0.5km
				2.1 - 4.75	> 0.5km
	Kuo et al. (1993)	Satellite	2 - 11km	1.7	< 1km
				3.4	> 1km
	Nair et al. (1998)	Satellite	< 1.5km	0.9 - 1.65	< 700m
				3.76 - 6	> 700m
	Benner and Curry (1998)	Radar	< 5km	0.6 - 2.3	< 0.3 - 1.7km
		Photographs from space shuttle		2.4 -4.6	> 0.3 - 1.7 km
	Neggers (2003a)	LES	< 3km	1.7	< 700 m

Table 1: Summary of suggested functional forms

2.7 Summary

The double power law with distinct break at a certain cloud size is the most popular functional form proposed recently (see Table 1). However, there is still a great deal of confusion. An exponential distribution has been suggested following studies of both satellite data and photographs from aircraft. Exponential distributions of cloud area and mass flux have also been found in LES output (Cohen 2001). Most recent studies however have found no evidence to support an exponential cumulus cloud size distribution. The lognormal distribution has not been proposed for some time, however the results of LeMone and Zipser (1980) have been used in more recent studies (e.g. Donner 1993). The single power law relation has only been suggested once and following a study of cloud *clusters* rather than individual clouds.

The results of Cahalan and Joseph (1989) (Figure 5) clearly indicate a probable double power law distribution. Some interesting questions arise when their data is plotted on a log-linear plot (Figure 6). An exponential distribution on this type of plot would be characterised by a straight line. Limiting attention to clouds larger than ~ 0.4km, it could be argued that a reasonably straight line could be fit to the data. Could it be that studies that have concluded that cumulus clouds are exponentially distributed did so because the cloud sizes investigated fell to one side of a scale break, and that the correct function describing the distribution is in fact a double power law?

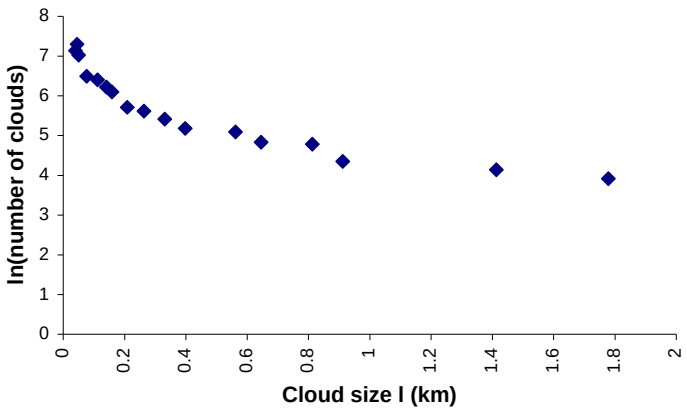


Figure 6: Cahalan and Joseph (1989) data on a log-linear plot

Most advocates of the double power law have determined the break to lie around 1km (Table 1). Plank (1969) and Hozumi et al. (1982) investigated ranges of cloud sizes that included the proposed break diameters, however uncertainty in their methods has already left their results questionable. Nevertheless, subsequent studies supporting the exponential distribution should consider whether the range of cloud sizes investigated has influenced the result. A similar theory could explain why Machado and Rossow (1993), when looking at large cloud clusters, found only a single power law functional form.

A perfectly exponential distribution of clouds with sizes in the range studied by Cahalan and Joseph (1989), plotted on log-log axes, also raises some interesting questions (Figure 7). Whilst clearly a curve, it seems possible that given only a selection of the data points, two straight lines (implying a double power law) could be perceptible. Could it be that the double power law is actually a consequence of limited data sets, and that the true exponential curve on the log-log has not been revealed? A glance at Figure 5 casts doubt on this theory. Two straight lines seem more likely than a curve, and the straight line for clouds smaller than 2km in diameter appears to be a particularly good fit. Also, the number of studies supporting the double power law and the range of scenarios and cloud sizes analysed, together with some quite plausible suggestions for the scale break, suggest that the double power law is probably the true functional form describing the distribution of cumulus cloud sizes.

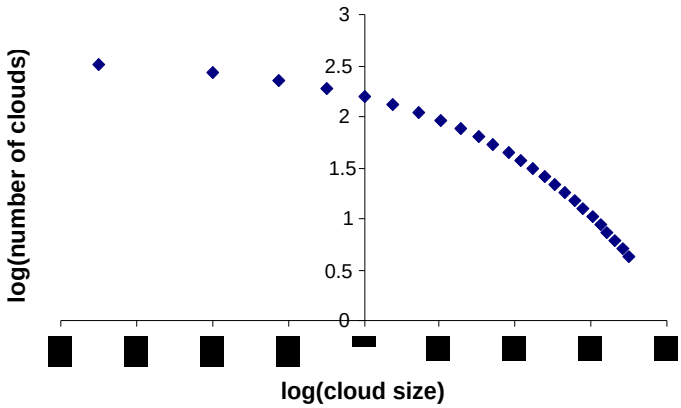


Figure 7: log-log plot of an exponential distribution of cloud sizes similar to those studied by Cahalan and Joseph (1989)

However, whilst the double power law relation appears most likely, there is confusion between those who advocate it. A wide range of power law exponents has been suggested, and there is no agreement on the characteristic cloud size at which the break in power law occurs. The double power law, with one exception, has been suggested following analysis of satellite data. There are uncertainties inherent in the techniques used, mainly stemming from the resolution of the satellite images. Small clouds may not be detected since they reflect less radiation. Also reflectance within individual clouds can vary significantly (Wielicki and Welch 1986). As such the threshold selected may lead to the area of some clouds being underestimated. Identification of cloudy pixels can be difficult because some surfaces have albedos similar to clouds, however the majority of satellite images used in the studies were taken over the ocean where this is less of a problem.

2.8 Objectives of this study

The double power law with scale break has been supported by Neggers et al. (2003a) using LES output. The data have been analysed using techniques used to analyse satellite images. The objectives of this study are:

1. To use the data from a large-eddy model (LEM) in cloud resolving model (CRM) mode to investigate the cumulus cloud fields produced at a range of heights, and deduce whether the cloud size distribution can be represented by a mathematical function.

2. To determine whether some of the confusion and disagreement surrounding a universal functional form for the cumulus cloud size distribution arises because of differences in analysis techniques. CRM output will be used to produce a binary cloudy/non-cloudy image similar to that produced when satellite data is analysed. Cloud sizes will be determined using a segmentation technique employed in observational studies involving satellite images, and size distributions investigated. Results and the analysis procedure will be compared to Neggers (2003a) and other studies. This will hopefully give some insight into whether the apparent functional form is related in some way to the data processing method implemented.

3. To determine (by comparison of results) whether the simulated satellite distribution is representative of the distributions obtained at particular heights, or whether the distribution looks different.

3. Use of LES in cloud studies

LES models are used to investigate convective processes by simulating full three-dimensional fields of thermodynamic variables and vertical and horizontal momentum. They are able to produce large datasets at resolutions that may be difficult to realise in field experiments with real clouds (Neggers et al. 2003a) and are particularly useful when there is difficulty in obtaining observations. Also, the user of the LES defines and controls all conditions of the simulation and as such it can be used to reproduce observed conditions.

In an LES model the large (resolved) scale motions are calculated explicitly, while the small (sub-grid) scale motions must be parameterised. LES models involve the solution of prognostic equations at each point in a computational domain. The usefulness of LES models in the study of clouds and their characteristics is of course dependent on their ability to realistically simulate cloud fields. A number of studies have been conducted to test whether they are capable of doing this. A lack of appropriate in-cloud measurements has left a number of important LES results unverified however most current LES models agree on the basic structure of shallow cumulus layers (Neggers et al. 2003b).

Neggers et al. (2003a) compared LES results directly with actual in-cloud measurements made during the small cumulus microphysics study (SCMS) in Florida in 1995 and other studies. Since the development of the LES in time is required to be as close as possible to the actual cloud field development, available observations (from surface instruments and radiosondes) are used to construct the LES case (through their use as initial conditions). A conditional sampling technique is used to compare vertical profiles of properties such as temperature, obtained from the LES, with actual measurements. Neggers et al. showed that the LES predicted properties of temperature, moisture and liquid-water content were in close agreement with those observed. Also, the vertical component of the in-cloud turbulent kinetic energy and the cloud size distribution matched the observations. In a separate paper, Neggers et

al. (2003b) compared LES results to those using high-resolution satellite data. To enable a straightforward comparison, exactly the same method was used to derive the cloud size densities and a comparable number of clouds were investigated. They showed that LES was capable of reproducing the break in power law found to describe the size distribution of cumulus clouds, and calculated a similar power law exponent for the smallest clouds observed (up to approximately 800m in effective diameter). At larger cloud sizes the rate of reduction in cloud size density with increasing size is greater in the LES case than suggested by the satellite data. This is probably because the domain size of the LES restricts the growth of largest clouds. The time delay between observations and satellite image might also explain some of the small discrepancies between the results. Overall, the study shows that LES is capable of reproducing some significant results.

Siebesma and Jonker (2000) showed that LES models were capable of reproducing the morphology of individual cumulus clouds. Cuiypers and Duynkerke (1993) used an LES model to simulate the partly cloudy convective boundary layer and found that vertical profiles of variances and fluxes agreed with experimental data. Siebesma et al. (2003) compared the results of ten LES models to observations made during the Barbados Oceanographic and Meteorological Experiment (BOMEX), which investigated trade-wind cumulus clouds. They found that all of the models were able to reproduce the observed steady state, although there was some difficulty in simulating cloud cover. Cohen (2001) found that mean thermodynamic and convective profiles from LES data were in agreement with previous observational and numerical studies.

Although LES is subject to the constraints of numerical models, these example results show that they are able to produce realistic cumulus cloud fields, and support the credible use of LES in the development of cumulus cloud parameterisations. Future research on sensitivity to resolution and impact of sub-grid scale parameterisations will hopefully lead to increased confidence in their use.

4. The model used in this study

4.1 Introduction

Cohen (2001) noted that a total amount of mass flux, *M*, distributed randomly amongst *N* clouds, implies an exponential distribution of mass fluxes. This is analogous to the Boltzman distribution of molecular energies. It was investigated whether this theoretical distribution was simulated by the Met Office large eddy model (LEM) in CRM mode. Data from the same model will be analysed in this study. A detailed description of the model is given in Cohen's thesis however a brief description is given below.

4.2 Description of model

The model explicitly resolves cloud-scale dynamics whilst also including a parameterisation of sub-grid scale turbulent eddy transports of heat, moisture and momentum and parameterisation of the main microphysical processes. The anelastic and Boussinesq approximations to the prognostic momentum and thermodynamic equations are solved for the domain of size 128x128x21km^3. The horizontal resolution is 2km. There are 50 vertical levels and a stretched grid is employed; the vertical resolution is 100m in the boundary layer and 500m in the upper troposphere.

4.3 Radiative-convective equilibrium

The CRM simulates the tropical atmosphere in an idealised state of radiative-convective equilibrium. Figure 8 illustrates how a constant, horizontally homogeneous cooling rate is applied to the atmosphere up to 400mb, from where it decreases linearly to zero at 200mb. This is a simplified version of the actual physical situation, ignoring all cloud-radiation feedbacks.

Equilibrium is reached when the net mass flux transported by convection causes exactly enough compensating subsidence to balance the heat loss by radiative cooling. This state is reached in around 10-20 days depending on the cooling rate. Horizontal snapshots of cloudy properties (e.g. vertical velocity) can be taken at a range of times within the equilibrium period.

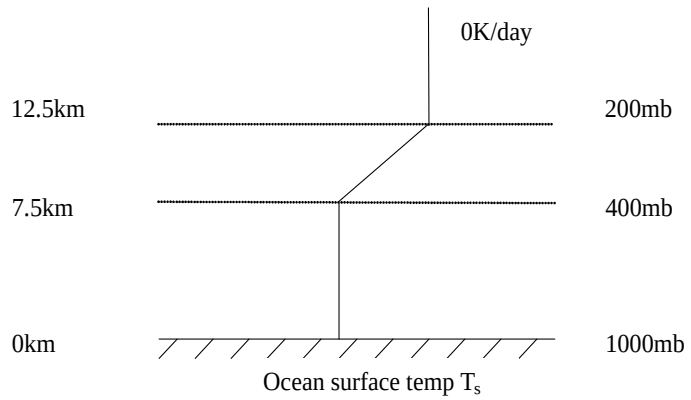


Figure 8: Schematic of the radiative cooling rate applied. Cooling rate of, for example, 4K/day at surface reducing to zero at 200mb

An ensemble of randomly distributed cumulus clouds is required to test the exponential distribution predicted by the theory in Section 4.1. A such, some further simplifications are employed to avoid possible organisation of the convection:

- no mean wind shear
- the atmosphere is assumed to be non-rotating, in contrast to the LEM used by Neggers et al. (2003a) and detailed in Cuiypers and Duynkerke (1993).

4.4 Test of data analysis procedure

Initially, data from the model was used to reproduce the exponential distribution of mass flux found by Cohen (2001) at a height of 2.4km (corresponding to the base of the deep convective layer). This was done as a useful check on the analysis procedures. Example histograms are shown in Figures 9 and 10 and the result is consistent with the distribution predicted by statistical mechanics.

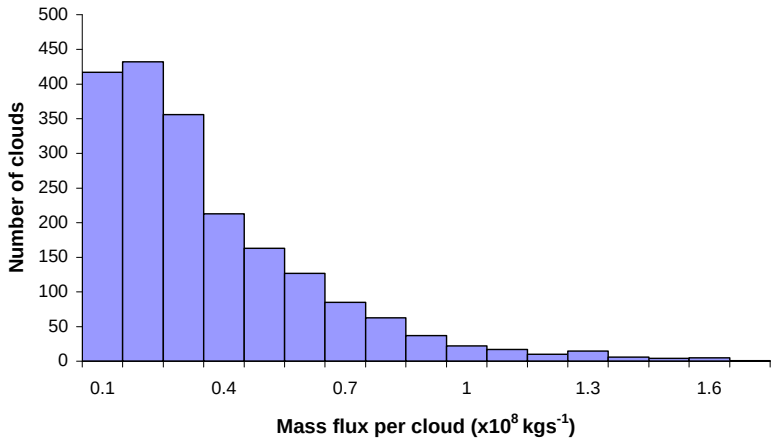


Figure 9: Histogram of mass flux per cloud at a height of 2.4km, using a 16K/day cooling rate in the CRM simulation. Clouds are identified where the vertical velocity, *w*, is greater than $1ms^{-1}$ (see Section 4.5) and this threshold has reduced the number of clouds in the first bin.

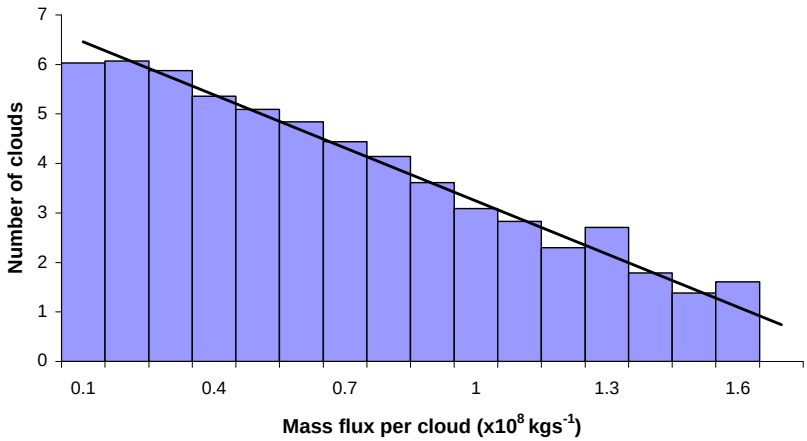


Figure 10: Same as Figure 9, but with a logarithmic scaling of the y-axis. A 'best fit' line has been added to show more clearly the exponential distribution.

The mass flux for each cloud at a vertical level is calculated using Equation 4.1:

$$m = \rho V_i A_i \qquad (4.1)$$

where ρ is the density at that particular vertical level,

V_i is the total vertical velocity for the cloud

(i.e. the sum of the velocities of each grid-square contributing to the cloud)

A_i is the area of one grid-square

A routine scans through the model domain and identifies cloudy grid-points (see Section 4.5). Adjacent cloudy grid-points are assumed to be part of the same cloud. Note that clouds with negative mass fluxes have not been included, since these are assumed to represent convective cells nearing the end of their life cycles.

4.5 Definition of a cloudy grid-point

A number of definitions can be used to determine whether a grid-point is cloudy or non-cloudy. The definition used when calculating the mass fluxes used in the histograms of Figures 9 and 10 will be used for the remainder of this study.

Updraught and downdraught grid-points have vertical velocities $w>1\text{ms}^{-1}$ and $w<1\text{ms}^{-1}$ respectively. Cloudy grid-points are identified where there is an updraught i.e. where $w>1\text{ms}^{-1}$. This definition appears somewhat simplistic but has been used in previous studies (LeMone and Zipser 1980, Xu and Randall 2001). The definition may lead to the number of clouds in the domain being underestimated sinc e weaker updrafts that may correspond to cloudy grid -points are not considered. Sensitivity to the w threshold used will be discussed in Section 5.6.

Figure 11 is a snapshot of the vertical velocities for each grid-square in the horizontal domain at a height of 2.4km. The snapshot has been taken at some point after radiative-convective equilibrium has been reached. Figure 12 indicates which of the grid-squares are cloudy, based on the definition outlined previously. The clouds appear to be randomly distributed, as expected based on the model setup.

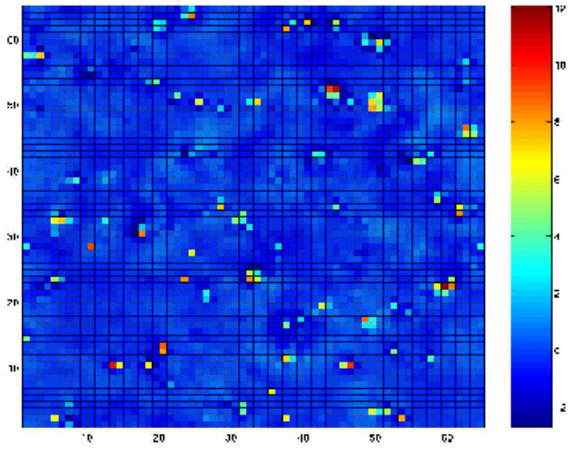


Figure 11: Snapshot of vertical velocity, w (ms^{-1}), field in horizontal domain for 16K cooling rate, at a height of 2.4km

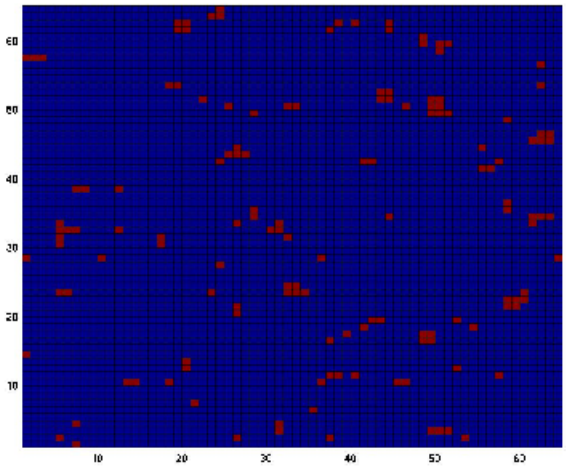


Figure 12: Cloudy/non-cloudy image based on data used to produce Figure 11 and w>1ms^{-1} cloudy definition. Cloudy grid-squares are red.

4.6 Calculation of cloud areas

The domain is scanned and cloudy grid-points identified. Adjacent cloudy points are assumed to be part of the same cloud. Each cloud is completed when all cloudy grid-points connected to the original have been located. The size of the cloud in grid-squares is then computed and stored. Each grid-point corresponds to a corner of one grid-square. The area of each cloud, A_i, is simply found by multiplying the number of grid-squares by the area of one grid-square (i.e. 4km^2).

The definition of cloud size used in this study is the same as that used by Neggers et al. (2003a):

$$l_i = \sqrt{A_i} \qquad (4.2)$$

where the linear size l is equivalent to the length of a square with the same area as the cloud.

During the analysis of simulated satellite images, A_i refers to the vertically projected area of the cloud (Figure 13), since we are looking at two -dimensional projected images taken from a high altitude.

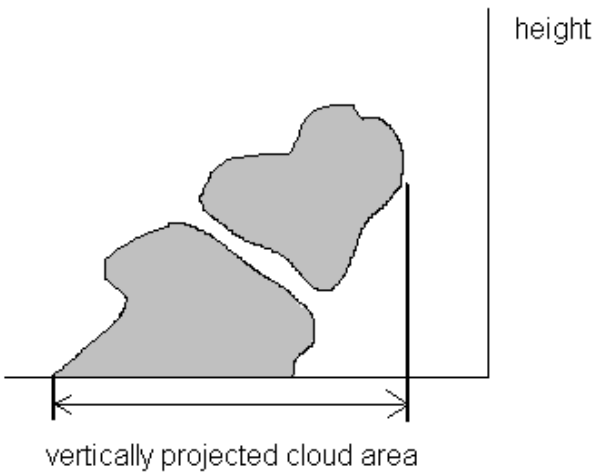


Figure 13: Schematic illustrating concept of vertically projected cloud area.

Once the linear size of each individual cloud has been calculated, the clouds are sorted into bins and histograms are plotted.

4.7 Analysis of simulated satellite image produced using model output

The CRM output has been analysed in the same way as the many recent studies involving satellite data (e.g. Sengupta et al. 1990). It is first necessary to produce a binary cloudy/non-cloudy image, similar to the images produced when a thresholding technique is applied to satellite data.

Vertical velocity values at each point of the three -dimensional domain are extracted from the CRM. An algorithm has been constructed to identify whether for each column of th e domain, there exists a cloudy grid -point (i.e. where $w>1ms^{-1}$). This cloudy point is assumed to be visible to a satellite looking down on the domain. When a column is found to contain a cloudy point, the corresponding grid -square of the two-dimensional domain equivalent to the area being viewed by the satellite is flagged as being cloudy.

An algorithm has been coded to calculate the sizes of the clouds in the 'satellite image'. An eight -segmentation method has been used as described in Figure 14. Example 'satellite images' are given later in this report (Section 5.4).

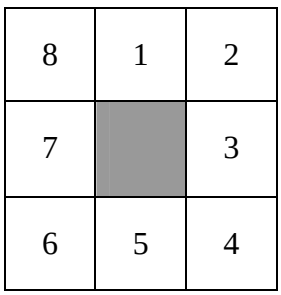


Figure 14: Eight-connected segmentation involves looking at the squares surrounding a cloudy grid-square (centre), including diagonally adjacent squares.

The clou d field image is scanned row by row, starting in the bottom left corner. Cloud effective diameter is calculated using Equation 4.2. The cloud sizes are stored and histograms plotted to determine the size distribution.

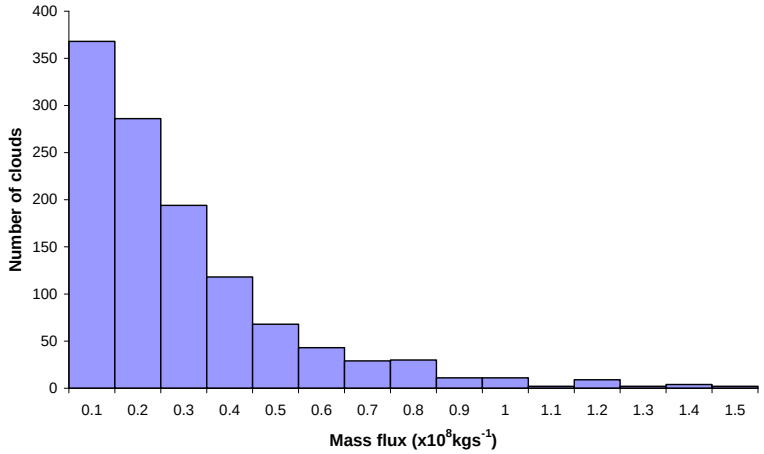
Boundary conditions have been appli ed such that opposite domain edges are considered to be adjacent. Cloudy grid -squares on one edge might therefore be connected to cloudy grid -squares on the opposite edge. This is to avoid the underestimation of cloud size s as a result of the boundaries and is consistent with the boundary conditions used in the CRM.

It is clear that clouds seen as unique entities by the satellite might actually consist of overlapping segments of clouds on different vertical levels. An actual satellite sees overlapping c louds and has difficulty distinguishing between parts of the cloud at different altitudes, seeing in effect just one cloud. (Although different wavelengths could be used to determine cloud heights, this has not been attempted in general in equivalent studies of satellite images.) The area of the apparently unique clouds is in fact the vertically projected area of the overlapping clouds at various heights (see Figure 13).

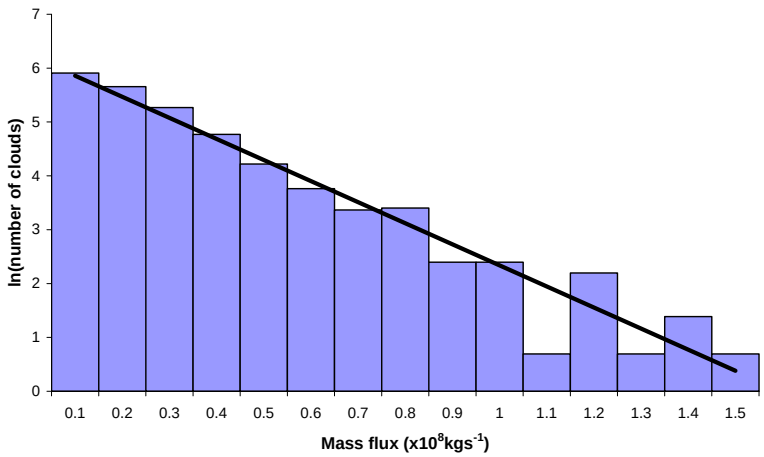
5. Results

5.1 Distribution of mass fluxes

It has been shown already that Cohen (2001) used the CRM to demonstrate that mass flux appears to be exponentially distributed at a height of 2.4km and with an imposed cooling rate of 16K/day. Additionally, approximately exponential distributions of mass flux have been found during this study for a number of other heights (e.g. Figure 15) and using a smaller cooling rate of 4K/day (e.g. Figure 16).

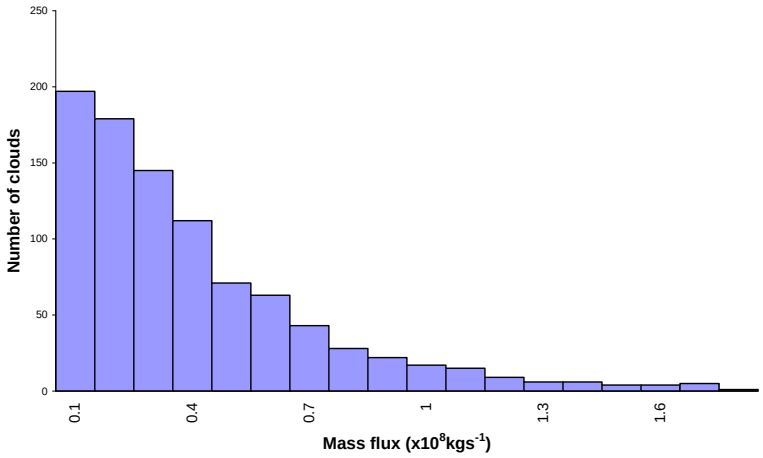


(a)

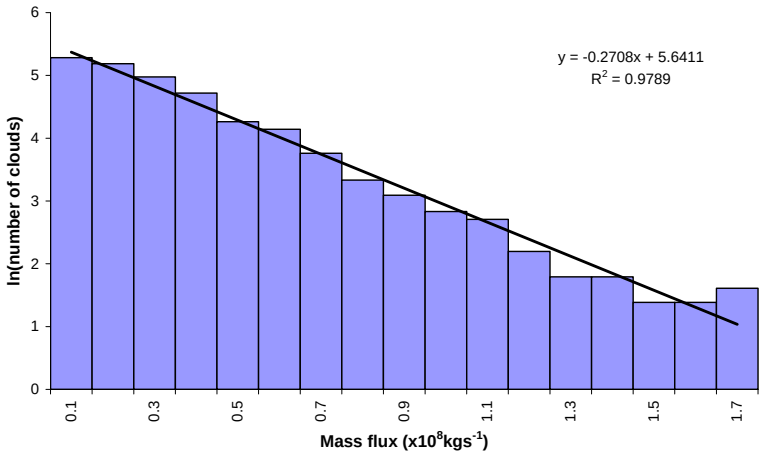


(b)

Figure 15: Distribution of mass flux per cloud at height of 3.9km and with an enforced cooling rate of 16K/day. (b) same as (a) but with logarithmic y-axis and best-fit line



(a)



(b)

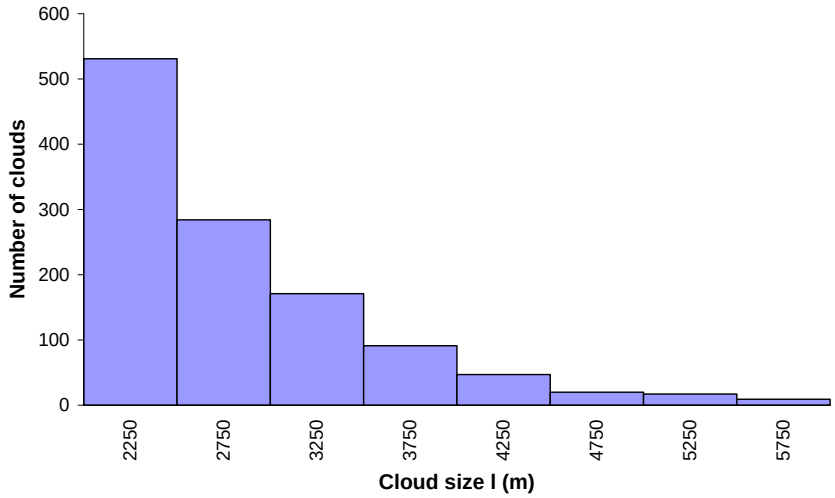
Figure 16: Distribution of mass flux per cloud at height of 5.3km and with an enforced cooling rate of 4K/day. (b) same as (a) but with logarithmic y-axis and best-fit line

5.2 Distribution of cloud sizes

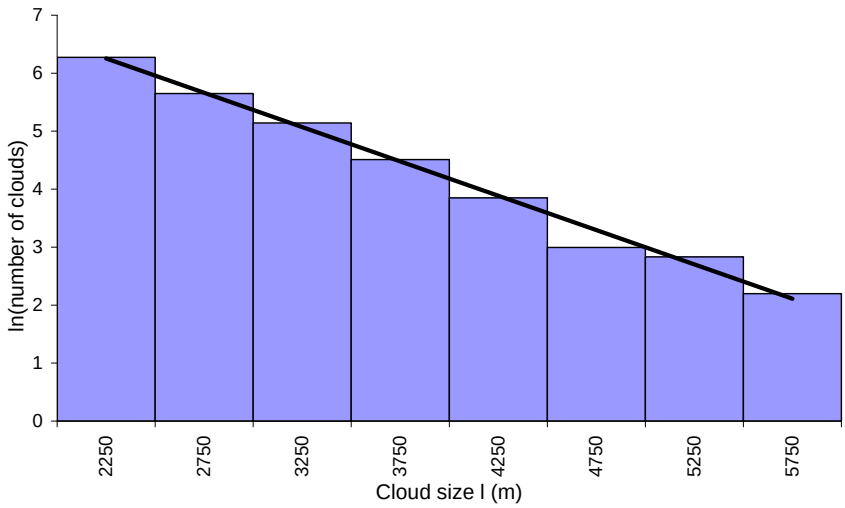
Cooling rates of 4K/day and 16K/day were used to investigate the size distribution of the cumulus cloud fields produced by the model. Cohen (2001) also showed that, using a cooling rate of 16K/day, cloud *areas* appear to be exponentially di stributed. The only height investigated was 2.4km however, and an exponential distribution of cloud areas does not imply that cloud diameters follow the same functional form. In fact, a perfectly exponential distribution of area would imply that diameter is *not* exponentially distributed. In this study, the length of an equivalent square (as used by Neggers et al. 2003a) is the measure of cloud size, rather than area.

The 16K/day rate is somewhat unrealistic, however it has been used because it produces a large number of clouds for statistical reliability. Clouds at a number of vertical levels were investigated. Once the system is in radiative -convective equilibrium the cloud fields can be sampled at a number of time intervals and size statistics saved to file. The samples are approximately 9 hours apart, such that there is no correlation between the convective cloud fields at successive times. The maximum number of files available was used to produce the histograms, to maximise the statistical reliability of the results. Although more clouds are produced for the greater forcing rate of 16K/day, there were more files available for the 4K run resulting in a greater number of cloud statistics.

Figures 17 and 18 show clearly that the cloud sizes at 2.4 km can reasonably be described by an exponential distribution (Equation 2.3), using the 16K/day and 4K/day cooling rates respectively. A plot of the logarithm of the cloud number against cloud size for such a distribution should result in a straight line, as seen in part (b) of the figures. This distribution was observed, at least approximately, at all vertical levels investigated. Figures 19 shows log -linear plots for 5.6km. There are fewer clouds at this height and so the statistical data is less reli able however an approximately exponential distribution can be seen.

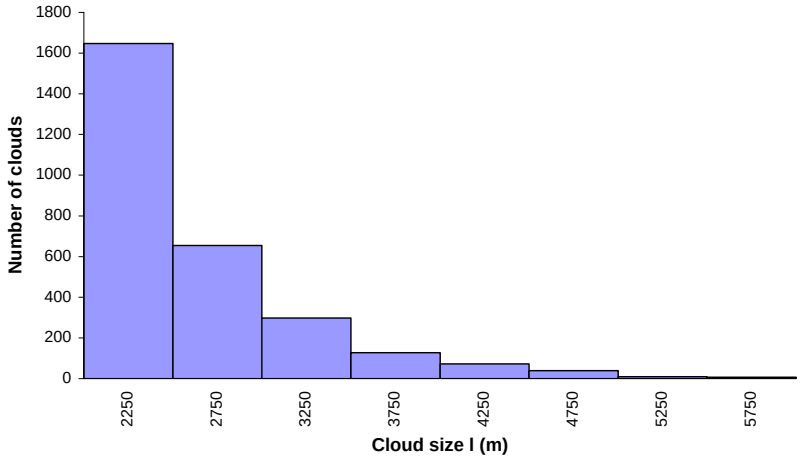


(a)

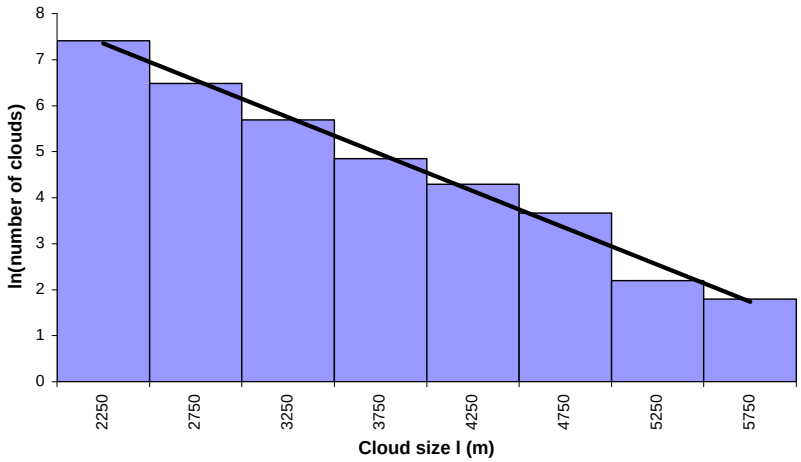


(b)

Figure 17: Histograms of cloud sizes at height of 2.4km for 16K/day cooling rate (b) same as (a) but with logarithmic y-axis

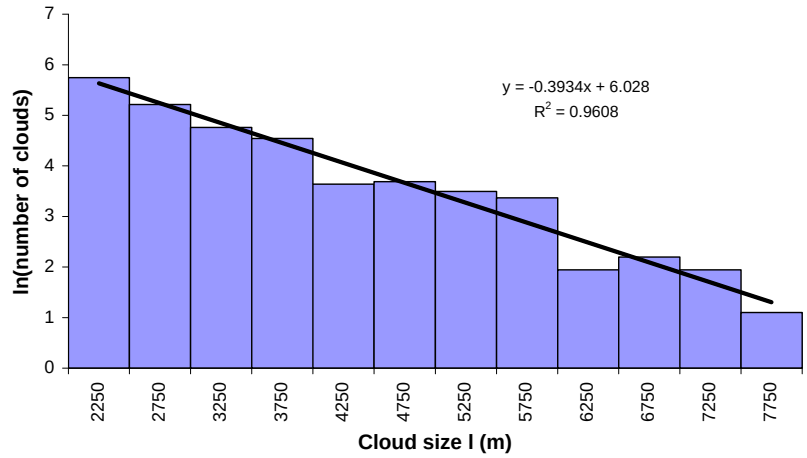


(a)

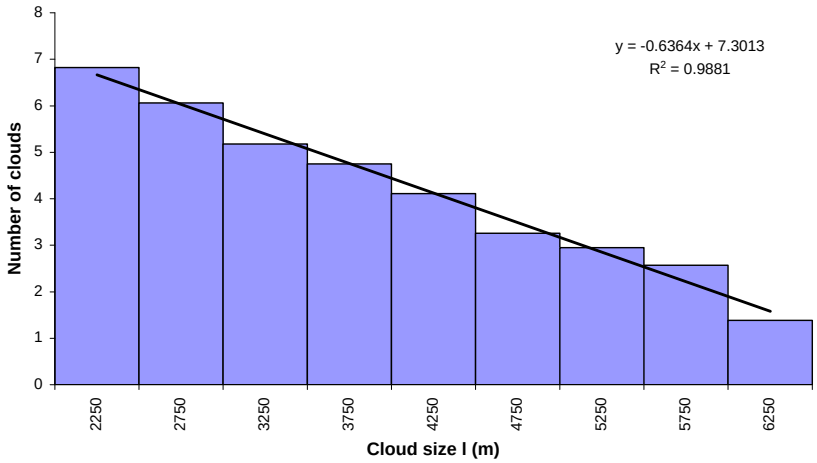


(b)

Figure 18: Same as Figure 17 but based on the 4K/day radiative cooling rate



(a)



(b)

Figure 19: Histograms for 5.6km. (a) based on 16K/day cooling rate and (b) based on 4K/day cooling rate. Units of x (in linear fit equation) is km.

5.3 Vertical profile of exponent

Histograms of cloud sizes have been produced for each of the model vertical levels. The data at each level has been fit to an exponential distribution and the exponent *b* (see Equation 2.3) calculated using linear regression. The number of clouds found at very low levels (less than approximately 1km) and very high levels (above 9-10km) is small, and so the statistics are unreliable. The R-squared value has been used as an indication of whether the statistics have produced a distribution of clouds that is close enough to being exponential for the value of *b* to be meaningful[2]. Somewhat arbitrarily, the value of *b* has not been recorded when the R-squared value is less than 0.95. Vertical profiles of *b* for the cooling rates of 4K/day and 16K/day are shown in Figures 20 and 21. Both plots indicate some oscillation in the value of *b* with height below approximately 3km, before a generally constant decrease to around 7-8km. In addition, the values of the constant in the 4K/day cooling case are significantly larger than in the 16K/day case. Possible explanations for this will be discussed in the Section 6.

[2] The R-squared value is a value between 0 and 1 and gives an indication of how well the 'best-fit' line fits the data. A value near 1 indicates a good fit to the data.

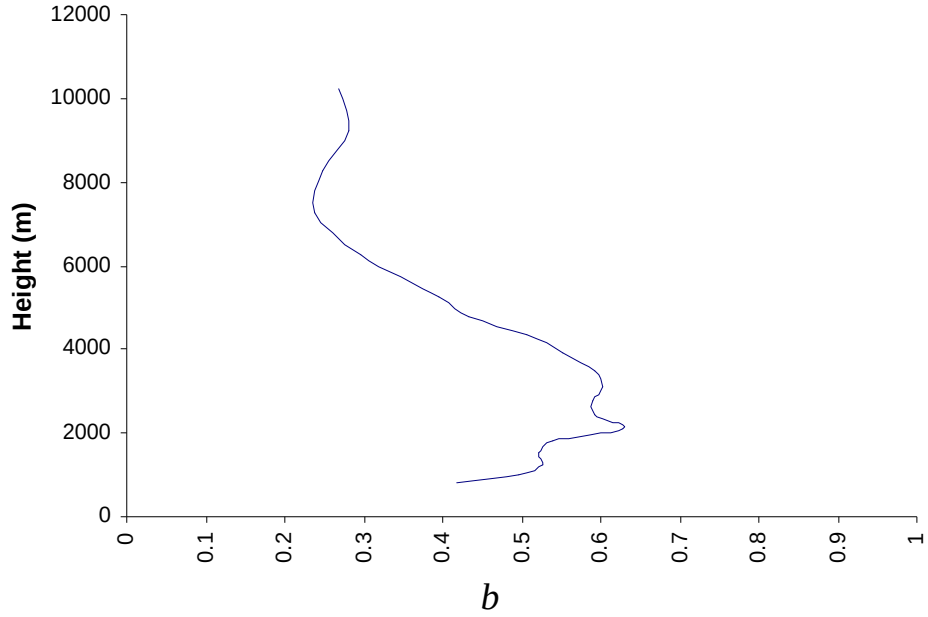


Figure 20: Vertical profile of the exponent *b* in Equation 2.3 (describing exponential distribution) for 16K/day cooling rate

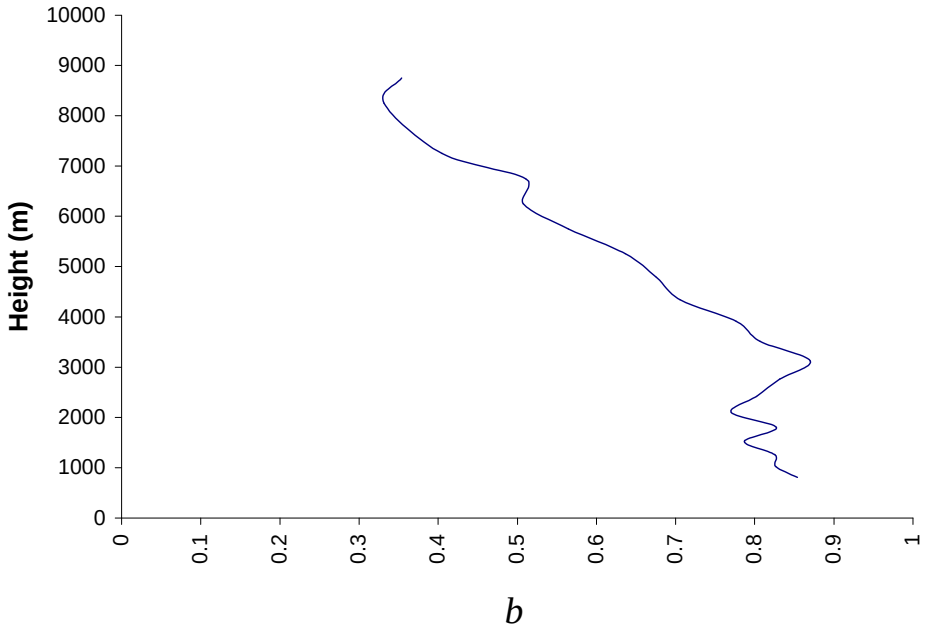


Figure 21: Vertical profile of the exponent *b* (km^{-1}) in Equation 2.3 (describing exponential distribution) for 4K/day cooling rate

5.4 Simulated satellite images

CRM output has been used to produce simulated versions of the binary cloudy/non-cloudy images produced following threshold analysis of satellite images. The method has been described in Section 4.7. Figure 22 was produced using the maximum vertical velocity in each column of the model. Figure 23 was produced by applying the w>1ms^{-1} cloudy grid -point definition to the data used to produce Figure 22. Cloudy grid-points in Figure 24 have been identified where w>0.8ms^{-1}. The resulting increase in the number of cloudy points has increased the size of some clouds. Sensitivity to the w threshold used will be discussed in Section 5.6. An example snapshot for the 4K/day radiative cooling rate is shown in Figure 25. There are clearly less clouds in this case because of the smaller applied forcing. Once again, a number of files each containing data relating to a particular time in the radiative - convective equilibrium period were available for use. All files wer e used for maximum statistical reliability.

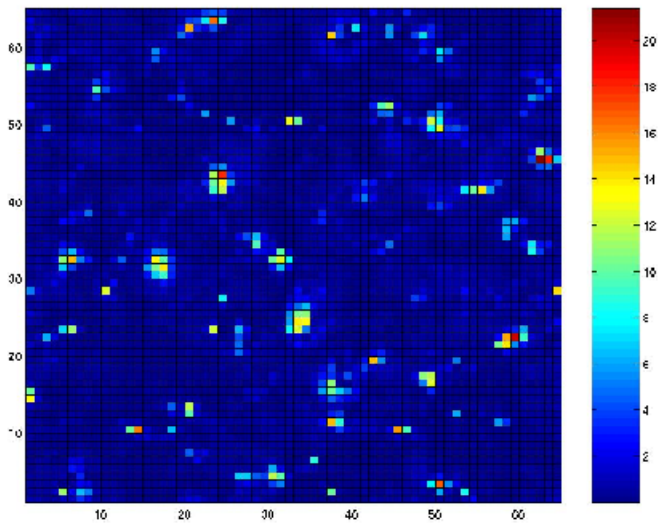


Figure 22: Horizontal domain of maximum vertical velocity (ms^{-1}) for all 50 vertical levels, produced using cooling rate of 16K/day

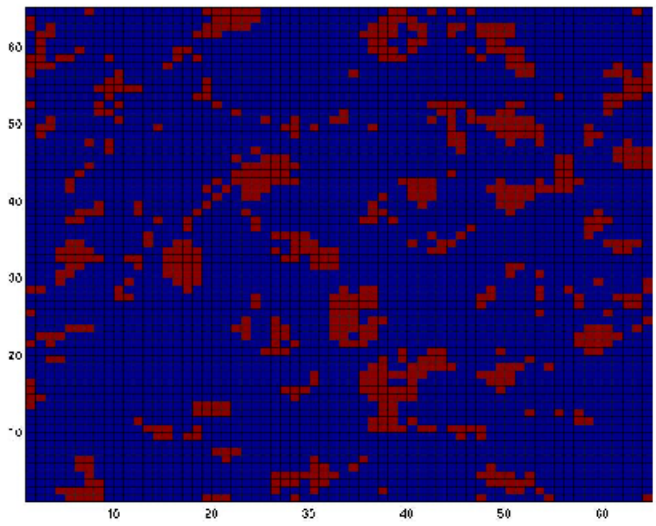


Figure 23: Simulated satellite image produced with cooling rate of 16K/day based on same data as Figure 22. Cloudy grid-square identified (and coloured red) where vertical velocity w>1ms^{-1}

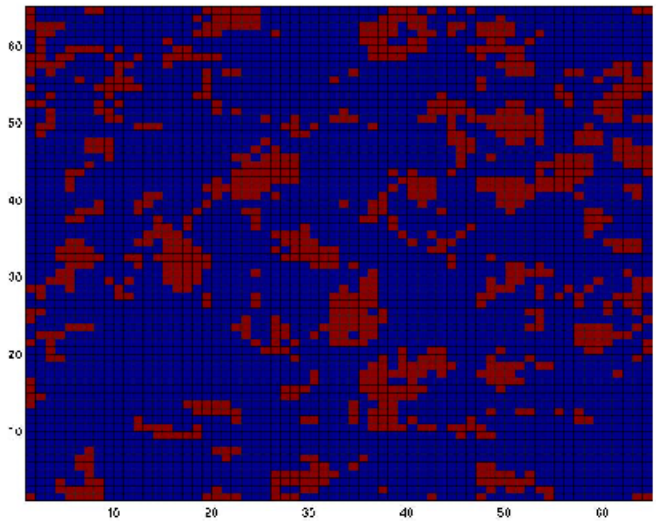


Figure 24: Same as Figure 23 but with cloudy grid-squares identified where w>0.8ms^{-1}

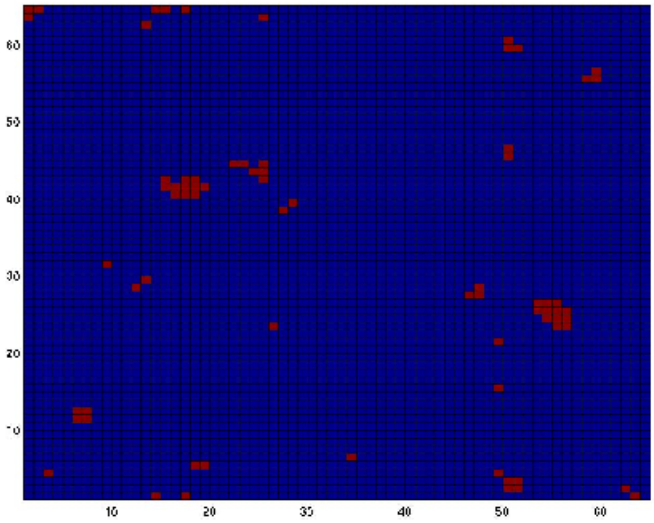
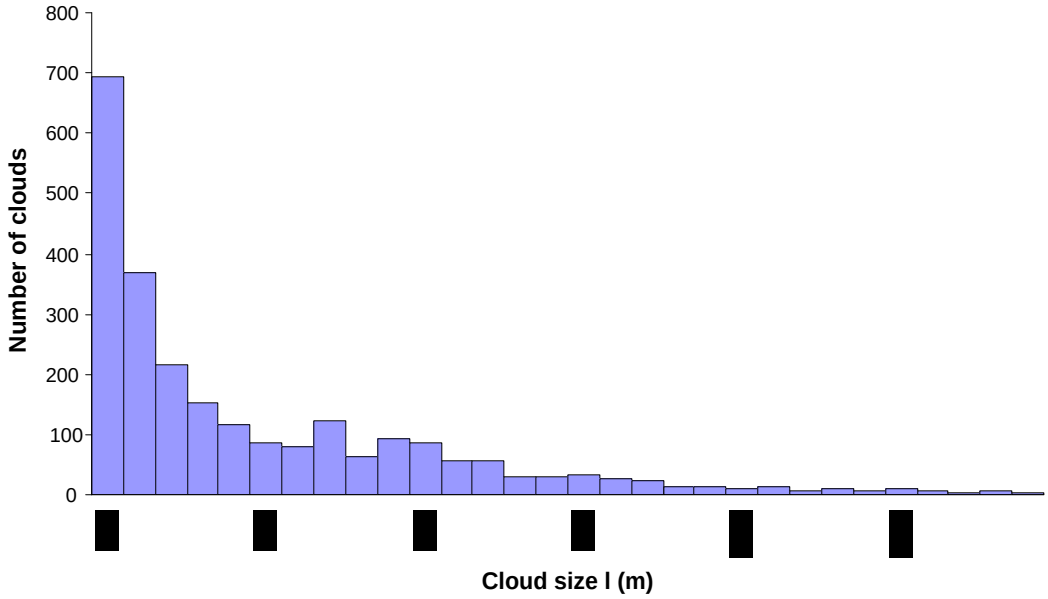
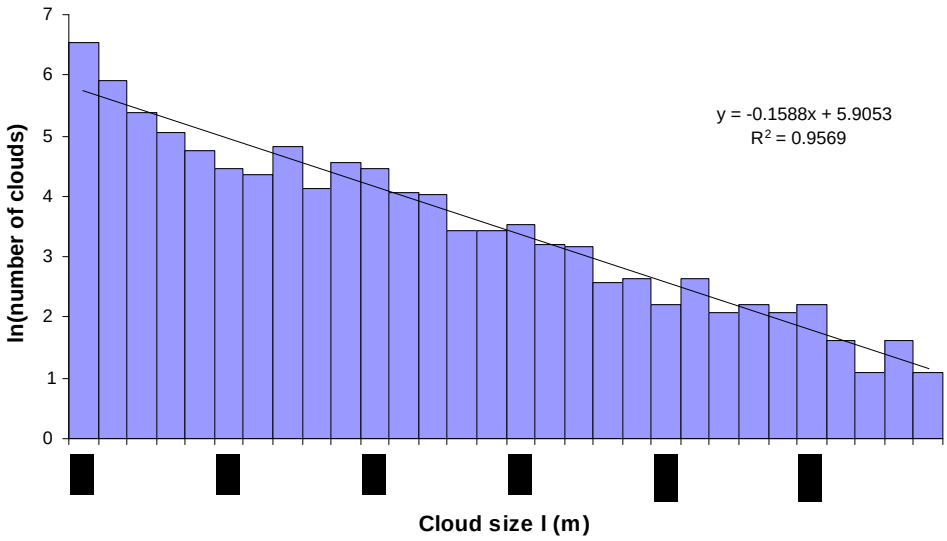


Figure 25: Simulated satellite image produced with cooling rate of 4K/day. Cloudy grid-square identified (and coloured red) where vertical velocity w>1ms^{-1}

Cloud sizes calculated using the eight-connected segmentation technique have been used to produce histograms based on the imposed cooling rates of 16K/day and 4K/day (Figures 26 and 27 respectively). It can be seen that for both forcings, analysis of the CRM output in the same way as satellite images have been analysed results in an approximately exponential distribution of cloud sizes. The R-squared values have been included to give an indication of how well the data fits this distribution. The 16K/day cooling rate has produced larger clouds in the simulated satellite image than the 4K/day case. The value of the coefficient of x in the equations included on the log-linear plots (i.e. the value of the constant b in Equation 2.3) is roughly twice as large in the 4K/day case.

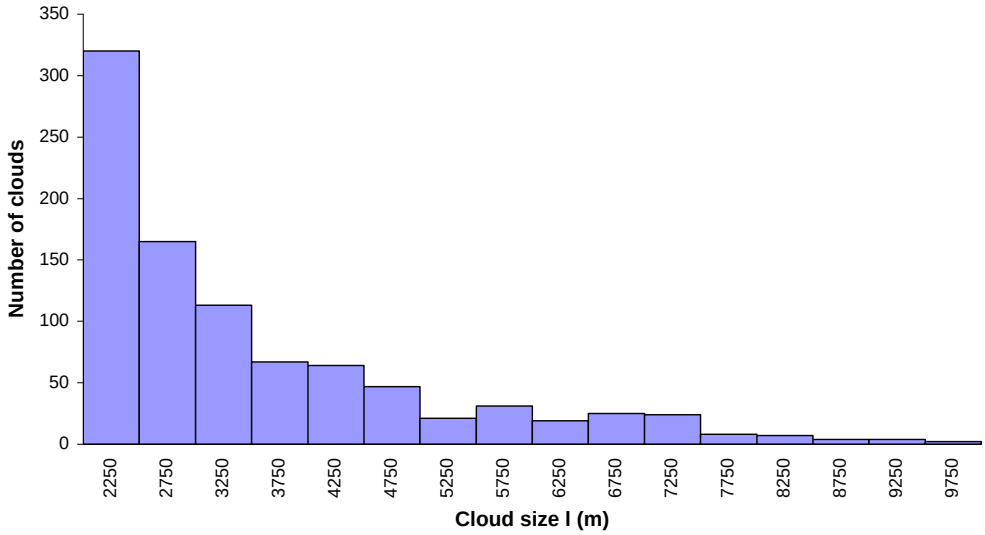


(a)

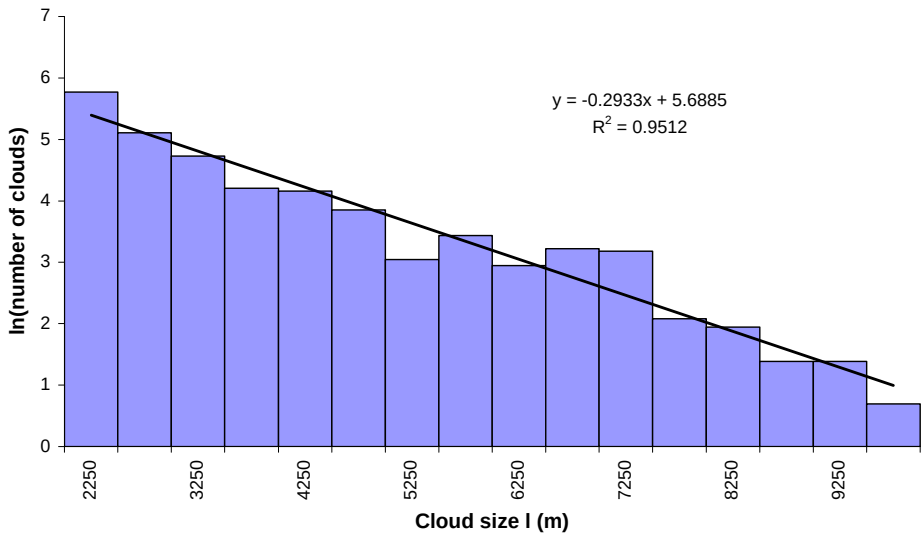


(b)

Figure 26: Histograms produced using cloud sizes from simulated satellite image, using cooling rate of 16K/day. (b) same as (a) but with logarithmic y-axis.



(a)



(b)

Figure 27: Histograms produced using cloud sizes from simulated satellite image, using cooling rate of 4K/day. (b) same as (a) but with logarithmic y-axis.

5.5 Consideration of power law distribution

For completeness, an attempt has been made to fit the cloud size data from the simulated satellite images to a power law distribu tion (Equation 2.4). The logarithm of the number of clouds has been plotted against the logarithm of cloud linear size for both the 16K/day and 4K/day radiative-cooling rates.

A true power law distribution as described by Equation 2.4 would result in a straight line on a log -log plot. In neither the 16K/day nor 4K/day (Figure 26) cooling rate cases is the straight line fit better that for the equivalent plot for an exponential distribution (see Figures 26 and 27), based on the R -squared values. It cou ld be concluded therefore that the exponential distribution is a better fit to the data than the single power law distribution.

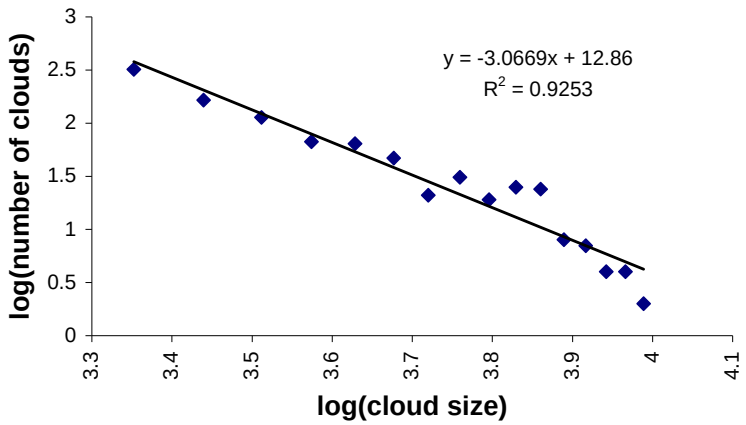


Figure 28: Logarithm of cloud number against logarithm of cloud size for 4K/day cooling rate

The log-log plot produced using a cooling rate of 16K/day is shown in Figure 29. Whilst a single straight line is clearly not visible, it could be argued that a double power law could reasonably describe the distribution. The double power law distribution is characterised on a log-log plot by a single line with a distinct change in gradient, i.e. two straight lines, with the break occurring at a particular cloud size. To a lesser extent this can also be seen in Figure 28.

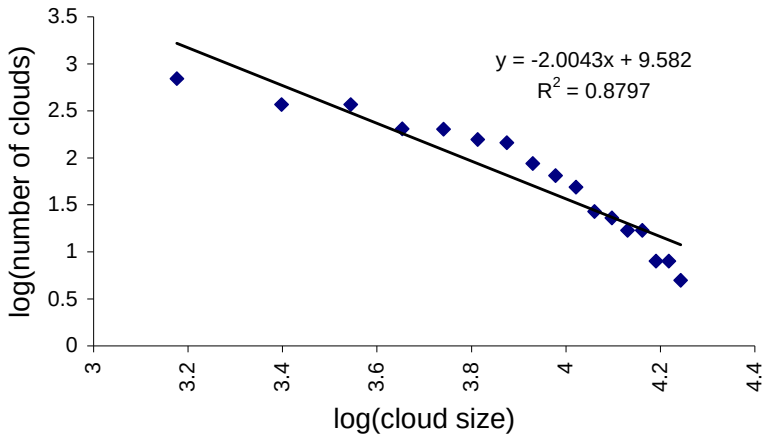
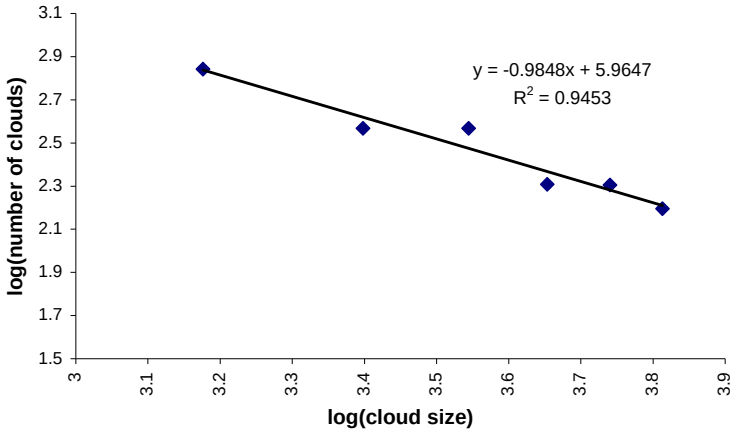
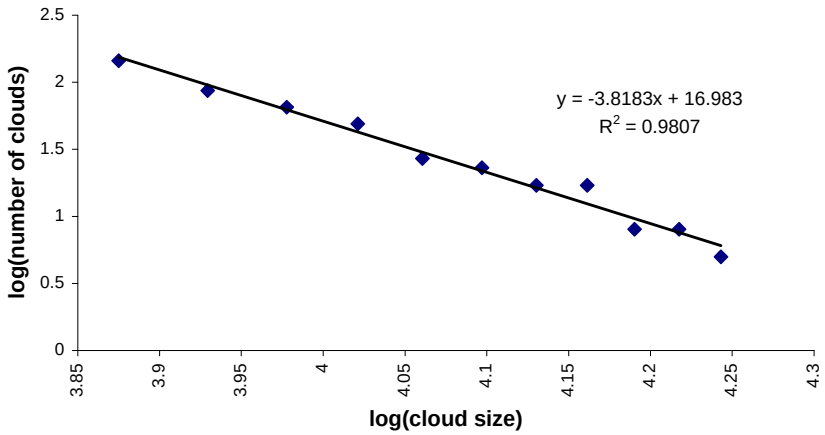


Figure 29: Logarithm of cloud number against logarithm of cloud size for 16K/day cooling rate

Individual plots for the data on either side of the possible break in Figure 29 (at approximately 7.5km) are shown in Figure 30. Best-fit lines have been included on the plots and the corresponding R-squared values are also displayed. The fitted line for the clouds with linear sizes greater than 7.5km has an R-squared value very close to 1, suggesting that the power law is a good fit to the data for clouds in this range. The line for clouds less than 7.5km in linear size fits the data less well, although this could be because there are fewer data points for clouds in this range. This suggests that it would be possible to conclude that the cloud size distribution might be described by a double power law with scale break at approximately 7.5km, and increase in power law exponent from $\alpha \approx 1$ to $\alpha \approx 4$. This change in exponent is similar to those reported in Table 1.



(a)



(b)

Figure 30: Best line fits through two apparent straight lines in Figure 29. (a) corresponds to cloud linear sizes less than 7.5km and (b) greater than 7.5km

5.6 Sensitivity tests

5.6.1 Sensitivity to definition of cloudy grid-point

Using both radiative cooling rates, simulated satellite images were produced by identifying cloudy grid-points where vertical velocity $w>0.8\text{ms}^{-1}$. This lowering of the threshold for cloudy grid-squares is equivalent to the assumption that weaker updrafts also imply cloud. A vertical velocity of 0.8ms^{-1} or more, particularly over an ocean (as simulated by the model) where there is no orography, is unlikely to be as a result of anything other than significant convective activity.

The resulting cloud size distribution for the 4K/day cooling case is shown in Figure 31. It appears exponential, and the similar R-squared value calculated indicates that the fit is almost as good as in the $w>1\text{ms}^{-1}$ case. The maximum cloud size is larger and the number of relatively large clouds has been increased, due to the increased number of grid-points identified as being cloudy. Overall, fewer single grid-square clouds (i.e. with linear size l=2km) are present and as such the number of clouds in the first bin has been reduced. (Whilst more single grid-squares have been identified as cloudy, they have been incorporated into larger clouds.) This has resulted in a reduction in the constant b. The 16K/day cooling rate case also reveals an exponential distribution and the value of the constant b has again been reduced.

The shape of the distribution does not therefore appear sensitive to the precise vertical velocity threshold applied however the constant b displays some sensitivity to cloudy grid-point definition. These results are summarised in Table 2. (In a future study, it would be interesting to test whether the results are sensitive to an entirely different cloudy grid-square definition, based on, for example, the presence of liquid or ice cloud water).

5.6.2 Sensitivity to segmentation technique

The simulated satellite images have been analysed using a 4-connected segmentation technique (Figure 3) to assess the sensitivity of the resulting distribution and associated constant. Both cooling rates have again resulted in approximately exponential distributions.

Comparison of Figure 32 with the corresponding 16K/day cooling rate simulated satellite image analysed with the 8-connected segmentation technique (Figure 26) reveals that the new technique has reduced the number of relatively large clouds and increased the proportion of clouds falling into the first bin. This is because there are more single pixel clouds in this case, since diagonally adjacent cloudy grid-squares are not assumed to be connected and as such are considered unique clouds (provided they are not *directly* adjacent to another cloudy grid-square). This has resulted in an increased value for the constant b in Equation 2.3. The 4K/day cooling rate has produced a similar increase in b for the same reasons. This result conflicts to a degree with the findings of Kuo et al. (1993) who concluded that results from both techniques were almost indistinguishable.

The new R-squared values indicate that the 4-connected segmentation technique produces data that fits the exponential distribution less accurately (Table 2). The increase in cloud number in the first histogram bin has contributed significantly to this. It seems likely that 8-connected segmentation is better able to capture the irregular nature of cloud sizes.

Cooling rate	w>1ms^{-1} 8-connected	w>0.8ms^{-1} 8-connected	w>1ms^{-1} 4-connected
16K/day	0.957	0.921	0.943
	0.16	*0.13*	*0.18*
4K/day	0.951	0.942	0.929
	0.29	*0.25*	*0.31*

Table 2: Summary of sensitivity test results, including R-squared values and *b* values (in *italics*)

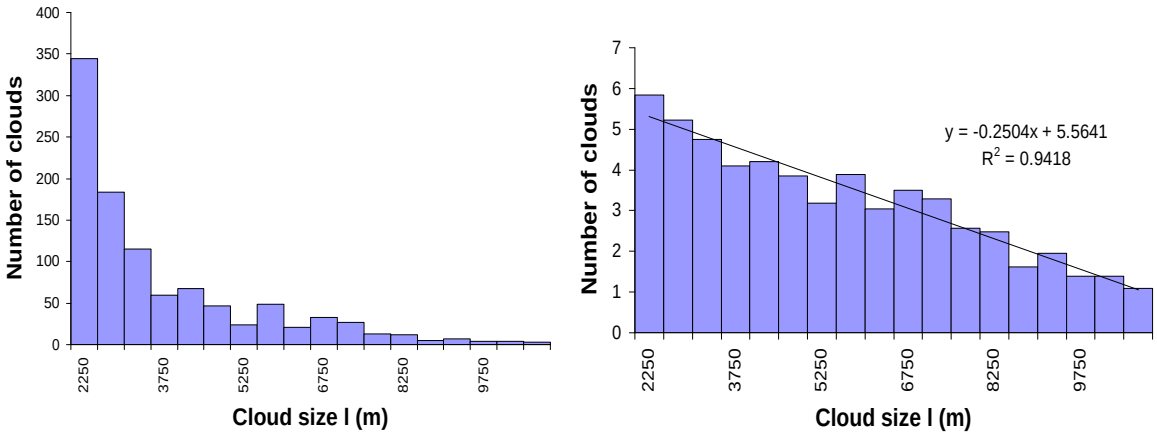


Figure 31: Cloud size distribution for simulated satellite image produced for 4K/day cooling rate and cloudy grid-point defined where w>0.8ms^{-1}

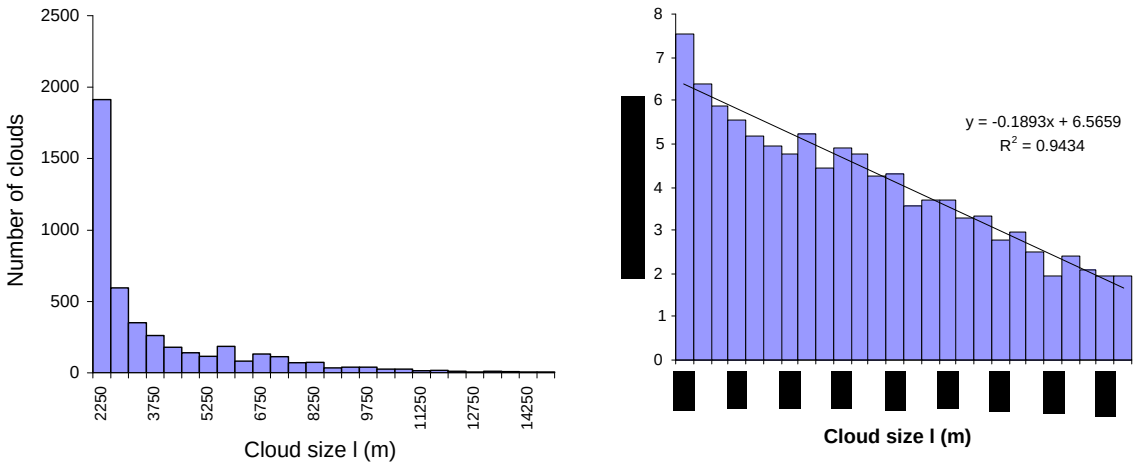


Figure 32: Cloud size distribution for simulated satellite image produced for 16K/day cooling rate and 4-connected segmentation technique for determining cloud sizes

5.6.3 Sensitivity to histogram bin width

Cumulus cloud sizes have been sorted into constant 500m class intervals for the purposes of this study. For consistency, this has remained constant throughout. However, to investigate whether the resulting distribution is sensitive to class interval size, histograms with bin widths of 1km have also been used. It has been found that the shape of the distribution is not sensitive to the choice of bin width. The value of the constant *b* in Equation 2.3 is, however, sensitive to bin width. The value roughly doubles when bins 1km wide are used.

6. Discussion

6.1 Distribution shape

The exponential distribution of mass flux and cloud size is the same as that expected using the statistical approach described by Cohen (2001), and analogous to the Boltzman distribution of molecular energies, i.e.

$$d\overline{n}(m) = \frac{\langle N \rangle}{\langle m \rangle} e^{-m/\langle m \rangle} dm \qquad (6.1)$$

where $d\overline{n}$ is the average number of clouds in the ensemble producing a mass between m and m+dm, $\langle N \rangle$ is the ensemble average total number of clouds in the system and $\langle m \rangle$ is the ensemble average mass flux per cloud.

The cloud size distribution is also in agreement with that determined experimentally by Plank (1969), Hozumi et al. (1982) and Wielicki and Welch (1986). As summarised in Table 1, the majority of recent studies have concluded that the power law is a general feature of convective cloud systems. Neggers et al. (2003a) reported double power law behaviour following analysis of LEM output. Nober (2003) found that cloud spectra produced using his simple Convective Cloud Field Model (CCFM) generally showed power law behaviour. The different results from this study are possibly because of the simplicity of the model, which incorporates a number of assumptions (detailed in Section 4), for example a non-rotating background. The grid-spacing of the model used by Neggers et al. was 50m x 50m (horizontally). This is a significantly finer resolution than the model used in this study (2km x 2km). This coarse resolution enables fast model runs but is clearly unable to resolve clouds smaller than 4km^2 in area (i.e. the area of one grid-square). This has severely limited the lower extreme of cloud sizes.

The domain size used by Neggers et al. was 6.4km x 6.4km. Clearly, the number of large clouds that are able to exist in the domain is limited. A larger domain size of 12.8km x 12.8km was also used to test the SCMS data, however this too restricts the number of larger clouds. Despite the coarse resolution, the large domain used in this study has enabled the simulation of many large clouds, contributing to a wide range of cloud sizes for analysis.

The discrepancy between the distributions obtained through analysis of satellite images could either be because of the errors introduced by the satellite image analysis or again due to the simplicity of the model used in this study. However it is possible that differences in the data analysis carried out in the various studies might explain some of the variation in results.

6.2 Vertical profile of exponent

Figures 20 and 21 show that following some fluctuation with height below 3-4km, the value of the constant *b* in Equation 2.3 is seen to decrease steadily. Analysis of cloud field 'snapshots' reveals an increasing number of larger clouds with increasing height and this is also clearly reflected in the cloud size histograms. As cloud plumes rise latent heat of condensation is released providing the energy required for further vertical growth. Horizontal growth occurs through mixing with surrounding air. This is mainly through entrainment as the plume rises, and detrainment when the plume reaches an inversion (or the level of neutral buoyancy). In addition, individual clouds may merge to form larger clouds, or cloud clusters. These processes explain the higher proportion of larger clouds at upper levels, reducing the rate of decay of cloud number with cloud size and resulting in a smaller value of the constant *b*. A similar decrease in *b* with height was observed for the mass flux distributions.

The 16K/day cooling rate is unrealistic and has been used to produce a large quantity of statistics. It has clearly resulted in a large number of cloudy grid-points on each level and has also produced a higher proportion of larger clouds than the lesser cooling rate. In particular, the histograms show that there is less of a drop in cloud number between the first and second histogram bins than in the 4K/day case. In other words, the decay in cloud number with increasing cloud size is reduced, corresponding to a smaller value for the constant *b* at each vertical level. The increased number of cloudy grid-points increases the likelihood of cloudy grid-points being connected, and hence increases the number of larger clouds.

6.3 Mean cloud size

The exponential distribution of cloud sizes is analogous to the decay of radioactive material:

$$N = N_0 e^{-t/\lambda} \qquad (6.2)$$

where λ is known as the mean lifetime. By comparing Equation 6.2 with Equation 2.3 it can be seen that λ is equivalent to $1/b$. It can be deduced that the inverse of *b* represents some characteristic or mean cloud size for the distribution. A smaller *b* would therefore suggest a larger characteristic cloud size than a larger *b* value.

6.4 Simulated satellite images

An exponential distribution of cloud sizes has been found for both the 16K/day and 4K/day radiative cooling rates, based on the simulated satellite images. Equations 6.3 and 6.4 describe the distributions for 4K/day and 16K/day respectively:

$$n(l) = 295e^{-0.29l} \qquad (6.3)$$

$$n(l) = 862e^{-0.16l} \qquad (6.4)$$

Once again, the constant *b* in Equation 2.3 has a smaller value for the 16K/day cooling case and the histograms (Figures 26 and 27) show the reduced decay in cloud number with size for this cooling rate. The exponents are slightly smaller than the smallest values reported by Plank (1969) and Hozumi et al. (1982).

Comparison of the exponents obtained from the simulated satellite images to the vertical profiles of Figures 20 and 21 show that they are smaller than any value obtained for the size distribution at a single vertical level. Rodts et al. (2003) compared direct measurements from flights through cloud fields with Landsat satellite data and found that the satellite image was dominated by larger clouds than was suggested by the aircraft data, although this will at least in part be because flights are unlikely to traverse the full length of each cloud. It seems likely that the overlapping of clouds at different vertical levels has led to the reduction in exponent value. Overlapping clouds result in a larger vertically projected cloud area 'viewed' by the satellite, and the number of smaller clouds is also reduced, hence the rate of decay of cloud number with size is reduced. This could also explain why the exponent values are smaller than those derived in earlier studies. In addition, the model cannot resolve clouds smaller than 2km in linear size, effectively reducing the number of smaller clouds in the distribution.

The variation of *b* with height, together with the reduction introduced when overlapping clouds are considered, has important implications for cloud parameterisation schemes. Frank and Cohen (1985) used a cloud model to produce an ensemble of clouds based on the statistical distribution obtained by LeMone and Zipser (1980). They noted that variations in the assumed distribution of cloud sizes resulted in significant variations in the vertical heating profiles simulated using their model. Future studies should consider the height at which the ensemble is specified, since this would determine the value of any constants used to generate the distribution. Distributions of mass fluxes specified would be determined by the height at which the convective updraft plume was initialised. In addition, in studies involving distributions derived from satellite data it would be important to consider the reduction in *b* introduced by the overlapping of clouds on different vertical levels (should an exponential distribution be specified) and the effect this overlapping has on other distributions, for example the lognormal distribution.

Currently, research is being conducted into the incorporation of statistics into a convective model through the use of an exponential distribution of mass fluxes. The model uses the distribution to calculate the probability of occurrence of a particular mass flux on a vertical level (Plant and Craig 2004). The parameterisation scheme specifies the distribution at one level and subsequently derives distributions at higher levels. The variation of *b* with height from this study could be used as a useful test of such parameterisation schemes

7. Conclusions

In Section 2 it was suggested that some of the variation in cumulus cloud size distribution study findings might be explained by the differences in data analysis techniques implemented. This study has confirmed that it is sometimes possible to fit more than one functional form to cloud field data. Whilst one function may be a better fit to the distribution than the other, unless a fit to each function has been attempted it is not known which result best represents the distribution.

The double power law is generally thought to most likely describe the distribution of cumulus cloud sizes. Many studies support this theory, including most recently a study of LES model data (Neggers et al. 2003a). The uncertainties introduced by the various studies have been discussed in some detail in Section 3. Although they are significant, the volume of evidence in support of the double power law lends the theory much credibility. Despite this, there is still disagreement concerning the value of the power law exponents and size of cloud at which the break in power law occurs. This study has shown that the choice of segmentation technique, cloudy grid-point definition and class size interval can all significantly affect the calculated value of any exponents. This should be considered when comparing values derived during the various studies.

The exponential distribution is generally the best fit to the CRM data analysed in this study. This is true for clouds at various vertical levels and for simulated satellite images of cloud fields, and has been confirmed for different radiative cooling rates. The result is contrary to those most recently reported. The exponential fit to the data is generally good, although it has been shown that in the case of the simulated satellite images, similar to those analysed by Neggers et al., a double power law can be fit reasonably well (particularly for the highly populated cloud fields produced using a cooling rate of 16K/day). There is also some uncertainty in the results for cloud size distributions at particular heights. Whilst good exponential fits were found for most vertical levels, fits to a single power law distribution were also often good and

sometimes better (based on the R-squared value). Figure 33 shows how the R-squared values vary with height for a cooling rate of 16K/day. Only R-squared values greater than 0.95 have been included. There is some oscillation, however it appears that the exponential distribution most nearly describes the cloud distributions at lower levels, while higher up where the clouds are larger, the power law distribution provides the better fit to the data. This result is not conclusive however it is consistent with the findings of Plank (1969) whose exponential distribution failed to accurately describe the distribution of cloud sizes in the afternoon, when larger more mature clouds dominated the cloud fields. It is also interesting to note that Machado and Rossow (1993) reported a single power law for large clusters.

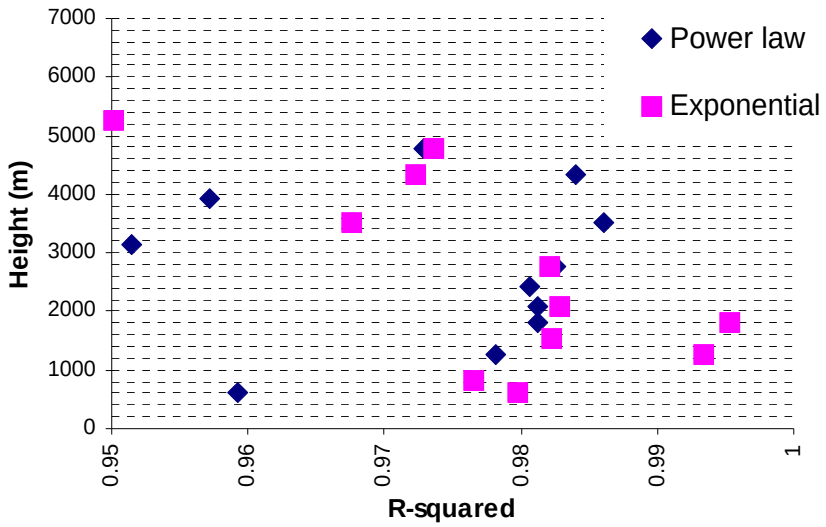


Figure 33: Variation of R-squared values with height for 16K/day cooling rate

It was also suggested in Section 2 that exponential distributions might mistakenly be reported when the range of cloud sizes is limited, such that the entire spectrum of clouds investigated falls to one side of the power law scale break size. An arbitrary distribution of clouds perfectly described by a single power law has been generated. Using this distribution, and looking only at clouds greater than a certain size, a log-linear plot has been produced and is shown in Figure 34. Using linear regression, a best-fit line has been superimposed also. The corresponding R-squared value indicates that the straight line is a good fit to the data. This plot can be compared to

the histograms based on model output earlier in this study, and the similarity casts some doubt on the exponential distribution result. The model resolution imposed a lower limit of 2km on the linear size of clouds investigated in this study. It seems possible that a double power law is the true functional form describing the cloud size distribution, but that the break in power law occurs at a linear size less than 2km. This break size is not inconsistent with those that have been reported (Table 1). It is smaller than the possible break size suggested for the simulated satellite distribution produced using a cooling rate of 16K/day (see Figure 30), however an exponential distribution was most likely in this case.

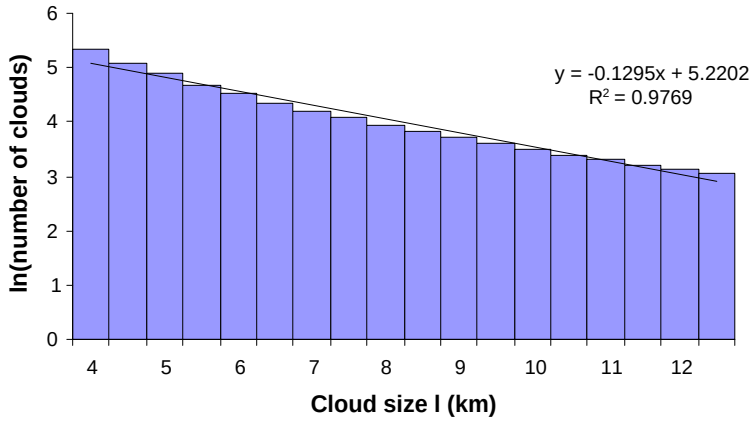


Figure 34: Distribution created using power law with exponent α=2 on a log-linear plot

7.1 Future work

The determination of a universal functional form for the distribution of cumulus cloud sizes is clearly complex. It is not currently possible to positively rule out one any of the proposed possibilities. In future studies, attempts to fit experimental data to more than one functional form would be beneficial. Comparison of the fits would indicate which form, if any, most nearly represents the actual size distribution. This has not been done in the majority of previous studies, although Benner and Curry (1998) did consider both exponential and power law distributions. The double power law distribution appears the most likely candidate. Research into the causes of the proposed scale break could further reduce the uncertainty in this. LEMs will be a useful tool in this research, since they can be configured to reproduce a wide range of scenarios relatively easily.

It has been found that investigations based on a narrow range of cloud sizes may lead to incorrect conclusions being drawn. As such, data for a wide range of cloud sizes should be used in future research.

In addition, further analysis of satellite data could lead to increased confidence in one of the universal functional form candidates. Different visible/TIR thresholds could be used to investigate the variation of distributions with height, and results could be compared with this study. Improvements in remote sensing instrument resolutions and analysis techniques would reduce uncertainty in the results of these studies.