UC Santa Barbara
UC Santa Barbara Previously Published Works

Title

An Application of Fractal Box Dimension to the Recognition of Mesoscale Cloud Patterns in Infrared Satellite Images

Authors

Carvalho, Leila MV

Dias, Maria AF Silva

An Application of Fractal Box Dimension to the Recognition of Mesoscale Cloud Patterns in Infrared Satellite Images

Leila M. V. Carvalho and Maria A. F. Silva Dias

Department of Atmospheric Sciences, Institute of Astronomy and Geophysics, University of São Paulo, Sao Paulo, Brazil

ABSTRACT

Mesoscale cloud patterns are analyzed through the application of fractal box dimensions. Verification of fractal properties in satellite infrared images is carried out by computing box dimensions with two different methods and by computing the fraction of cloudy pixels for two sets of images: 174 are considered the "control series," and 178 (for verification) are considered the "test series." The main instabilities in the behavior of such dimensions are investigated from the simulation of circles filling space in several spatial distributions. It is shown that the box dimensions are sensitive to the increase of the area covered and to the spatial organization—that is, the number of cells, the spatial clustering, and the isotropy of the distribution of pixels. From a principal components analysis, the authors find six main patterns in the cloudiness for the control series. The three main patterns related to enhanced convection are the massive noncircular spread cloudiness, the highly isotropic distribution of cloud in several cells, and the most circular pattern associated with mesoscale convective complexes. The six patterns are separated into a cluster analysis, and the properties of each cluster are averaged and verified for the test series. This method is a simple and skillful procedure to recognize mesoscale cloud patterns in satellite infrared images.

1. Introduction

The cloud amount, its distribution, intensity, and organization over an area depends on the dynamic and thermodynamic factors interacting on several scales. Model results and observations have shown the importance of such interactions for the complexity of cloud patterns. On the other hand, cloudiness features, including cloud amount, top height, anvil extension, and intensity of convective and mesoscale updrafts and downdrafts, can act upon the evolution of meteorological variables and on the dynamic patterns themselves (Betts 1976; Ogura and Cho 1973; Albrecht et al. 1986; Frank and Cohen 1985; Johnson 1976).

One dramatic example of the importance of cloud organization upon its evolution can be found during the mesoscale convective system (MCS) genesis. The merging of anvils in the first stage of an MCS leads to the formation of elevated stratiform precipitation, an upward shift in the mean vertical motion (Gamache and Houze 1982; Dudhia and Moncrieff 1987), and an associated upward shift of the level of maximum heating (Esbensen and Wang 1984; Silva-Dias et al. 1984), favoring its upscale growth to mesoscale proportion (An-

thes and Keyser 1979; Hack and Schubert 1986), which is characteristic of the development of a mature, midlatitude mesoscale convective complex (MCC).

The objective of this paper is to propose a new method of analysis of cloud geometry distribution obtained from satellite infrared (IR) images that is able to recognize mesoscale patterns. Weather forecasters and climatologists have classified different mesoscale systems from satellite IR images in a subjective way. We discuss an automatic discriminating method applied to just one channel (the infrared) to detect different organizations of clouds over a limited area.

The fractal geometry (Mandelbrot 1977, 1982) provides a model to investigate cloud complex shapes and complex cloud distribution over an area. One general characterization of fractals is that they are the end result of physical processes that modify shape through global, local, and internal action. Therefore, the fractal dimension of a surface corresponds quite closely to our intuitive notion of roughness and can be used to resolve the problems of texture segmentation and classification (Peleg et al. 1984; Pentland 1984).

Scaling properties in clouds have been found from satellite images, indicating self-similarities in several scales (Lovejoy 1982; Machado et al. 1992). However, the studies of Cahalan and Joseph (1989) and Sengupta et al. (1990), using enhanced data resolution with the Landsat multispectral scanner, have shown different power laws (fractal dimensions), depending on the na-

Corresponding author address: Leila Maria Véspoli de Carvalho, Department of Atmospheric Science, IAG/USP, Rua do Matão 1226, 05508-900, São Paulo, SP, Brazil.
E-mail: leila@model.iag.usp.br

ture of clouds. Lovejoy et al. (1987) and Gabriel et al. (1988), using an analysis technique called ''functional box counting,'' have shown that both radar rain and satellite cloud fields are multifractal over various ranges in scale. Schertzer and Lovejoy (1987), Lovejoy and Schertzer (1990), Olsson et al. (1993), Tessier et al. (1993), Cahalan et al. (1994), and Marshak et al. (1997) present some examples of multifractal properties applied to observed rain and cloud fields. The fractal structure of clouds with a simple atmospheric model is examined in Yano et al. (1996), who show that the fractality of simulated, large-scale cloud systems depends on cumulus parameterization. Yano et al. thus show the necessity of quantifying the relationship between fractal properties and mesoscale cloud system dynamics.

In this paper, clouds viewed by satellite IR images are considered in a mesoscale context. Multiple dimensions are found, and the relationship between the fractal box dimension and the patterns of satellite cloud images is investigated. Within the analyzed spatial scale we will show that the distribution of clouds is scaling.

To interpret the multiple dimensions of cloud distribution, we focus our analysis on two main aspects. First, we analyze the box dimensions of actual IR satellite images of clouds in a limited area. Second, we compute box dimensions for the simulated distributions of ''well-behaved'' structures, such as circles. The simulated structures give us a basis for the discussion of the fractal behavior of the cloud distribution over the area. The main aspect highlighted by the fractal box dimension is its ability to identify different structures according to their level of aggregation, the circular patterns, and the isotropy of their pixel distribution. As discussed in Pentland (1984), the fractal box dimension is related to the complexity of the objects over an area, but the concept of roughness of a homogeneous surface (i.e., a single cloud) attributed to fractal dimensions can be broadened for the distribution of a set of fractal structures (several clouds).

This paper is organized as follows. Section 2 describes two procedures to compute box dimensions. Section 3 discusses the differences in the behavior of those dimensions and the possibility of using the dimensions to identify cloud patterns in two series of satellite IR images that show some of the main mesoscale systems in the continental area during different periods of the summer. Section 4 proposes a method to identify patterns in satellite images by using box dimension and multivariate statistical methods such as principal components and clustering analysis. We show that by combining both methods a large variety of geometrical patterns of mesoscale systems are identifiable, among them the different evolutionary stages of MCCs.

2. The fractal box-dimension definition

The scaling dimension of the square is based on representing the square as an almost-disjoint union of small unit squares. The Hausdorff dimension of an object is based on covering the object by small disks or balls for a minimum cover. The box dimension of a subset X of the plane is defined similarly by counting the number of cells of a grid with constant Δs that intersect X, $N(\Delta s)$. Then X has a dimension D if $N(\Delta s)$ satisfies the power law

$$N(\Delta s) \approx c(1/\Delta s)^D \qquad (1)$$

asymptotically in the sense that

$$\lim_{\Delta s \to 0} N(\Delta s)\Delta s^D = c. \qquad (2)$$

The box dimension D can be computed from (2) as

$$D = \lim_{\Delta s \to 0} [-\log N(\Delta s)/\log \Delta s]. \qquad (3)$$

The fractal dimension as defined by Eq. (3) is usually identified with the Hausdorff–Besicovitch dimension and is known as the capacity dimension D_O (see, e.g., Tsonis 1992). Nevertheless, there is a fine distinction between the Hausdorff–Besicovitch dimension and the capacity dimension: while the former is obtained by covering the set minimally with hypercubes that may be different in size, the latter is obtained with the same process except that the hypercubes are the same size (Essex and Nerenberg 1990).

In this work, we present two methods to compute the box dimension. In the first, each image is considered a set $S(Tc) = \{x, y \mid T(x, y) \geq Tc\}$, where Tc is a threshold in counts. The image is covered by square boxes that vary their scale from 2 to 2^4 pixels. This upper limit is defined as the break point relative to the length of the side of each image (98 pixels) in order to obtain just one value of D from Eq. (3). The box dimension obtained from this method will be referred to as D_b. This method is similar to that Lovejoy et al. (1987) apply to radar reflectivity for the dimension they call $D2(T)$, where T is a threshold.

A satellite IR image can be considered a surface of varying height as the gray level (or temperature) changes in space. In this sense, a box dimension can be obtained by considering cubes of height h, which is defined as the difference between the maximum and the minimum count level in some scale Δs for a given area S:

$$N(\Delta s) = \frac{\overline{n}(\Delta s)S}{\Delta s^2}, \qquad (4)$$

where $N(\Delta s)$ is the number of cubes of side Δs and

$$\overline{n}(\Delta s) = \frac{\sum (h/\Delta s)}{n} = \frac{\sum \frac{\Delta s + \Delta z}{\Delta s}}{n}$$

$$= \frac{\sum 1 + \text{INT}\frac{(C_{max} - C_{min})}{\Delta s}}{n}$$

$$= 1 + \frac{\left[\sum \text{INT}\frac{(C_{max} - C_{min})}{\Delta s}\right]}{n}. \qquad (5)$$

Here C_{max} and C_{min} are the maximum and minimum value of counts, respectively, and n is the number of boxes in the scale Δs. Since different units are used in the z direction [$z = C(x, y)$, where C is the count value, equivalent to temperature], the box-counting dimension computed from Eqs. (4) and (5) is equivalent to the "graph" dimension (Mandelbrot 1977) and is referred to in this paper as D_g. Its value ranges between 2 (the topological dimension of the plane) and 3 (the topological dimension of the cube). It should be emphasized that the proposed method of computing D_g includes all boxes inside the studied area, not just those that include "clouds" or simulated circles.

3. Application of the box dimension to satellite IR images

Lovejoy (1982), Wielicki and Welch (1986), Gabriel et al. (1988), Lovejoy and Schertzer (1990), Tessier et al. (1993), Cahalan and Joseph (1989), Machado et al. (1992), and Mapes and Houze (1993) are some examples of studies that have applied scaling properties to satellite images of clouds. In this section we will show how the scaling properties of mesoscale cloud fields inferred by box dimension can be applied to the recognition of fractal mesoscale patterns in satellite IR images.

We will show that the fractal box dimensions are sensitive to different distributions of cloudy pixels over an area. Mesoscale convective systems appear in different spatial organizations, according to the dynamic and thermodynamic behavior of meteorological variables such as wind field, temperature, humidity, convective available potential energy, and so on. We will attempt to identify the cloud distribution organization from a few parameters extracted only from satellite IR images, among them the box dimensions.

a. The satellite dataset

To carry out this study, some areas were chosen in the South American tropical and subtropical zones. The IR images were obtained from the *Meteosat-3* satellite in two different periods during the summer of 1994 with a 1-h time step. The images correspond to two periods of consecutive days 20–21 and 24–25 January and were found to be enough to sample several MCSs in the Tropics.

During summer, several convective systems, such as MCCs, squall lines, and isolated convective cells, develop in tropical continental areas. We investigate such MCSs from satellite IR images, sampled over four regions, corresponding to 98 × 98 satellite pixels, whose coordinates are given in Table 1. These regions were named based on the region covered by most pixels in each area. The samples cover the Amazon Basin and the regions of north-northeast, south, and southeast Brazil, southern Bolivia, northern Argentina, Paraguay.

TABLE 1. Studied area coordinates.

Local	Upper latitude	Lower latitude	Upper longitude	Lower longitude
Paraguay	$-22.31°$	$-32.41°$	$-49.27°$	$-60.07°$
Southeastern Brazil	$-14.03°$	$-23.15°$	$-46.58°$	$-56.65°$
Amazon Basin	$-3.49°$	$-11.73°$	$-57.23°$	$-65.81°$
Northern Brazil	$-0.19°$	$-8.13°$	$-41.66°$	$-51.86°$
Bolivia*	$-16.16°$	$-25.25°$	$-57.84°$	$-67.19°$

* Included in the test series (see text).

For each of those areas, a time series of images was available for the two periods. To develop the best method to recognize mesoscale patterns we chose as the "control series" the sample series of 24 and 25 January 1994. These images were chosen with the purpose of sampling the most common patterns of cloudiness in summer over the continental tropical and subtropical area. One of the most important mesoscale systems found in this sample was an MCC at the latitude of Paraguay and northern Argentina.

The time series of 20 and 21 January was used as a "test series" to verify the results obtained with the control series in independent satellite images. It includes the samples over Bolivia (Table 1). These images were obtained in the same areas as the control series but with no previous knowledge of the characteristics of the developing MCSs. The control series contains 174 images, while the test series contains 178. Each sequence focuses on the evolution of a given cloud scene.

The algorithm used to compute the box dimension—and therefore all of the variables used to analyze the satellite data—is based on count value, not temperature. This choice was made in order to generalize the method to other possible fractal, noncloud structures observed in images. The equivalence between temperature and count value for the Meteosat satellite can change slightly from day to day. This is one reason to limit the time series of satellite data to only two consecutive days.

For all of these time series the following variables were computed: dimensions D_b and D_g, and the percentage of area covered by values greater than count 185 considered as cloud scenes, or simply the cloud fraction for the satellite resolution pixel (hereafter referred to as Ac). Note that the original count value (Cv) of Meteosat was transformed to 255 Cv.

Count 185 was chosen to represent the cloudy areas since the corresponding temperature in the studied days is about 240 K—that is, 239.9 K (day 24), 240.0 K (day 25), 240.3K (day 20), and 240.4 (day 21)—and therefore corresponds to cloud scenes. These differences in the temperature threshold can be neglected for our purposes.

b. Box dimensions and fractal cloud patterns

1) THE CONTROL SERIES OF DAYS 24 AND 25

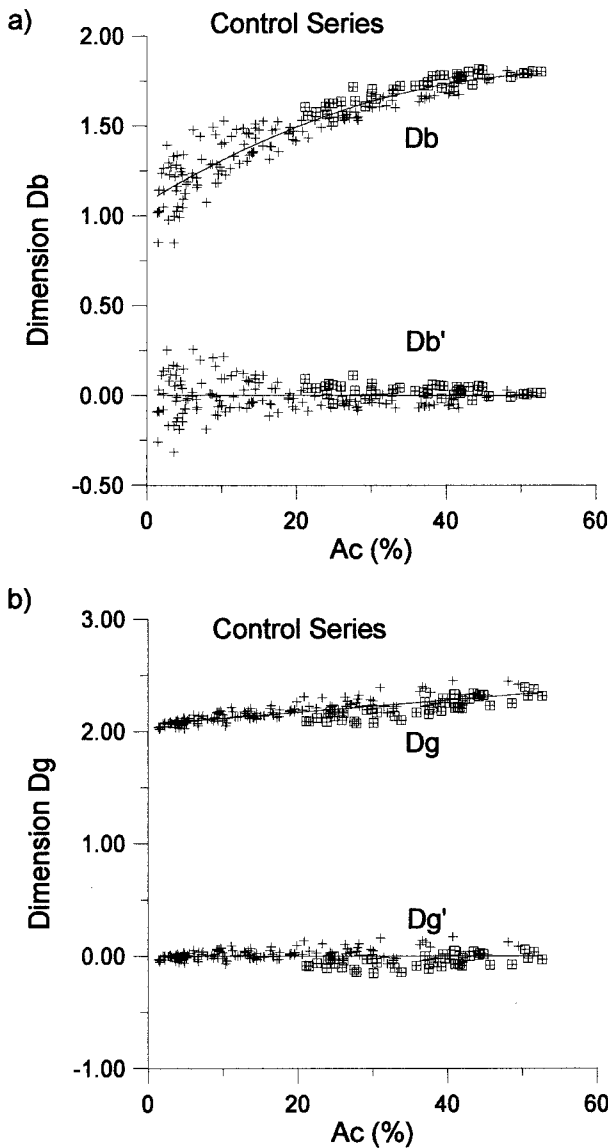It is well known that in satellite IR images both the pattern and the cloud fraction inside an area change.

a)

b)

FIG. 1. (a) Dimensions D_b and D'_b against Ac for the control series. (b) The same as (a) but for the dimensions D_g and D'_g. Squares represent the data from the Paraguay area.

TABLE 2. Coefficients of polynomial fit between D_b and D_g vs Ac.

Series	Dimension	Polynomial degree	Residuals explained (%)	Coefficient
Control series	D_b	0	0	1.07514
		1	83	0.025528
		2	87	−0.000226772
Control series	D_g	0	0	2.06278
		1	66	0.00544504
Test series	D_b	0	0	1.06169
		1	56	0.02887
		2	62	−0.00323322
Test series	D_g	0	6	2.0632
		1	63	0.0010686
		2	68	−0.00017617

Table 2. In Fig. 1, D_b is plotted by subtracting the polynomial fit.

Figure 1b shows a linear tendency between D_g and Ac. A polynomial fit with degree greater than 1 (not shown) explains only 1% more of the variance, indicating that the straight line can be considered the best fit for this pair of variables. The coefficients of the straight line fitted to the data may be seen in Table 2. The fluctuation values of D_b and D_g computed after extracting the line fit will be referred to as D'_b and D'_g.

To understand the behavior of fractal box dimensions, idealized pictures were built to represent different spatial distributions of pixels over the area. These pictures were constructed using circles as the basic objects. Although the box-counting technique can be applied to images constituted of circles [a power law can be found according to Eq. (2)], the circle is not self-similar: that is, it cannot be written as an almost-disjoint union of smaller circles. Nevertheless, the circle, a "well-behaved" object, was chosen for two main reasons: 1) the resemblance of circular structures with natural convective clusters in several scales and 2) the understanding of how the box dimension changes, which is due exclusively to the distribution of cloudy pixels over the studied area.

We should emphasize that all experiments performed with circles, and therefore the fluctuations of box dimensions, have validity only for the same area and the same resolution of the satellite IR images. The "circles" are obtained by an integer approximation of the equation $r = (x^2 + y^2)^{1/2}$, where r is the radius of the circle and x and y are the coordinates with respect to the center of the circle. The integer approximation implies a nonsmoothed aspect, as shown in Fig. 2. Furthermore, we have limited range of scales allowed in the case of real and simulated images: the limit of $\Delta s \rightarrow 0$ in Eq. (3) is a theoretical limit since satellite images impose this limit due to their resolution.

Figure 2 shows circles of increasing radius, keeping constant the area in which they are embedded. The external circle radius ranges from 10 to 50 pixels, with a

Two questions arise: Does the box dimension change when the fraction of objects or pixels is increased (decreased), split off (compacted), or spread out (constituting a well-defined circle) over an area? Is there any particular behavior of the box dimension related to such patterns? In this section we will discuss how box dimensions can distinguish fractal patterns of clouds and how they depend on the cloud fraction in the images.

Figures 1a and 1b show the plot of D_b and D_g against Ac, respectively. There is a tendency of increasing the value of box dimension as Ac increases. Figure 1a shows a 2° polynomial curve as the best fit between D_b and Ac. The value of the coefficients and the percentage of variance explained by the polynomial fit are listed in
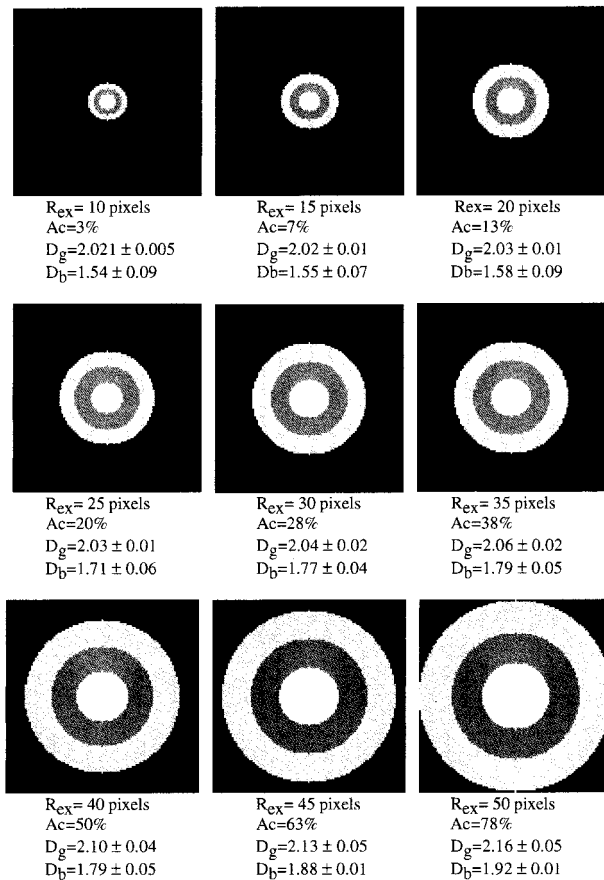
FIG. 2. Circles with external radius increasing from 10 to 50 pixels with a step of 5 pixels. Beneath each picture are the values of the external radii R_{ex}, Ac, D_b, and D_g and the standard deviation error.
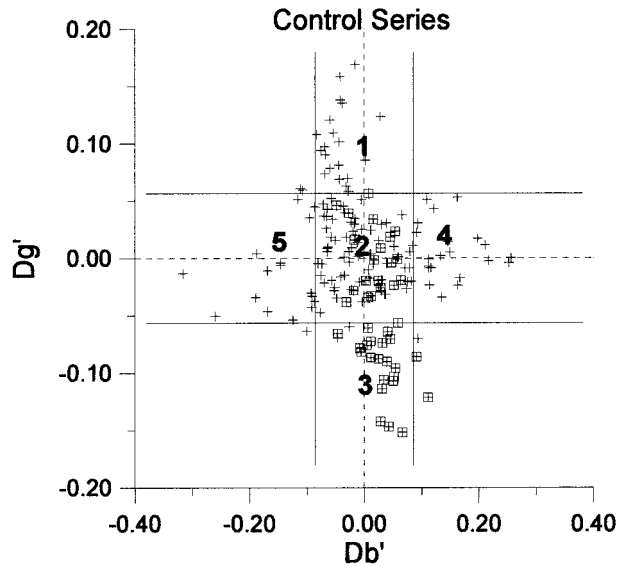

FIG. 3. Diagram of $D_g' \times D_b'$ for the control series. Solid lines indicate one standard deviation of the mean value. Numbers indicate the categories (see text). Squares represent the data from the Paraguay area.

step of 5 pixels between images. The side of the square area is equal to 98 pixels. The values of D_b and D_g computed according to Eq. (3) and their standard deviation are shown in the same picture.

Note that D_b tends to increase as the fraction of pixels, or radius of the circles, increases. Dimension D_b ranges from about 1.54 to 1.92, while the percentage of covered pixels ranges from 7% to 78%, indicating that D_b tends to the Euclidean dimension of the circle as the circle fills the image. The standard deviation of D_b decreases, and the linear correlation coefficient increases as the area is filled by the circle.

It must be noted that the box dimension D_b, according to the definition, does not depend on the internal gray level since the box-counting technique counts only those boxes of side Δs that contain at least one pixel with a value greater than the chosen threshold.

The increasing values of D_g as the circle radius is increased are also seen in Fig. 2. The standard deviation error of D_g increases, and the linear correlation coefficient decreases from 0.9999 to 0.9987 as the area is filled for the circle. This means that it begins to deviate from the straight line on a log–log plot, indicating that

the proposed methodology of computing D_g is able to detect a slight tendency of breaking the self-similarity or scaling properties as the picture is filled with a circle. Dimension D_g ranges from 2.02 to 2.16 in this simulation—that is, an increase of about 7%—while the increase in D_b is about 25%. It seems therefore that D_b is more sensitive to the increase of Ac and tends to the Euclidean topological dimension faster than D_g. This is a direct result of the greater increase of $N(\Delta s)$ for smaller scales Δs, compared to the larger ones, as the circle fills the studied area [see Eq. (3)].

This property of box dimensions seems to be related to the results pointed out by Lovejoy et al. (1987). They found a systematic decrease in the box dimensions as the threshold T of radar reflectivity increases. As T increases, the fraction of radar pixels greater than or equal to T decreases, which leads to a resulting decrease of the box dimensions. A similar result was found by Gabriel et al. (1988) for IR and VIS radiance.

Hamburger et al. (1996), studying fractal properties of randomly placed n-dimensional spheres ($n = 1, 2, 3$), have shown that fractal dimensions (the box-counting and Minkowski-sausage techniques were applied) are not universal and depend on density. Using analytical and numerical calculations, they have also shown that in the regime of a low-volume fraction occupied by the spheres apparent fractal behavior is observed for a range of scales between physically relevant cutoffs. Hamburger et al.'s results can explain why the circles considered in our experiments have fractal dimensions if we take into account the range of scales Δs used for the solution of Eq. (1). Nevertheless, even being considered "false fractals," our simulated pictures have be-

havior that is similar to true fractals (clouds) for the range of scales considered.

If the changes in fractal box dimensions are related to how the cloudy pixels are distributed over the area, the values of D'_b and D'_g for the satellite images obtained by subtracting the dependence on Ac from the box dimensions (Fig. 1) could indicate different structures or fractal patterns in the cloudiness. Figure 3 shows a plot of D'_b against D'_g, assuming this proposition. We observe the "cross" aspect of the plot, suggesting that the extremes of the cross are possibly related to different fractal patterns. The Paraguay time series, which represents the MCC life cycle, is indicated with a different symbol.

To investigate these properties of the box dimension, we separate the images belonging to each sector of the cross in the range of values of D'_b and D'_g, established as follows:

- category 1 $-\sigma_b < D'_b \le \sigma_b$ and $D'_g > \sigma_g$,
- category 2 $-\sigma_b < D'_b \le \sigma_b$ and $-\sigma_g < D'_g \le \sigma_g$,
- category 3 $-\sigma_b < D'_b \le \sigma_b$ and $D'_g > -\sigma_g$,
- category 4 $D'_b > \sigma_b$ and $-\sigma_g < D'_g \le \sigma_g$,

and

- category 5 $D'_b \le -\sigma_b$ and $-\sigma_g < D'_g \le \sigma_g$,

where $\sigma_b = 0.086$ and $\sigma_g = 0.056$ are the sample standard deviations for D'_b and D'_g, respectively.

Figure 4 contains eight images with values of D'_b or D'_g greater than one standard deviation from the mean. Panels (a) and (b) show two typical images of category 1, (c) and (d) show two typical images of category 3, (e) and (f) category 4, and (g) and (h) category 5. As expected, images in each category have similar patterns: the category 1 images show a cloud pattern in which the cloudy pixels are spread out almost homogeneously over the area in several cells; that is, no circular, centered convective organization is found. In category 3, which has remarkably lower values of D'_g, the images show highly organized convection in well-defined circular structures, most of which belong to an MCC event (Figs. 4c–d). Other circular structures in this category, such as a storm in the Amazon Basin, do not correspond to an MCC event. Because we have extracted the tendency on Ac, we can find some broad range of Ac in all of the categories.

Using Machado et al.'s (1992) method to detect connected areas of cold cloudiness, we obtained the eccentricity e (minor axes/major axes) of the largest cloud. This is shown in Fig. 4 (category 3: Fig. 4c, $e = 0.78$; Fig. 4d, $e = 0.87$). The eccentricity of the largest contiguous cloud cell for category 1 is equal to 0.52 (Fig. 4a) and 0.58 (Fig. 4b). Nevertheless, for the smaller cells in all categories the method detects high eccentricity. The number of contiguous cells occupying a fraction of the area Ac $> 0.3\%$ (an arbitrary threshold of Ac) is equal to 10 and 15 in the examples shown in Fig. 4a and Fig. 4b, respectively, and to 3 and 1 in the examples shown in Fig. 4c and Fig. 4d, respectively.
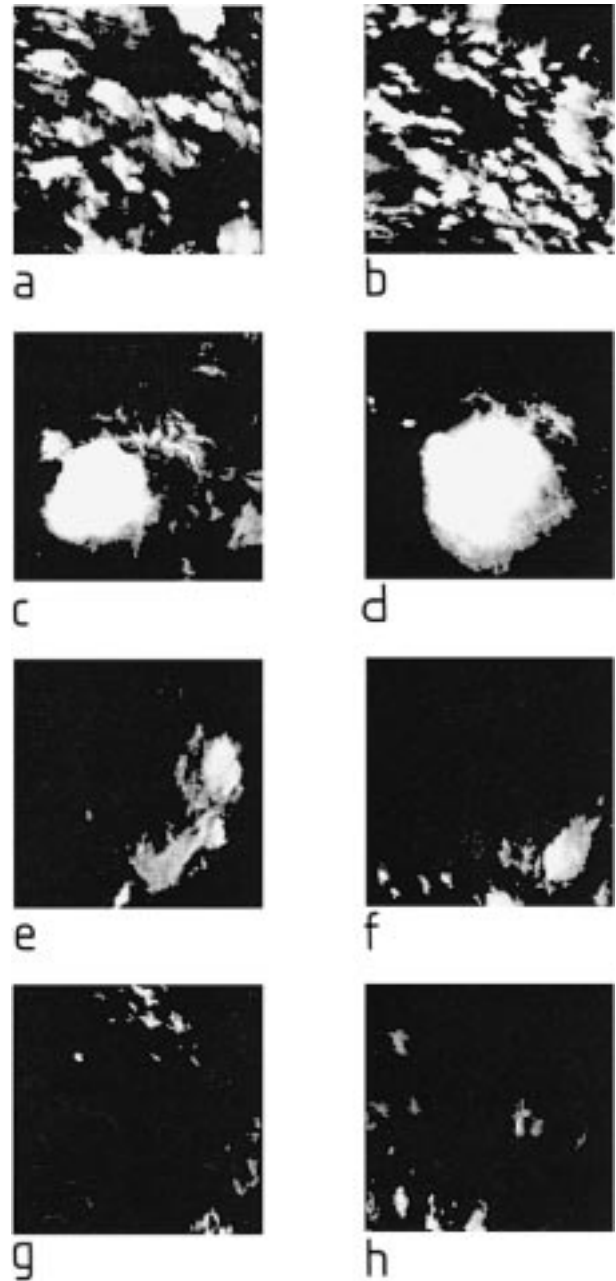

FIG. 4. Some examples of satellite IR images belonging to categories 1 [(a) and (b)], 3 [(c) and (d)], 4 [(e) and (f)], and 5 [(g) and (h)] for the control series: (a) southeast Brazil, day 25, 2200 UTC; (b) Amazon Basin, day 24, 2000 UTC; (c) Paraguay, day 25, 0100 UTC; (d) Paraguay, day 25, 0400 UTC; (e) southeast Brazil, day 24, 1100 UTC; (f) northern Brazil, day 25, 1500 UTC; (g) southeast Brazil, day 25, 1500 UTC; and (h) northern Brazil, day 24, 1000 UTC.

The general properties of such categories will be discussed based on the simulations of the circles.

Since box dimensions increase as the fraction of covered pixels is increased, it is interesting to understand what happens to D_b and D_g if the object is divided but
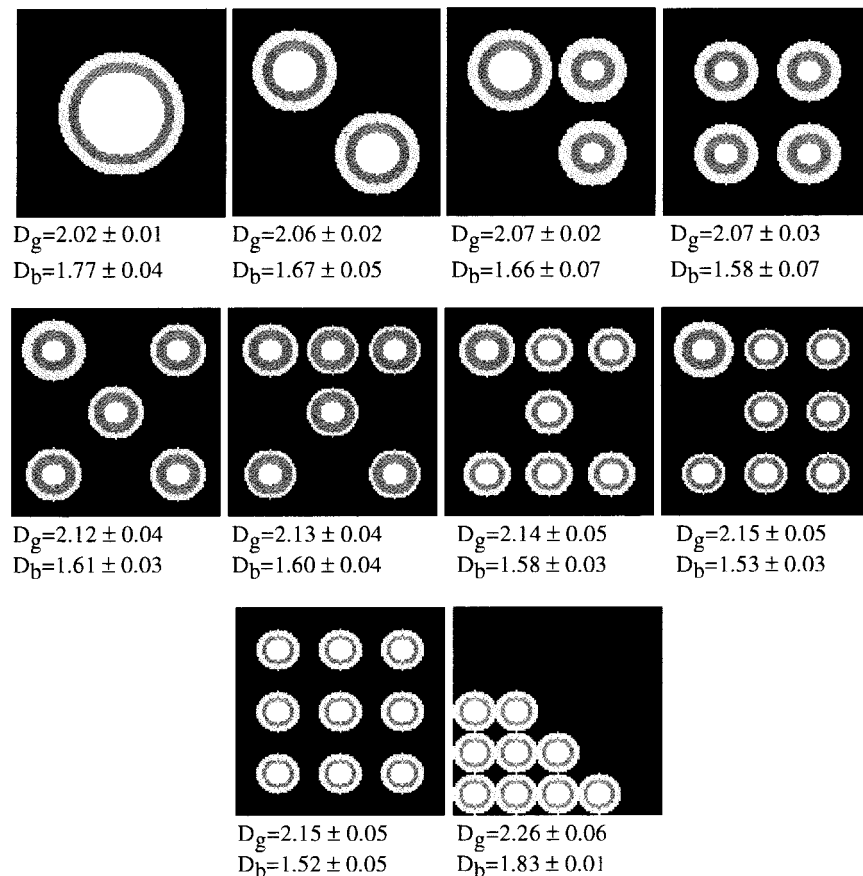
$D_g = 2.02 \pm 0.01$
$D_b = 1.77 \pm 0.04$

$D_g = 2.06 \pm 0.02$
$D_b = 1.67 \pm 0.05$

$D_g = 2.07 \pm 0.02$
$D_b = 1.66 \pm 0.07$

$D_g = 2.07 \pm 0.03$
$D_b = 1.58 \pm 0.07$

$D_g = 2.12 \pm 0.04$
$D_b = 1.61 \pm 0.03$

$D_g = 2.13 \pm 0.04$
$D_b = 1.60 \pm 0.04$

$D_g = 2.14 \pm 0.05$
$D_b = 1.58 \pm 0.03$

$D_g = 2.15 \pm 0.05$
$D_b = 1.53 \pm 0.03$

$D_g = 2.15 \pm 0.05$
$D_b = 1.52 \pm 0.05$

$D_g = 2.26 \pm 0.06$
$D_b = 1.83 \pm 0.01$

FIG. 5. Increasing number of circles keeping constant the same equivalent area of a single circle with $r = 30$ pixels. The values of D_b and D_g and the respective standard deviation error are presented beneath each picture. The last two pictures show the same nine circles with different configurations in the space: the picture on the left shows circles distributed isotropically, and the one on the right shows the circles grouped in the left corner of the area.

the percentage of covered pixels is kept constant and the number of circles is changed. Figure 5 illustrates how this simulation was performed. The first panel in the upper left-hand corner shows a circle with an external radius equal to 30 pixels. The other panels in Fig. 5 were constructed by splitting the fraction of pixels of the first panel into n parts ($n = 2$–9). To obtain a close approximation to the fraction of covered pixels in the first panel, some of the other panels in Fig. 5 contain circles that do not all have the same radius. This is due to the integer approximation associated with the number of pixels in each radius. The results for D_b and D_g are shown in Fig. 5

Dimension D_b tends to decrease as the total fraction of covered pixels is divided into smaller circles. This behavior is opposite that observed when the fraction of pixels of a single object is increased over the area. On the other hand, D_g increases as the number of circles increases—the opposite behavior when compared with D_b. The behavior of D_g as the number of circles is increased is explained by the fact that for smaller scales (smaller Δs) $N(\Delta s)$ is increased because the number of

boxes with a nonzero difference of counts increases [see Eq. (4)]. On the other hand, D_b increases as the nonempty pixels are concentrated in a single circle because the number of empty squares for larger Δs increases, indicating lower values of $N(\Delta s)$ for these scales. This results in an increasing slope of the fitted straight line between $\log[N(\Delta s)]$ and $\log(\Delta s)$.

Another useful property of box dimension is illustrated in the last two frames of Fig. 5: D_b and D_g increase as the space is filled anisotropically. This behavior can be explained considering that for larger Δs, $N(\Delta s)$ is smaller when the structures are anisotropically distributed over the area, implying an increase in both D_b and D_g.

The images typical of category 4 (Figs. 4e,f) show some centralized structure, with no evident circular shape, shifted from the center of the area. The eccentricity of the largest cell in the two examples is equal to 0.54 and 0.63 (Figs. 4e and 4f, respectively). The images typical of category 5, at the other extreme, though apparently not so different from the category 4 images, seem to have no preferential portions of the area occupied by clouds (Figs. 4g,h). Compared to cat-
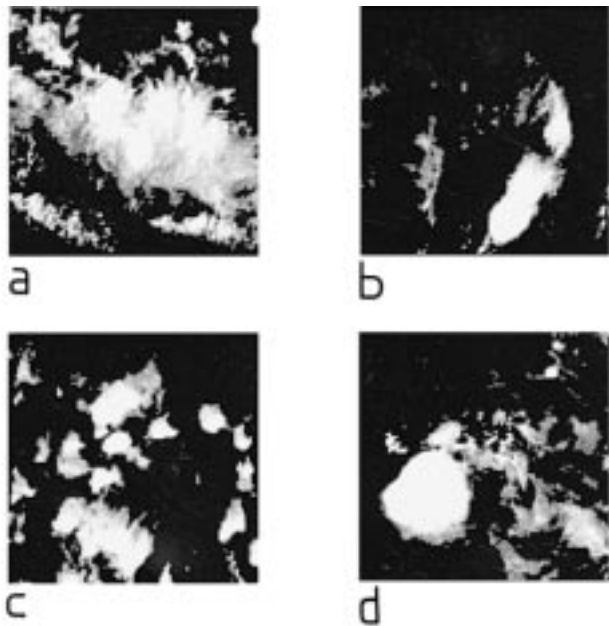
a   b

c   d

FIG. 6. The same as Fig. 4 but for category 2: (a) Paraguay, day 25, 1400 UTC; (b) southeast Brazil, day 24, 0800 UTC; (c) northern Brazil, day 24, 2300 UTC; (d) and Paraguay, day 24, 2300 UTC.

FIG. 7. The same as Fig. 3 but for the test series. Categories 6 and 7 are included in this case.

egory 1, these images with sparsely distributed clouds do not include, markedly, more cloudy pixels; that is, the cloud cells in category 5 are smaller and/or more sparse than those in category 1.

In Fig. 3 the images at the center of the cross, in the most densely filled category 2, show a more complex structure compared with the images shown in the other categories (Fig. 6). The shapes found in category 2 range from large structures, probably some of them representing the decaying MCC, to complex structures, whose edges indicate complex fractal structures rather than the well-behaved geometric patterns found in other groups.

These results illustrate four features of fractal box-dimension analysis:

1) Circular patterns in the range of scales analyzed from these images (see Table 1) are the most readily identifiable patterns that can be obtained from the fractal box-dimension properties since single, perfect circles have the box dimension D'_g approaching the topological dimension $D_T = 2$, as shown from simulated images. This explains the circular structures prevailing in category 3, in which the lowermost values of D'_g appear.

2) As a consequence of feature 1, the spread distribution of cloud cells with no circular organization is well distinguished by this method. This agrees with what was discussed with simulated circles in Fig. 5, in which it is demonstrated that D_g increases as the cloudy pixels are spread throughout the area with fixed cloud fraction.
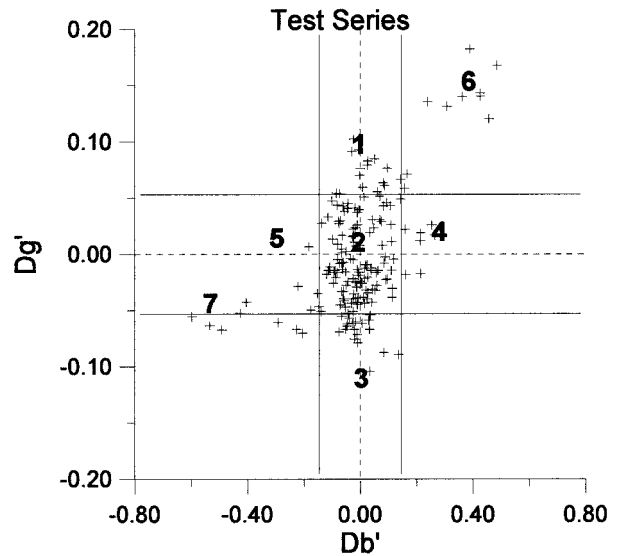
3) Anisotropic distribution of cloudy pixels over the area tends to have larger values of D_b, which explains the right side of the cross in Fig. 3.

4) The variance of D_b is well explained (about 87%) by Ac according to the second degree polynomial fit for the cloud scenes (see Table 2). Since we removed the tendency on Ac, only about 13% of variance remained is not explained by the polynomial fit. Figure 2 shows that D_b is in fact more sensitive to Ac than D_g.

2) THE TEST SERIES OF DAYS 20 AND 21

The same variables, D_b, D_g, and Ac, were computed for the test series to investigate the fractal properties of other independent series. Similar to the control series, the test series showed increasing values of box dimension with the increase of Ac. The polynomial fit is shown for each dataset, and the corresponding values of the coefficients are presented in Table 2.

The polynomial coefficients fitted for each curve are different from those of the control series, as well as the percentage of residuals explained by each fit (Table 2). We found that a 2° polynomial fit between D_g and Ac is more adequate than the linear one. These modifications led to differentiated averages and standard deviations for D'_b and D'_g, as indicated in Fig. 7. It must be noted that the lowermost branch of D'_g values is less evident in this case, while the lowermost branch of D'_b is more extensive. Moreover, another category of large values of D'_g and D'_b is found isolated in the upper-right corner of Fig. 7, and a new category of small values in the lower-left corner.

We verify that the test series shows the overall behavior found for the control series. Figure 8 shows one
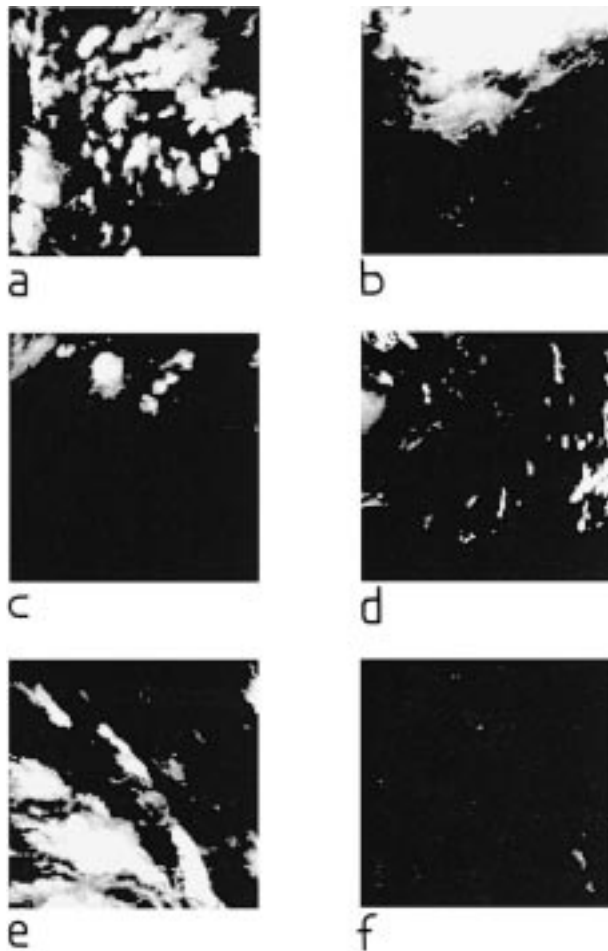
FIG. 8. The same as Fig. 4 but for the categories in Fig. 7. Panels (e) and (f) represent categories 6 and 7, respectively: (a) northern Brazil, day 20, 2000 UTC; (b) Paraguay, day 21, 2300 UTC; (c) Bolivia, day 20, 1200 UTC; (d) southeast Brazil, day 20, 1600 UTC; (e) Amazon Basin, day 21, 0100 UTC; and (f) Amazon Basin, day 20, 1300 UTC.

typical image for categories 1, 3, 4, and 5, illustrating such similarities. Note that categories 6 and 7 have been added in this case (see Fig. 7). The lowermost values of D'_g are related to circular (or semicircular), compact structures (Fig. 9b) and that the uppermost values of D'_b and D'_g are related to structures occupying almost half of the area in a highly anisotropic fashion. It is clear that the clouds appeared compacted in the lower-left corner of the area (Fig. 8e). The simulations shown in Fig. 5 illustrate this behavior: as the nine circles are compacted into the left corner of the area, both D_b and D_g are increased. Therefore, the results found here are consistent with those discussed in the preceding analysis, indicating that anisotropic distribution has higher D'_b and D'_g.

It is clear that for meteorological purposes it would be more useful to identify and separate those images that correspond to enhanced convection instead of iden-

tifying only similar fractal patterns regardless of how much total cloud flux of mass, heat, moisture, and so forth they represent. Therefore, the fraction of cloudy pixels in the area should be included in order to cluster fractal patterns of interest to meteorological studies.

4. Fractal patterns and convective activity

The results from the previous section suggest that the information of cloud fraction in the area should be included for the definition of fractal patterns in order to elucidate the differences among the convective intensities as well as their patterns. This information is included in both the fractal dimensions D_b and D_g and in a more straightforward property, the fraction of cloudy pixels (Ac). Another variable that may be included to investigate the intensity of the convection is the lowermost temperature (equivalent to the higher count, hereafter referred to as "C_{max}"), as well as the fraction of area occupied for these higher tops (hereafter referred to as "P_{max}"). Several authors (see, for example, Adler and Fenn 1979; Scofield 1987) have shown the importance of these parameters in the determination of convective activity and the likelihood of precipitation on the ground. In this section, we will discuss two statistical methods to classify similar patterns in the cloudiness based on such variables and in the fractal box dimensions: the principal components (PC) analysis and the cluster analysis.

a. The PC analysis

Each image of the four time series defined in the control series can be thought of as an individual observation that defines some patterns that could be related to others in the same time series and even in other time series. Our major premise is that such patterns can be defined by the variables D'_b, D'_b, Ac, C_{max}, and P_{max}. To perform the PC analysis correctly, the values of C_{max} and P_{max} were multiplied by defining a single variable, hereafter referred to as "Pxcm." This was done because the highest value of count does not vary substantially from image to image, but if the fraction of the coldest pixels is included, the differences between the images are emphasized.

Therefore, the control series is represented by the matrix

$$\mathbf{x} = \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{vmatrix},$$

where x_{ij} is one of the variables described above. The number of variables p is equal to 4 (D'_b, D'_g, Ac, Pxcm), and the number of observations n is equal to 174 images.

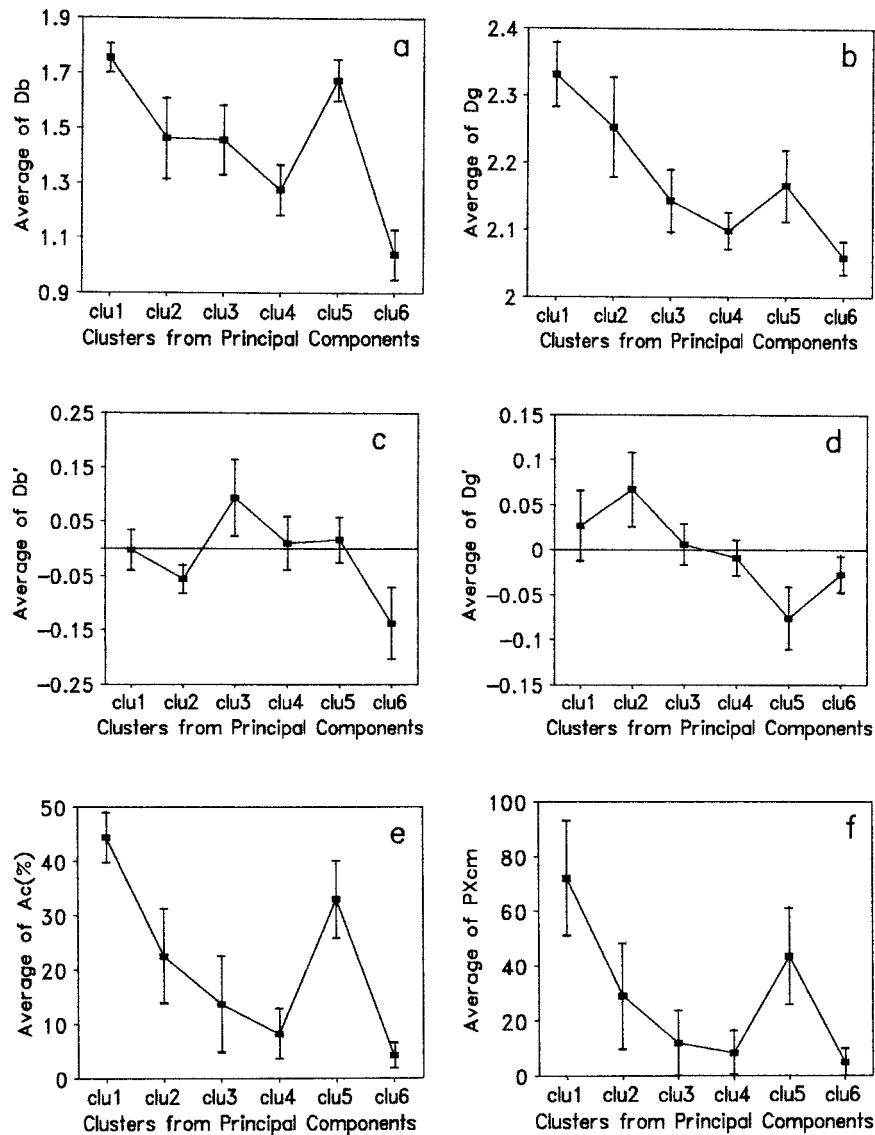A correlation matrix is generated from the matrix $\mathbf{x}$,

FIG. 9. Average value for each cluster obtained from PC analysis for the control series: (a) D_b, (b) D_g, (c) D_b', (d) D_g', (e) Ac, and (f) Pxcm.

which yields p eigenvalues λ_j. Each eigenvalue is related to a corresponding eigenvector with p elements. They represent the new base, and each component explains a variance according to the related eigenvalue. The resulting weights of the four PCs found and the variance explained by each one are shown in Table 3. The weights in each component can be thought of as representing the *relative* weight and variation of the variables that describe the pattern of each principal component. Table 3 also illustrates the eigenvalues λ_j with the corresponding shift $\delta\lambda$, according to North et al. (1982). It indicates that $n = 174$ samples resolve λ_1, λ_2, λ_3, and λ_4.

The greatest absolute weights (those values corresponding to at least 70% of the major weight found)

are boldface in Table 3 in order to identify the variables that have the greatest contribution in each PC. This procedure indicates that the first PC has its variance explained mainly by the relative high weight on Ac and Pxcm. The positive sign of those variables indicates that they increase (or decrease) in the same "direction" in the first component. The second PC, on the other hand, has its greatest weight on D_b' and D_g', with the opposite sign.

The third PC has the main contribution in variables D_b' and D_g', with the same sign, indicating that the two variables fluctuate in the same direction on these orthogonal axes. The fourth PC, which explains only about 3.0% of the total variance, has been considered meaningless for our purposes.

TABLE 3. The weights of variables on the eigenvectors of principal components and the percent of variance explained for each component.

	Weights of variables on principal components (PCs)				Percent variance	Cumulative percent	Eigenvalue $\lambda \pm \delta\lambda$
PC no.	Ac	Pxcm	D_b'	D_g'			
1	**0.7050**	**0.7067**	−0.0495	0.0336	47.03	47.03	1.881 ± 0.202
2	−0.0927	−0.0390	**−0.7030**	**0.7093**	29.09	76.12	1.164 ± 0.124
3	−0.0038	0.0199	**0.7093**	**0.7046**	20.88	97.00	0.835 ± 0.090
4	**0.7060**	**−0.7071**	−0.0137	0.0375	3.00	100.00	0.120 ± 0.013

Therefore, we can conclude that the first three PCs can define six main patterns, if we assume that the images can have a high magnitude on some of these PCs, with a positive or negative value. We can define two opposite patterns with high relative magnitude on the first PC, two opposite patterns with high relative magnitude on the second PC, and so forth. Thus, the verification of the PC analysis should be carried out through the analysis of the subset of images of the control series corresponding to the six main patterns found. This was performed through an automated clustering analysis.

b. The clustering model

1) THE CONSTRUCTION OF CLUSTERS FROM PC ANALYSIS

The clustering method provides the most adequate process of grouping similar data in a subset (see, e.g., Kaufman and Rousseuw 1990). Based on Euclidean distances, with standardized data, the clustering method is able to find, from an objective procedure, the similar patterns represented by the eigenvectors obtained from the PC analysis. The following steps were performed from the results of the control series.

1) From the matrix of expansion coefficients of the PC analysis, we chose the six main ''seeds'' through the relative value of each of the first three PCs, shown in Table 4. Only the first three PCs were considered because they represent almost 97% of the accumulated explained variance.
2) From those seeds, a clustering analysis was applied to the whole set of images for the control series in order to find the best correlated images according to their patterns, as represented by the PCs.
3) The subsets of the images obtained from the clus-

TABLE 4. Seeds used as centroids of clustering analysis.

	Expansion coefficients		
Cluster	PC1	PC2	PC3
1 (PC1 > 0)	**3.4114**	0.1168	0.6297
2 (PC2 > 0)	0.0416	**1.8044**	0.5517
3 (PC3 > 0)	−0.4836	−0.4294	**1.0383**
4 (PC1 < 0)	**−1.6832**	−0.4832	−0.1160
5 (PC2 < 0)	0.5766	**−2.4582**	−0.5789
6 (PC3 < 0)	−1.2786	0.4442	**−1.6983**

tering analysis were investigated according to the average properties of the variables included in the PC analysis and according to the patterns they represent.

Figure 9 illustrates the average values for each variable included in the PC analysis, according to the clusters defined before. Clusters 1 and 5 presented the greatest values of Ac, Pxcm, and D_b. Note that D_g has its greatest value in cluster 1, decreasing toward cluster 6, with a peak in cluster 5. This behavior is similar to those found for Ac and Pxcm, though the peak in cluster 5 exceeded the one in cluster 3. Dimension D_b' showed a marked maximum in cluster 3 and its lowest value in cluster 6. On the other hand, D_g' showed a peak in cluster 2 and its lowest value in cluster 5. In general, opposite behaviors of the variables are expected between clusters 1 and 4, 2 and 5, and 3 and 6 since these pairs of clusters were obtained through asymmetric properties described by the PC analysis.

Figure 10 displays the results of clustering analysis for the control series showing some typical images belonging to each cluster. The layout of the images in Fig. 10 was chosen to emphasize the differences between the clusters seeded with opposite signs of PC. The properties of such images can be summarized as follows.

- Cluster 1: The patterns of this cluster resemble those of the most enhanced convection, spreading out for the studied area (Figs. 10a,b). Some decaying but massive MCC images were found in this group. The shape defined by count 185 shows noncircular patterns. The high gray levels indicate the intensity of convection in the area.
- Cluster 2: The pattern of the images in this cluster was defined mainly by the high values of D_g'. This cluster grouped mainly category 1 images. The negative weight of D_b' on PC2 (the second eigenvector) seems to reinforce the property discussed before in which the partition of Ac in many isotropically distributed cells showed relatively high values of D_g' and low values of D_b' (Figs. 10e,f).
- Cluster 3: (Compared to the images in clusters 1 and 2). The images in this cluster show fewer cloud pixels, and the pixels tend to group together (Figs. 10i,j). These images present an anisotropic distribution, revealing the behavior discussed before, in which such distribution of cloudy pixels in the area is described
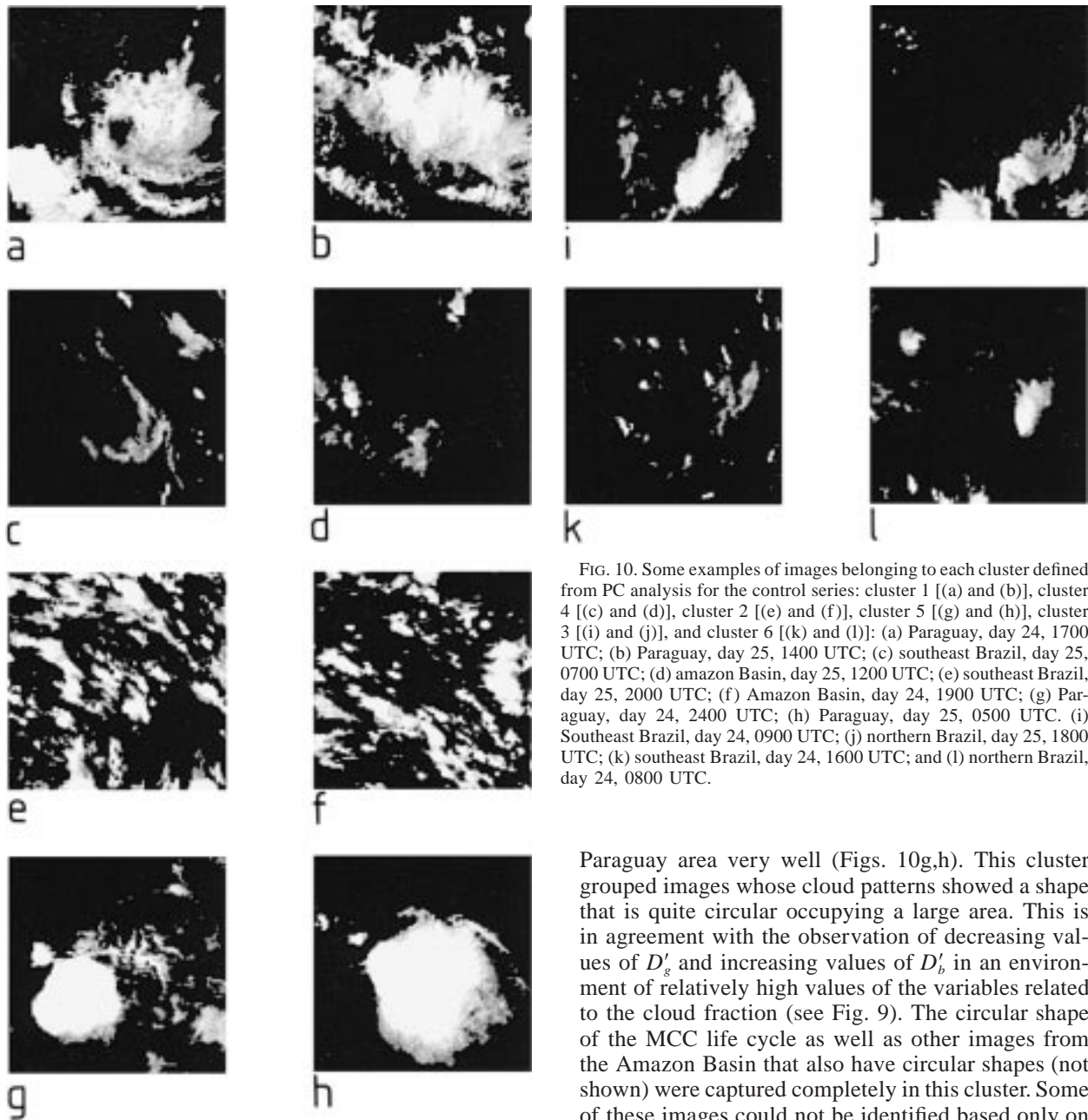
a

b

i

j

c

d

k

l

Fig. 10. Some examples of images belonging to each cluster defined from PC analysis for the control series: cluster 1 [(a) and (b)], cluster 4 [(c) and (d)], cluster 2 [(e) and (f)], cluster 5 [(g) and (h)], cluster 3 [(i) and (j)], and cluster 6 [(k) and (l)]: (a) Paraguay, day 24, 1700 UTC; (b) Paraguay, day 25, 1400 UTC; (c) southeast Brazil, day 25, 0700 UTC; (d) amazon Basin, day 25, 1200 UTC; (e) southeast Brazil, day 25, 2000 UTC; (f) Amazon Basin, day 24, 1900 UTC; (g) Paraguay, day 24, 2400 UTC; (h) Paraguay, day 25, 0500 UTC. (i) Southeast Brazil, day 24, 0900 UTC; (j) northern Brazil, day 25, 1800 UTC; (k) southeast Brazil, day 24, 1600 UTC; and (l) northern Brazil, day 24, 0800 UTC.

e

f

g

h

for high values of D_b' and D_g', which is in agreement with the observed weight of D_b' and D_g' in PC3 (see Table 3).

- Cluster 4: This cluster represents the images with properties that are asymmetric to cluster 1 (see Table 4). It grouped preferably those images that had low fraction of clouds, were sparsely distributed, or had no clear organization over the area (Figs. 10c,d). It is a result of the lower weight on Ac and Pxcm (Table 3).
- Cluster 5: With characteristics opposite those of cluster 2, this cluster identifies the MCC cycle over the

Paraguay area very well (Figs. 10g,h). This cluster grouped images whose cloud patterns showed a shape that is quite circular occupying a large area. This is in agreement with the observation of decreasing values of D_g' and increasing values of D_b' in an environment of relatively high values of the variables related to the cloud fraction (see Fig. 9). The circular shape of the MCC life cycle as well as other images from the Amazon Basin that also have circular shapes (not shown) were captured completely in this cluster. Some of these images could not be identified based only on the analysis of the standard deviation of D_b' and D_g', as discussed in section 3. Some of the images were found belonging to category 2 (Fig. 3), which represents the values below one standard deviation or closer to the mean values. Only with the aid of the variables that include information on cloud quantity and shape (Ac, Pxcm, Db', and Dg') could one find this important subset in the images of the control series. Therefore, this cluster represents the highly organized, intensive convection that is clearly opposite the more sparse distribution of cloudiness found in cluster 2 (Fig. 10).

- Cluster 6: This cluster has properties asymmetric to cluster 3 (Table 4). Therefore, it represents both im-
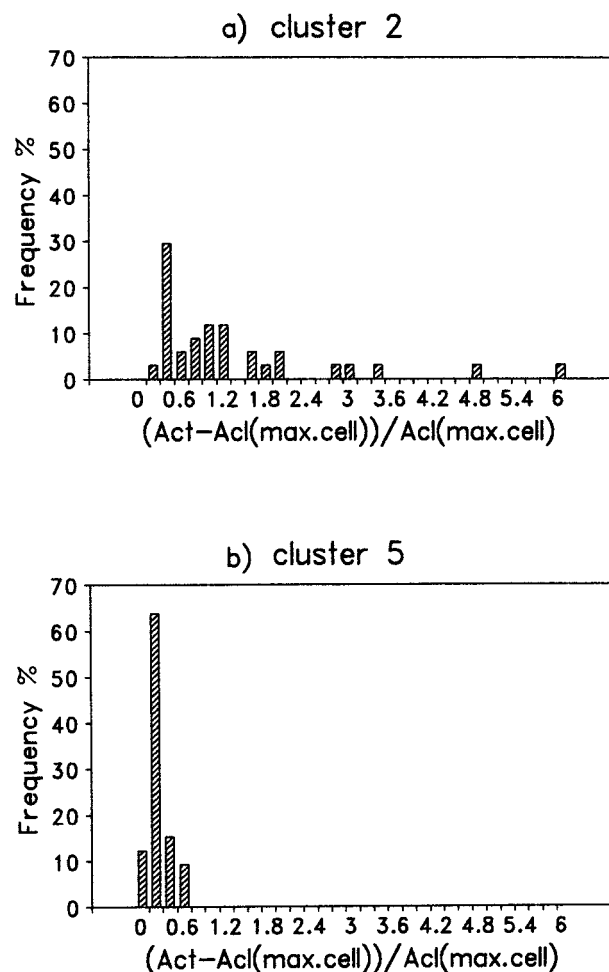
a) cluster 2

b) cluster 5

Fig. 11. Frequency distribution of α for clusters 2 and 5. Numbers at horizontal axes indicate the interval maximum.

ages that have fewer clouds (low values of PC1) and those that have clouds that are not grouped (low values of PC3). This cluster, of course, shares some properties with cluster 4, but this was considered important because it is able to select the less expressive subset of convective activity in the area from the others.

Examining clusters 2 and 5, again, we defined a parameter:

$$\alpha = \frac{\text{Act}(> 0.3\%) - \text{Acl(max cell)}}{\text{Acl(max cell)}},$$

where Act($>0.3\%$) is the total fraction of clouds with an area greater than 0.3%, and Acl(max cell) is the area fraction of the largest cloud in the image. The choice of 0.3% was arbitrary. Figure 11 shows the frequency distribution of α for clusters 2 and 5. The frequency distribution indicates how much remains of the total Act after removing Acl(max cell) and how much this residue represents the Acl(max cell). For instance, if we find only a single cell with Acl $> 0.3\%$ in the image, Act

is equal to Acl (max cell), and therefore, the ratio equals zero. Cluster 5 has a distribution concentrated in the values lower than 0.8 with a sharp peak around 0.2 and almost 13% of the data with zero value of α, which means just one cell. On the other hand, the broadening in the frequency distribution for cluster 2 (large right tail) indicates that the total Act is composed of several cells with Acl $> 0.3\%$. It is worth noting that some images in cluster 2 present large, noncircular, contiguous cells with complex noncircular shapes.

Figure 12 shows the distribution of the number of contiguous cells with Acl $> 0.3\%$ for clusters 1, 2, and 5, illustrating how the total Act is distributed in individual, contiguous cells over the area. Figures 11 and 12 emphasize the differences in mesoscale cloud patterns shown before and how the box dimensions can summarize properties, such as the number of cells, the total cloud area, and the cloud patterns distribution and shape.

These results suggest the following:

- Clusters 1 and 5 represent the most intense phenomenon regarding the convective activity over the area, which can be differentiated from their shape.
- Cluster 2 could be considered the middle of an idealized scale of convective intensity. It is between the highly intensive homogeneous distribution of cloudy pixels in cluster 1 and the random and sparse distribution in cluster 6.
- Cluster 3, whose images show some kind of grouped structures, can be considered the middle of the same idealized convective intensity scale but between the highly circular and massive shapes in cluster 5 and, perhaps, the small amount of cloudiness in cluster 4. It should be noted, however, that large-scale and mesoscale meteorological fields related to these clusters should be considerably different, and therefore, could define differentiated patterns as well.

2) The generalization of the clustering model

Since the control series samples the most common patterns of cloudiness in the summer over the tropical and subtropical continental areas, we may expect that there is a high probability of finding images that belong to the clusters defined above for other samples obtained close to, or in, the same areas.

Based on such observations, a generalized clustering model was defined by seeding the clusters with the average value of the variables in each cluster previously obtained (Table 5). In this way, the clustering analysis is simplified into a single step.

3) Results of the generalized clustering model for the test series

The result of the generalized clustering model for the test series was obtained assuming the same six seeds as defined in Table 5 and seen in Fig. 13.
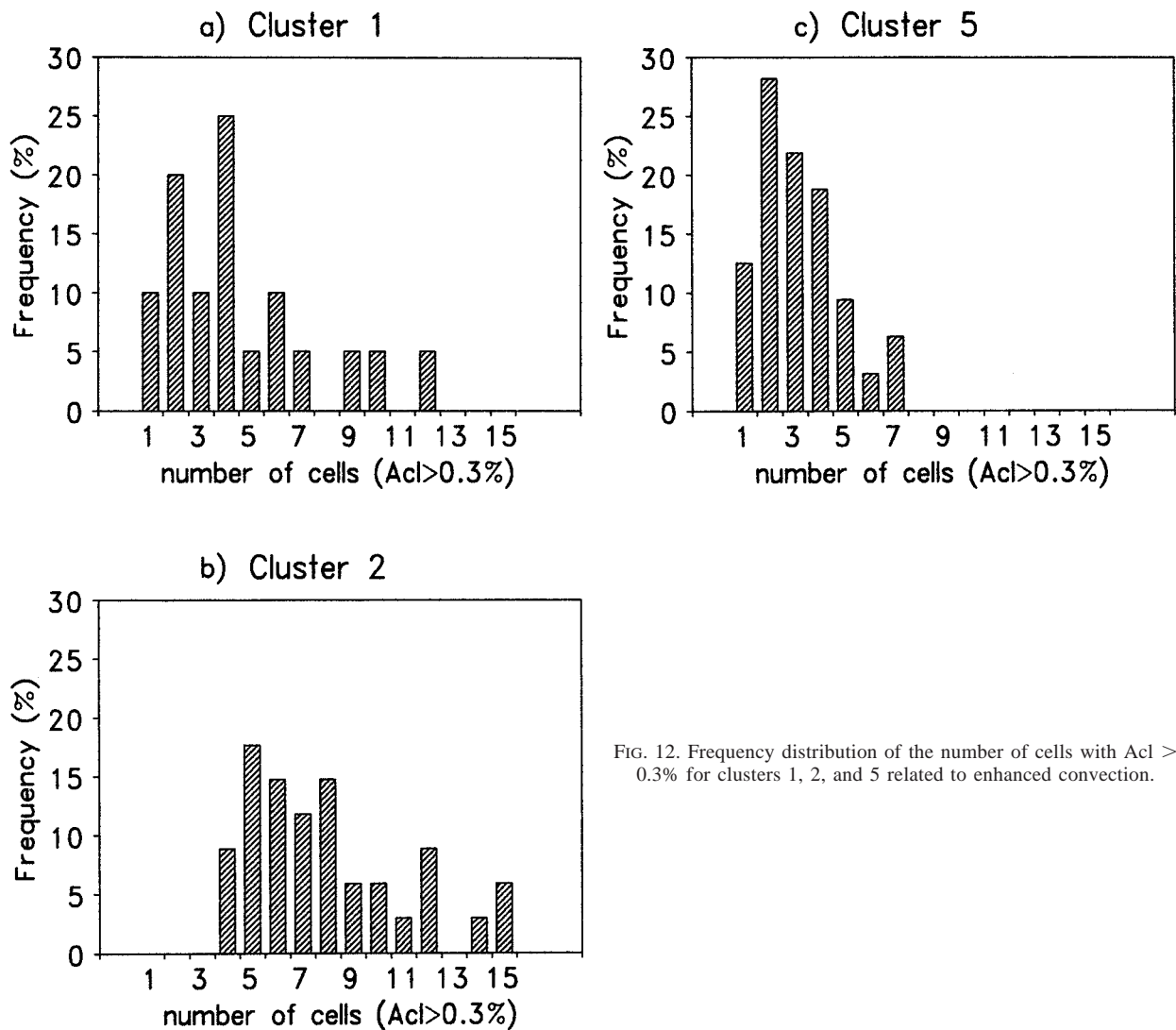
a) Cluster 1

c) Cluster 5

b) Cluster 2

FIG. 12. Frequency distribution of the number of cells with Acl > 0.3% for clusters 1, 2, and 5 related to enhanced convection.

It can be verified from the typical images in each cluster that the same patterns found in the control series also appear in the test series. 1) Cluster 1 includes images that represent the enhanced convection that covers a larger fraction of the area. Cluster 2 includes images with cloud cells that are distributed sparsely and uniformly over the entire area. Cluster 3 includes images with high values of D'_b and D'_g, as discussed before. It represents images with anisotropic distribution of

cloudy pixels over the area. Clusters 4 and 6 represent, again, images with fewer clouds in the area, while cluster 5 groups images with an almost circular, massive shape. The cluster 5 images represent part of the cycle of an MCC that developed in the boundaries of the studied area. As this MCC appears with some internal (relative to the count 185) globular structures, but with a high quantity of cloudy pixels, its images became representative of cluster 1 (not shown). This can be understood if we take into account that some significant change in D'_g is found when new "cells" are included inside the most external circle, as shown in Fig. 14. These new cells are noncentered if the center is taken as the symmetric axis of the most external circle. In this case, $N(\Delta s)$ increases for the smaller Δs as the number of circles increases in the area. As the structures appear symmetrically distributed over the area, the errors in D'_g determination become smaller.

It is important to comment that the clustering analysis

TABLE 5. Seeds of the generalized clustering model.

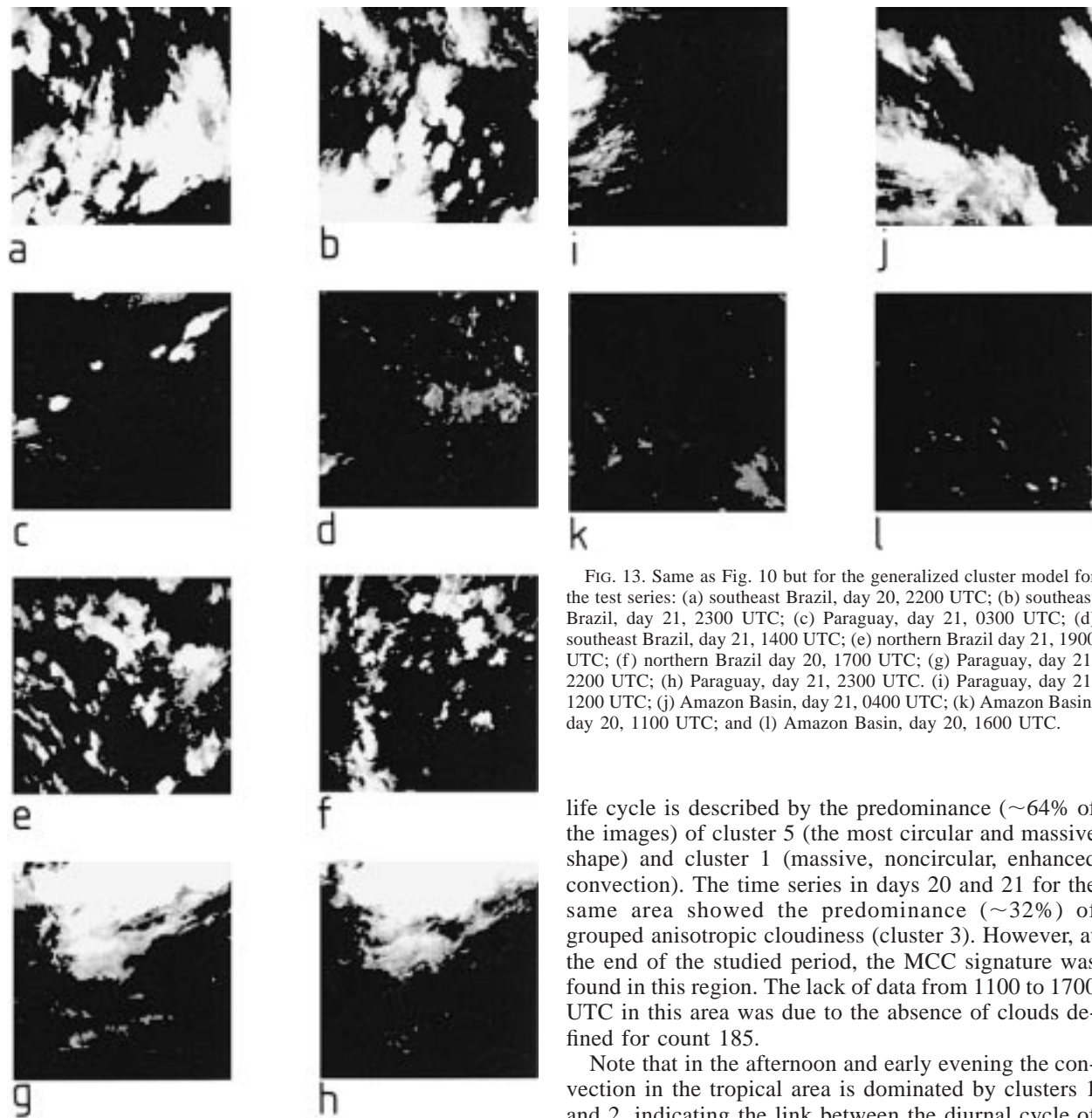| Cluster	D_b	D_g	Ac	Pxmc	D'_b	D'_g
1	1.753	2.331	44.375	72.071	−0.003	0.027
2	1.461	2.253	22.506	28.872	−0.057	0.067
3	1.456	2.143	13.620	11.808	0.093	0.006
4	1.274	2.098	8.193	8.269	0.010	−0.009
5	1.674	2.166	32.927	43.395	0.016	−0.076
6	1.039	2.059	4.194	4.776	−0.138	−0.027

FIG. 13. Same as Fig. 10 but for the generalized cluster model for the test series: (a) southeast Brazil, day 20, 2200 UTC; (b) southeast Brazil, day 21, 2300 UTC; (c) Paraguay, day 21, 0300 UTC; (d) southeast Brazil, day 21, 1400 UTC; (e) northern Brazil day 21, 1900 UTC; (f) northern Brazil day 20, 1700 UTC; (g) Paraguay, day 21, 2200 UTC; (h) Paraguay, day 21, 2300 UTC. (i) Paraguay, day 21, 1200 UTC; (j) Amazon Basin, day 21, 0400 UTC; (k) Amazon Basin, day 20, 1100 UTC; and (l) Amazon Basin, day 20, 1600 UTC.

for the control series from the seeds described in Table 5, according to our expectations, resulted in clusters quite similar to those obtained from the PC analysis (not shown due to redundancy).

4) THE LIFE CYCLE OF CLOUDINESS

Figure 15 summarizes the time evolution of the classification into the six clusters for each time series studied for the control and test series. The differences between the control and series over the Paraguay area can be seen in this figure. During days 24 and 25, the MCC life cycle is described by the predominance (~64% of the images) of cluster 5 (the most circular and massive shape) and cluster 1 (massive, noncircular, enhanced convection). The time series in days 20 and 21 for the same area showed the predominance (~32%) of grouped anisotropic cloudiness (cluster 3). However, at the end of the studied period, the MCC signature was found in this region. The lack of data from 1100 to 1700 UTC in this area was due to the absence of clouds defined for count 185.

Note that in the afternoon and early evening the convection in the tropical area is dominated by clusters 1 and 2, indicating the link between the diurnal cycle of the convection over the tropical areas (Silva-Dias et al. 1987) and the observed patterns of clouds. For the MCC life cycle on days 24 and 25 note that most circular patterns prevail late at night and in the early morning, in agreement with the observations of Velasco and Fritsh (1987).

5. Summary and conclusions

The analysis of the properties of box dimensions obtained from satellite cloud images from two different periods in several areas during the South American summer and elucidated from the simulated circles can be summarized as follows.
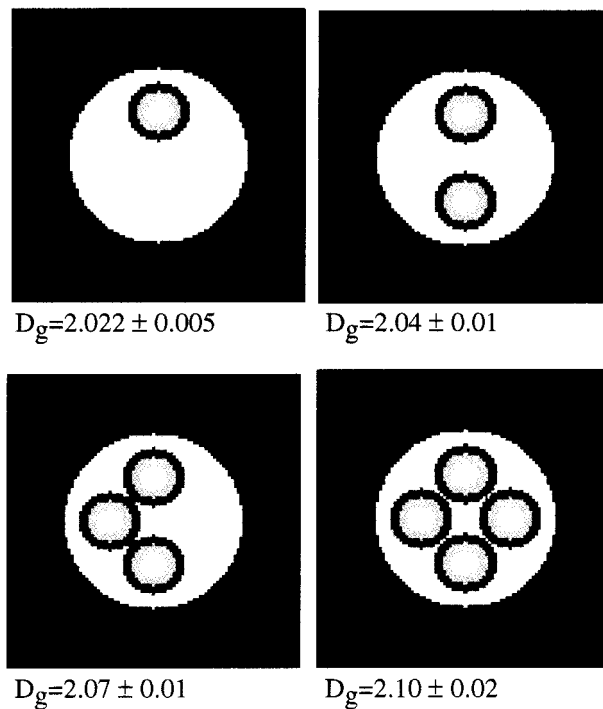
$D_g = 2.022 \pm 0.005$ $D_g = 2.04 \pm 0.01$

$D_g = 2.07 \pm 0.01$ $D_g = 2.10 \pm 0.02$

FIG. 14. Effect of increasing the number of internal circles with noncentered axes.

1) Dimensions D_b and D_g tend to increase as the fraction of objects increases over a fixed area.
2) Dimension D_b tends to increase, while D_g tends to decrease, as the cloudy pixels are grouped with fixed cloud fraction in the area.
3) Dimension D_g is very sensitive to the way the cloudy pixels are distributed. For some cloudy pixels occupying a single circular area, D_g assumes smaller values than if the cloudy pixels were distributed into several cells.
4) Anisotropic distribution of cloudy pixels over the area yields higher values of D_b and D_g. No conclusive behavior of D_g is found if new gray levels are added to a single cell.
5) Dimensions D_b and D_g depend on anisotropy due to the objects' distribution over the area. Some tests with other Euclidean shapes, such as ellipses (not included), showed that box dimensions can distinguish between circles and ellipses with equivalent area only if the ellipses have an eccentricity (minor axes/major axes) of less than 0.2. In this case the ratio $D_b$circle/$D_b$ellipse is definitely greater than 1.1. Otherwise, the ratio fluctuates around 1 ± 0.05.

Six main fractal cloud patterns became evident from a PC analysis since the first three components explained almost 97% of the variance of the patterns in the control series. These patterns were found from the positive and negative high weight on the first, second, and third PC.

The two main sets of images related to enhanced con-vection have been classified into the well-organized circular structures of the MCCs (cluster 5) and the most widespread noncircular convection (cluster 1), based on the high negative weight on the second PC and the high positive weight on the first PC, respectively. The identification of the MCC life cycle from the clustering model was a result of the differentiation between the geometry of the massive circular structures, with fractal dimension tending to the topological dimension $D_T = 2$, and the massive, widespread structures related to higher values of Ac and fractal dimensions D_g and D_b, indicating increasing complexity of the shape. The fraction of cloudiness distributed in a larger number of cells and/or the presence of cells with complex edges increases the values of fractal dimensions and can be considered the ''opposite'' geometry of the well-organized circular structures of the MCCs (cluster 2).

Less expressive convective activity was also classified (clusters 3, 4, and 6) and is necessary to ensure the separation between moderate and light convective activity. One of the patterns represents the anisotropy of cloud distribution over the area (cluster 3).

The generalization of the clustering model was performed throughout the investigation of the existence of the same six clusters in the test series. Similar patterns were found in the test series, validating the results found for the control series. Nevertheless, the new clusters revealed a differentiated proportion of images compared to the control series, indicating that the proposed clustering model can be generalized to any time series of images for the studied areas. The life cycle of convection can be captured from the clustering analysis. It shows that massive circular shapes require dynamically organized structures not common for most time series analyzed. Furthermore, the clustering analysis captured the diurnal cycle of convection and of cloud patterns.

From this analysis we demonstrated that the fractal box dimension can be a powerful tool in analysis of satellite images since it can discriminate a large range of fractal patterns in mesoscale convective systems. The signature of massive circular patterns in the behavior of the fractal box dimension can improve the diagnostics of such systems in an automatic way, without any other source of information. This may permit an improvement in climatological analysis over a wide range of time- and space scales since this analysis can be applied to large areas to monitor tropical thunderstorms and hurricanes.

Radiation versus cloud parameterizations have an implication on the surface energy budget and consequently on the life cycle of modeled precipitation systems (Betts and Ridgway 1988). Therefore, the patterns defined by D_b and D_g and the other parameters used to describe convective activity in the cluster method may also be a useful way to address the problem of subgrid-scale distribution of cloudiness in numerical models.
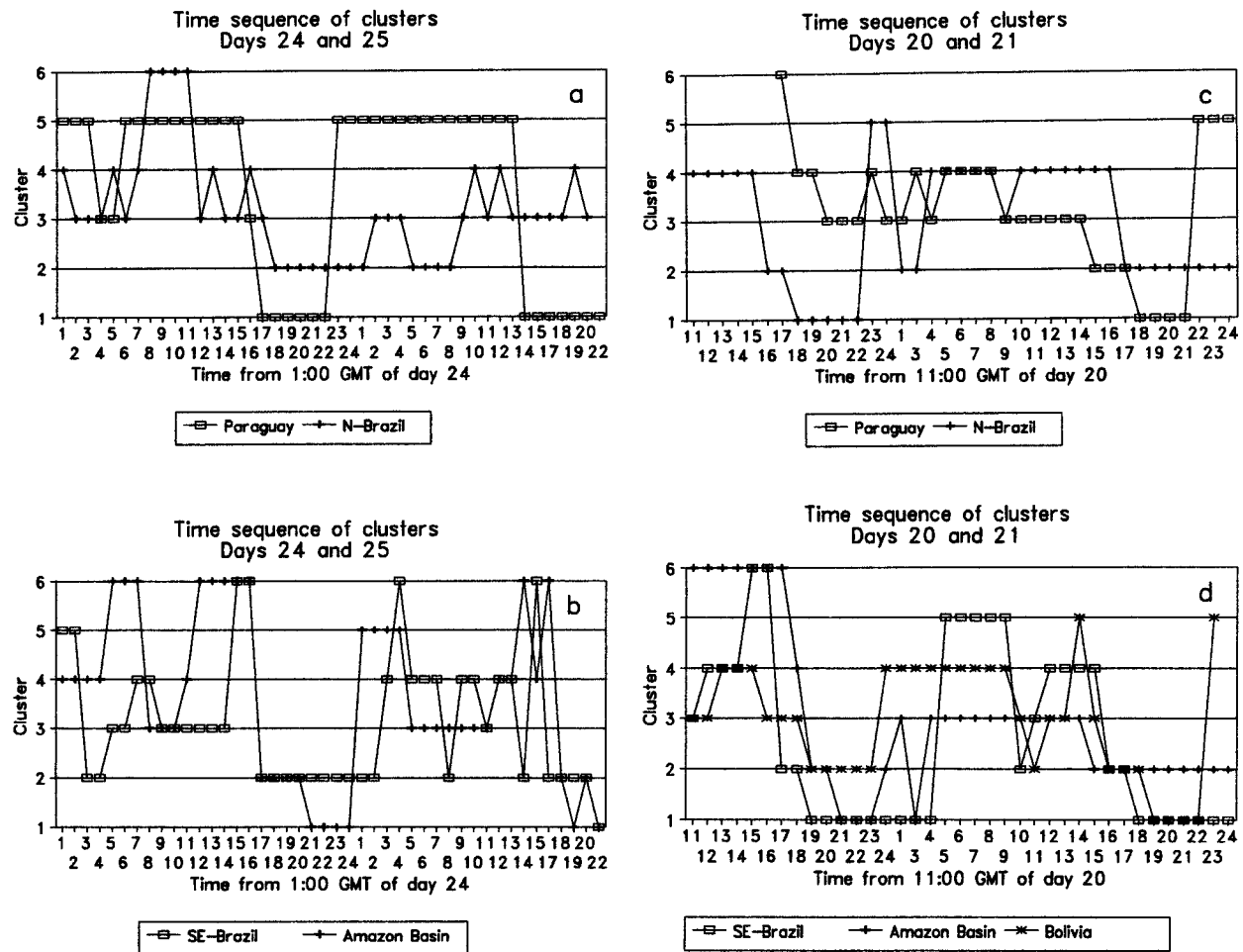
FIG. 15. Time sequence of clusters for each studied area: (a) control series for Paraguay and northern Brazil, (b) control series for southeastern Brazil and the Amazon Basin, (c) same as (a) but for the test series, and (d) same as (b) but for the test series.

REFERENCES

Adler, R. F., and D. D. Fenn, 1979: Thunderstorm intensity as determined from satellite data. *J. Appl. Meteor.* **18,** 502–527.
Albrecht, B. A., V. Ramanathan, and B. A. Boville, 1986: The effects of cumulus moisture transports on the simulation of climate with a general circulation model. *J. Atmos. Sci.,* **43,** 2243–2462.
Anthes, R. A., and D. Keyser, 1979: Tests of a fine-mesh model over Europe and the United States. *Mon. Wea. Rev.,* **107,** 963–984.
Betts, A. K., 1976: The thermodynamic transformation of the tropical subcloud layer by precipitation and downdraft. *J. Atmos. Sci.,* **33,** 1008–1020.
——, and W. Ridgway, 1988: Coupling of the radiative, convective, and surface fluxes over the equatorial Pacific. *J. Atmos. Sci.,* **45,** 522–536.
Cahalan, R. F., and J. H. Joseph, 1989: Fractal statistics of cloud fields. *Mon. Wea. Rev.,* **117,** 261–272.
——, W. Ridgway, W. J. Wiscombe, and T. L. Bell, 1994: The albedo of fractal stratocumulus clouds. *J. Atmos. Sci.,* **51,** 2434–2455.
Dudhia, J., and M. W. Moncrieff, 1987: A numerical simulation of

quasi-stationary tropical convection bands. *Quart. J. Roy. Meteor. Soc.,* **113,** 929–967.
Esbensen, S. K., and J. T. Wang, 1984: Heat budget analysis and the synoptic environment of GATE cloud clusters. Preprints, *15th Conf. on Hurricanes and Tropical Meteorology,* Miami, FL, Amer. Meteor. Soc., 455–460.
Essex, C., and M. A. H. Nerenberg, 1990: Fractal dimension: Limit capacity or Hausdorff-Dimension? *Amer. J. Phys.,* **58,** 986–988.
Frank, M. W., and C. Cohen, 1985: Properties of tropical cloud ensembles estimated using a cloud model and an observed updraft population. *J. Atmos. Sci.,* **42,** 1911–1928.
Gabriel, P., S. Lovejoy, D. Schertzer, and G. L. Austin, 1988. Multifractal analysis of resolution dependence in satellite imagery. *Geophys. Res. Lett.,* **14,** 1373–1376.
Gamache, J. F., and R. A. Houze, 1982: Mesoscale air motions associated with a tropical squall line. *Mon. Wea. Rev.,* **110,** 118–135.
Hack, J. J., and W. H. Schubert, 1986: Nonlinear response of atmospheric vortices to heating by organized cumulus convection. *J. Atmos. Sci.,* **43,** 1559–1573.
Hamburger, D., O. Biham, and D. Avnir, 1996. Apparent fractality emerging from models of random distributions. *Phys. Rev. E,* **53,** 3342–3358.
Johnson, R. H., 1976: The role of convective-scale precipitation downdrafts in cumulus and synoptic-scale interaction. *J. Atmos. Sci.,* **33,** 1890–1910.

Kaufman, L., and P. J. Roussew, 1990: *Finding Groups in Data: An Introduction to Cluster Analysis.* John Wiley and Sons, 340 pp.

Lovejoy, S., 1982: The area–perimeter relation for rain and cloud areas. *Science,* **216,** 185–187.

——, and D. Schertzer, 1990: Multifractals, universality classes and satellite and radar measurements of cloud and rain fields. *J. Geophys. Res.,* **95,** 2021–2034.

——, ——, and A. A. Tsonis, 1987: Functional box-counting and multiple elliptical dimensions in rain. *Science,* **235,** 1036–1038.

Machado, L. T., M. Desbois, and J. P. Duvel, 1992: Structural characteristics of deep convective systems over tropical Africa and the Atlantic Ocean. *Mon. Wea. Rev.,* **120,** 392–406.

Mandelbrot, B. B., 1977: *Fractals: Form, Chance and Dimension.* W. H. Freeman and Co., 365 pp.

——, 1982: *The Fractal Geometry of Nature.* W. H. Freeman and Co., 468 pp.

Mapes, B. E., and R. A. Houze, 1993: Cloud clusters and superclusters over the oceanic warm pool. *Mon. Wea. Rev.,* **121,** 1398–1415.

Marshak, A., A. Davis, W. Wiscombe, and R. Cahalan, 1997: Scale invariance in liquid water distributions in marine stratocumulus. Part II: Multifractal properties and intermittency issues. *J. Atmos. Sci.,* **54,** 1423–1444.

North, G. R., T. L. Bell, and R. F. Cahalan, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 699–706.

Ogura, Y., and H. R. Cho, 1973: Diagnostic determination of cumulus cloud population from observed large-scale variables. *J. Atmos. Sci.,* **30,** 1276–1286.

Olsson J., J. Niemczynowicz, and R. Bernsdtsson, 1993: Fractal analysis of high-resolution rainfall time series. *J. Geophys. Res.,* **98,** 23 265–23 274.

Peleg, S., J. Naor, R. Hartley, and D. Avnir, 1984: Multiple resolution texture analysis and classification. *IEEE Trans. Pattern Anal. Machine Intell.,* **6,** 518–523.

Pentland, A. P., 1984: Fractal-based description of natural scenes. *IEEE Trans. Pattern Anal. Machine Intell.,* **6,** 661–674.

Schertzer, D., and S. Lovejoy, 1987: Physical modeling and analysis of rain clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res.,* **92,** 9693–9714.

Scofield, R. A., 1987: The NESDIS operational convective precipitation estimation technique. *Mon. Wea. Rev.,* **115,** 1773–1792.

Sengupta, S. K., R. M. Welch, M. S. Navar, T. A. Berendes, and D. W. Chen, 1990: Cumulus cloud field morphology and spatial patterns derived from high spatial resolution Landsat imagery. *J. Appl. Meteor.,* **29,** 1245–1267.

Silva-Dias, M. F., A. K. Betts, and D. E. Stevens, 1984: A linear spectral model of tropical mesoscale systems: Sensitivity studies. *J. Atmos. Sci.,* **41,** 1704–1716.

——, P. L., J. P. Bonatti, and V. E. Kousky, 1987: Diurnally forced tropical tropospheric circulation over South America. *Mon. Wea. Rev.,* **115,** 1465–1478.

Tessier, Y., S. Lovejoy, and D. Schertzer, 1993: Universal multifractals: Theory and observations for rain and clouds. *J. Appl. Meteor.,* **32,** 223–249.

Tsonis, A. A., 1992: *Chaos: From Theory to Applications.* Plenum, 274 pp.

Velasco, I., and J. M. Fritsch, 1987: Mesoscale convective complexes in the Americas. *J. Geophys. Res.,* **92,** 9591–9613.

Wielicki, B. A., and R. M. Welch, 1986: Cumulus cloud properties derived using Landsat satellite data. *J. Appl. Meteor.,* **25,** 261–276.

Yano, J. I., J. C. McWilliams, and M. W. Moncrieff, 1996: Fractality in idealized simulations of large-scale tropical cloud systems. *Mon. Wea. Rev.,* **124,** 838–848.