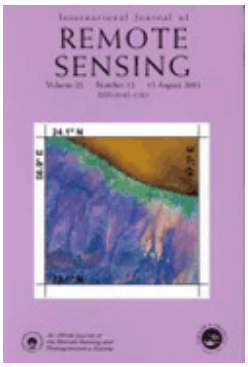
International Journal of Remote Sensing



Daytime cloud classification over South America region using multispectral GOES-8 imagery

Journal:	*International Journal of Remote Sensing*
Manuscript ID:	TRES-PAP-2009-0425
Manuscript Type:	Research Paper
Date Submitted by the Author:	07-Jul-2009
Complete List of Authors:	Bottino, Marcus; INPE, CPTEC Ceballos, Juan; INPE, CPTEC
Keywords:	CLASSIFICATION, CLOUDS
Keywords (user defined):	satellite, classification, cloudiness

Daytime cloud classification over South America region using multispectral GOES-8 imagery

MARCUS JORGE BOTTINO* and JUAN CARLOS CEBALLOS

Instituto Nacional de Pesquisas Espaciais - INPE
Center for Weather Forecast and Climate Studies - CPTEC
12630-000 Cachoeira Paulista - SP - Brazil

Abstract. Information about a minimal set of characteristic variables and types of scene was searched for GOES 8 Imager multispectral imagery over an extended area of South America, September 2002, 1600 UTC (near local noon over most part of the area). Thirteen variables were considered for each pixel: five of them described reflectance and brightness temperature in 4 channels, three variables assessed temperature difference related to channel 4; finally, five variables assessed local homogeneity (texture) in each channel. Thirty-two clusters were determined by a classification scheme ("dynamic cluster") based on minimal Euclidean distance. Factor analysis in principal components applied to cluster centroids shows that only five variables might be taken as non-redundant, namely reflectance in channel 1 and brightness temperature in channel 4 as well as their textures, together with difference between channels 5 and 4. Although factor analysis suggests to define about seven clusters (which in turn are consistent with image nephanalysis), principal components analysis makes evident that an objective minimal number of scenes is actually loosely defined but provides some useful criterions for definition of proper centroids, depending on user's convenience.

*Corresponding author. Email address: marcus.bottino@cptec.inpe.br

1. Introduction

The type and amount of cloud over a region is important information for meteorological analysis and for climate studies. A proper characterization provides important parameters for cloud representation in atmospheric general circulation models (Lau and Crane 1995, Norris and Weaver 2001, Gordon *et al*. 2005), in climate studies (Budyko 1969, Minnis and Harrison 1984, Philander *et al.* 1996), and in rainfall estimates (Uddstrom and Gray 1996, Germán *et al*. 2006). It is also useful for technological applications in agriculture and renewable energy generation (Chou 1991, Ceballos *et al.* 2004, among others). Last but not least, cloud masking is an important tool for remote sensing studies of earth surface such as biomass burning detection (França *et al*. 1995), sea surface temperature (Barton 1995) and vegetation index assessment (Defries and Townshend 1994).

Geostationary satellite imagery allows for higher frequency observation of extended areas from a constant position in space, thus becoming a current tool for continuous visual diagnostics (nephanalysis) of cloud type and cover. Images in the atmospheric window spectral band (11 μm) are the most usually inspected, looking at brightness temperature (or grey levels representative of emerging radiance) as well as observing image texture. Additional useful information is provided by channels associated to reflectance in visible spectrum or to absorption/emission by infrared vapour band. Concerning South America and neighbouring ocean area, two satellites are especially important. GOES satellite carries an Imager system providing information in five spectral intervals, while the Meteosat Second Generation SEVIRI system increases this number to 12 bands.

Considering the importance of GOES imagery for the South American region, the present paper analyzes some aspects of bulk information provided by the Imager. GOES 8 was in operation as GOES-E from 1996 with nadir at Amazon region (75ºW), being substituted by GOES 12 in March 2003. GOES 10 was located at GOES-W position (Pacific Ocean, 135º W) and was substituted by GOES 13 in September 2006. Following the EOPA Program (Earth Observation Partnership of the Americas), GOES 10 was shifted to 60ºW location in early 2007 with an expected life time until 2010. This new position allows for a continuous and specific monitoring of South American region. It is important to note that Imager channels in GOES 8 and 10 have the same spectral distribution (see details online at http://www.oso.noaa.gov/goes/goes-calibration/change-channels.htm). Therefore, cloud classification methods applied to GOES 8 Imager will remain useful for the next years. In addition, a long-term time series integrating similar spectral data base, provided by GOES 8 and GOES 10 imagery, will be highly useful for climatological studies over South America and neighbouring oceans. Only GOES 8 imagery characteristics will be considered hereafter.

Concerning cloud characterization, it is usual to present pixels in two-dimensional diagrams (brightness temperature *Tb* in thermal window channel vs. albedo or reflectance *R*) to illustrate physical features of cloud fields (Sèze and Rossow 1991, Porcú and Levizzani 1992, Rossow and Garder 1993). Nevertheless, these two variables yield partial information only. For instance, Cu and St fields presenting the same coordinates (*R*, *Tb*) exhibit quite different textures (Sèze and Desbois 1987, Welch *et al*. 1988). GOES imagery provides albedo in a narrow interval of visible solar spectrum (channel 1), and brightness temperature in four additional channels located in thermal infrared spectrum. The set of five-channel data describes remarkably different physical characteristics of surface, cloud and atmosphere. Considering diurnal scenes, VIS channel 1 allows to assess bidirectional reflectance for solar radiation while upwelling radiance detected in the channel 2 (at 3.9 μm) includes the contribution of reflected as well as emitted radiation (produced at upper wavelength tail of solar spectrum and lower tail of terrestrial thermal spectrum, respectively). This channel allows for detection of hot pixels associated to burning mass (Setzer and Verstraete 1994). The other channels inform about 1) radiance emitted in 6.7 μm (channel 3) water vapour band, which strongly absorbs and emits radiation, being representative of mid- and upper atmospheric levels; 2) radiance in the atmospheric window 10.8 μm (channel 4), hopefully providing information about lowest cloud or ground temperatures, but being affected by upper level vapour continuum absorption/emission; 3) radiance in a spectral range close to atmospheric window at 12 μm

(channel 5, allowing for correction methods in ground temperature assessment, and also cirrus detection).

Shortcomings of pixel analysis based on single channel information are evident. The most frequent problem is discrimination between Ci and developed Cu (both cold in atmospheric window), or between Cu fields and St cloud (both bright in VIS channel). Additional information provided by several channels, their differences and/or local texture provide valuable means of cloud discrimination. For instance, *Tb* difference between channels in 11 and 12 μm is roughly linear with precipitable water, being the base for the so-called split-window technique for ground temperature assessment, which can be used for AVHRR sensors in NOAA-n series as well as for GOES 8 and GOES 10 sensors (Barton 1995, Sun and Pinker 2003). This temperature difference is also affected by semitransparent cirrus due to ice-cloud spectral emissivity response (Inoue 1987, Giraud *et al*. 1997). The difference of *Tb* between channels 3 and 4 discriminates water vapor absorption in middle and low levels and is used to identify high well developed clouds (Tokuno and Tsuchiya 1993, Bottino *et al*. 2003). Also, *Tb* difference between radiances emitted at 3.9 and 11 μm is higher during night time in presence of fog and stratus (Ellrod 1995). A characteristic *Tb* difference may also be expected during day time. Uddstrom and Gray (1996) estimated the solar reflected parcel from 3.9 μm signal which can provide clouds microphysical information.

A variety of texture features may be assessed for an image segment (Welch *et al.* 1988, Baum *et al.* 1997). Gray-level vector methods used by Chen *et al*. (1989) for Landsat imagery assess texture aspects related with probability distributions of brightness over an image. This seems appropriate for defining general morphologic features of cloud field. However, these parameters are hardly associated to pixel-by-pixel analysis without a complex and expensive time computer effort.

This paper assumes that local information about an Imager pixel admits a set of simple numerical variables, such as signal intensity and local texture in five channels, as well as the difference between brightness temperatures in thermal channels. The pixel texture variables are assessed using only 3 × 3-pixel sets (eg.: Coakley and Baldwin 1984, Sèze and Desbois 1987). This implies in a reasonable array size (about 15 × 15 km), compatible with pixel resolution and local spatial variability of cloudiness.

Two basic issues are explored: 1) How many variables should be actually defined? It is important to avoid excessive computer time processing. Also, noisier information could be expected when redundant variables are included. 2) How many groups or types of objects are distinctly meaningful? Rough separation in a limited number of main types of objects should be possible: for instance, surfaces as land and water, and clouds as Cu, St, Ci, Cb. Nevertheless, a certain number of transitions between these main classes could be useful, especially in presence of a multi-layered (ML) cloud field.

To investigate these questions, factor analysis in principal components (PCA) is a useful tool to quantify variables relationships describing a series of data. As far as software tools for analysis exhibit limitations for very large samples, we have first sought a limited but representative set of covers by means of clustering procedures and then proceeded to apply PCA to cluster centroids (Ceballos and Bottino 1997). The clustering was performed using a non-supervisioned iterative classification method with the minimal Euclidean distance (hereafter referred as MED) as similarity measure. The method is described in section 2. The study was focused on the subtropical South American sector imagery described in section 3. In that section, we also present the procedures for classification and valuation the coherence of the method.

A first step in classification was performed using as many as 13 variables and finding several tens of centroids. Factor analysis allowed to select a smaller set of non-redundant variables. This restricted set yielded several tens of final centroids, expectedly corresponding to different classes or types of scene. In order to valuation the coherence of the method a set of several hundreds of image targets was randomly chosen, which were labelled by nephanalysis and compared with classification by MED. The results are discussed in section 4. Finally, the classes were grouped by factor analysis to define the main types of scenes.

2. A scene classification method

The proposed task is to describe (in the scale of one pixel or 3 × 3-pixels) multispectral and spatial variability properties which allow to distinguish various objects such as different types of clouds or ground surfaces. The analysis includes definition of a numerical method for scene classification, respecting the limitations of spatial resolution inherent to available imagery.

Classification of satellite imagery presents different statistical approaches, according to the purpose of the research. Generally speaking, they exhibit one of two main aspects: 1) supervisioned procedures start by the collection of samples within an image which correspond to known cover types, then drawing up a statistical description used to classify the whole image (Welch *et al.* 1988, Baum *et al.* 1997, Chen *et al*. 1989, Uddstrom and Gray 1996, Tag *et al*. 2000); 2) non-supervisioned procedures find out a statistical repartition function which is applied for image segmentation in different objects, whose nature is subject of further analysis (Desbois *et al.* 1982, Sèze and Desbois 1987, Sèze and Rossow 1991, Porcú and Levizzani 1992, Ceballos and Bottino 1997, Gordon *et al*. 2005).

It was chosen a non-supervisioned classification procedure known as "method of dynamic clusters" (hereafter referred as MDC). It performs classification of an image using the MED as similarity measure and an iterative definition of a given number of groups or clusters (Diday and Simon 1980, Desbois *et al.* 1982). MED is assessed considering pixels as vectors in *M*-dimensional space, with *M* = number of variables or data defining the pixel. *M* = 13 variables were defined as follows.

Five basic data are associated to each pixel in GOES-8 imagery. Information associated to channel 1 is a reflectance factor *F* for solar radiation incident on the top of the atmosphere [also called "bidirectional reflectance" (Stuhlmann *et al.* 1985)]. Reflectance *R* instead of reflectance factor was chosen as a variable because it appears nearly constant for clear-sky conditions (except for slant solar beam, *i.e.* extreme situations of sunrise and sunset), at least over land scenes (Ceballos *et al*. 2004). Relationship between *F* and *R* is

$$F = \pi L_\lambda / S_\lambda, \quad R = bF / \cos Z \tag{1}$$

L_λ being spectral radiance detected in channel 1, S_λ the spectral specific flux (of solar origin) incident on the top of atmosphere, and *Z* = zenithal solar angle. Factor *b* accounts for a correction due to anisotropy of reflecting surface; usually lies in the range 0.8-1.2 (Lubin and Weber 1995) and will be ignored hereafter, assuming a priori *b* = 1.

Channels 2 to 5 detect radiance of thermal emission origin which is usually quantified in terms of brightness temperatures (*T2*, ..., *T5*). It is to be noted that radiance in channel 2 includes reflected (solar) radiation during daytime.

Eight additional variables were considered, as follows. Two types of variables were defined: 1) differences of thermal channels (*T24*, *T34*, *T54*) related to *T4*; 2) texture in each channel (*X1*, …, *X5*), defined as the logarithm of variance in a set of 3 × 3 elements around a central pixel. The first type usually allows to separate thin from thick and well developed clouds; the second one quantifies local homogeneity of cloud field, helping to separate stratiform from broken cloud.

Pixels were considered vectors in a *M*-dimensional Euclidean space, be $\boldsymbol{p}_\mathbf{n}(x_{1n}, x_{2n}, ..., x_{M,n})$, $M = 13$, $n = 1,2, ..., N$ (*N* = total number of sampled pixels in a given image or group of images). A set of *K* groups can be defined following the "dynamic cluster" procedure. Defining *K* initial reference vectors or "seeds" $\boldsymbol{r}_\mathbf{k}(r_{1k}, r_{2k}, ..., r_{M,k})$, $k = 1, 2, ..., K$, the distance of the *n*-th vector $\boldsymbol{p}_n$ to the *k*-th reference vector $\boldsymbol{r}_k$ is given by

$$d_{nk} = [\sum_{m=1}^{M} (x_{mn} - r_{mk})^2]^{1/2} \tag{2}$$

Each pixel is associated to the nearest reference vector, leading to define *K* clusters with sizes $N_1, N_2, ..., N_K$ and centroids given by the respective average vector $\boldsymbol{r}_k^{(1)}$ defined by

$$r_{mk}^{(1)} = \frac{1}{N_k} \sum_{n=1}^{N_k} x_{mk,n}, \quad m = 1,2,...,M \tag{3}$$

These centroids are taken as a new set of seeds and the N vectors $\boldsymbol{p}_n$ are once more classified. The procedure is repeated generating successive centroids $\boldsymbol{r}_k^{(i)}$, i = 1, 2, ... It is generally expected for displacements of centroids to be shorter at each new step. The iteration stops at $i = I$, when all K displacements $|\boldsymbol{r}_k^{(I)} - \boldsymbol{r}_k^{(I-1)}|$ are smaller than a given value ε. Noting that components x_{mn} of a vector $\boldsymbol{p}_n$ refer to physically different magnitudes and might have quite different numerical values, standardized y_{mn} values were adopted in lieu of the original components:

$$y_{mn} = (x_{mn} - < x_m >) / s_m \quad (4)$$

where $<x_m>$ and s_m are the average and standard deviation of the m-th variable over the N-sized sample (or "universe"). Hereafter, vectors $\boldsymbol{p}$ are defined by a set of standardized components. The procedure leads to K centroids $\boldsymbol{r}_k$ considered characteristic of K different clusters. Subsequently, pixels of new images can be associated to one of K classes according to the nearest centroid.

3. Imagery data and analysis procedures

GOES-8 images available at CPTEC (Center for Weather Forecast and Climate Studies) were analyzed. A South American sector extended from Pacific to Atlantic oceans was considered (43°S to 20°S and 76°W to 35°W). Two different sets of ten days (one day every three days) were chosen in September 2002 at 1609 UTC (about 1200 local time for central longitude of the image). The first group (starting at 1st September) was chosen as training set and the other (starting at 2nd September) for validation procedure. Image sectors were matrices of 500 rows × 1500 columns. GOES images currently recorded exhibit the following characteristics: 1) they have full resolution in channels 2, 4 and 5, so that pixels correspond to sizes of about 4 × 4 km at satellite nadir (longitude 75°W); 2) 16 pixels in channel 1 (size of about 1 × 1 km) correspond to each pixel in channel 4, but only one of them is sampled; 3) pixel size in channel 3 had correspondence with two pixels in channel 4 (covering two rows). Linear interpolation was adopted for line duplication in this channel in order to exactly collocate all five channels.

Concerning the training set, one pixel was sampled from each two rows and two columns for defining variables *R1*, *T2, T3, T4, T5, T24, T34, T54*. The set of 3 × 3-pixels around it in the original image matrix was used for assessing textures *X1* to *X5*. It could be expected that the most part of different types of clouds were present for the total sample (amounting more than 1.8 million pixels).

It is not clear if choice criteria for selecting K seeds and the number K itself are responsible for a unique definition of final groups and centroids. Sèze and Desbois (1987) randomly chose the starting seeds, while Porcú and Levizzani (1992) selected a number of maxima in bi-dimensional histograms. It is always possible to choose a reasonable number of samples corresponding to as many as possible different cover types (this choice induces a supervisioned initial clustering, but further non-supervisioned iteration can account for a "natural" distribution of proper centroids). Baum *et al*. (1997) selected a number of features depending of predominant type of air mass, ranging from 13 to 24. This type of starting point seems more "natural" and convenient than the automatic choices abovementioned. Thirty-five seeds were defined selecting different types of scenes by visual inspection of the training set of images, then extracting a small set of pixels for each observed type and calculating the respective variable mean values. A stopping value ε = 0.04 was chosen for clustering iterations (considering standardized values as described in section 2). Using such initial seeds, not more than ten iterations were needed for fitting the convergence criterion.

Concerning a method for validation, Sèze *et al*. (1986) compared their clustering method with conventional ground-based observation of cloudiness (in oktas), finding a reasonable agreement for the total cloud cover. Nevertheless, ground validation of any satellite cloud classification technique is a delicate issue due to the considerably different ways of observing the clouds (Henderson-Sellers and McGuffie, 1990). Two other possible procedures might be the qualification of automatic results by comparison with 1) cloudiness as shown in synoptic charts (Porcú and Levizzani 1992); 2) visual nephanalysis of satellite imagery (Sèze and

Desbois 1987). Validation of specific cases may be performed choosing samples of targets, selected by previous nephanalysis (Uddstrom and Gray 1996; Baum *et al.* 1997; Tag *et al.* 2000). They can also be selected randomly with subsequent identification by nephanalysis (Welch *et al.* 1988). Present paper follows this last criterion, in two steps. Firstly, the MDC applied to the training set of images allowed to define $K = 35$ centroids or reference vectors. Images of this set were classified by MED related to the K vectors, *and the correspondence with scene types was found out by visual nephanalysis.* Secondly, a set of 320 3 × 3-sized targets was chosen within the set of validation images. Targets were regularly distributed within the images. Each one was identified by visual nephanalysis of its neighborhood surrounding image and was labelled according to six types of scene: surface (Sf, 135), cumulus (Cu, 74), stratus (St, 28), cirrus (Ci, 30), multilayered (ML, 44) and cumulonimbus (Cb, 9). The numbers indicate the observed target frequency. The term "multilayer cloud system" is used in association with complex sets of clouds present during synoptic disturbances, usually covering an extended area and presenting several cloudy layers. In order to compare the set of labelled targets with the objective classification method, each one of them was classified by MED according to the K classes. A contingency table allowed to compare the independent nephanalysis with the results of classification by MED method (see section 4).

4. Results

Table 1 presents the components of centroids found using 13 variables. Three centroids were discarded because they presented unacceptable values due to noise in one or more channels. Note that standardized variables were used in clustering procedure but absolute values x are reported. Relative frequencies g of classes lied between 1% and 6%, thus being all of the same order in the N-sized sample.

[Table 1. Centroids resulting from multispectral clustering of the training set images, September 2002.]

4.1. Detection of redundant variables

Principal components analysis (PCA) was applied to data in table 1, considering the $M \times M$ correlation matrix $\mathbf{R}$ ($M = 13$) for components of $K = 32$ centroids $\boldsymbol{r}_k$ (objects). This type of analysis provides tools for detection of redundant variables among the original M ones, as described in the Appendix A. Vectors $\boldsymbol{r}_k(y_{1k}\ y_{2k}\ \ldots\ y_{Mk})$ have images $\mathbf{Z}_k(\zeta_{1k}\ \zeta_{2k}\ \ldots\ \zeta_{Mk})$ obeying the relation

$$\boldsymbol{r}_k = \mathbf{F}\,\mathbf{Z}_k \qquad (5)$$

[equation (A5) in the Appendix A], where $\mathbf{F}$ represents a matrix of factor loadings. Factor analysis allows for reduction from the M-dimensional description of original variables y_m, to a new set of $J < M$ independent variables ζ_j describing basic modes or behaviours. This is equivalent to describe the m-th variable y of the k-th vector through lower-ordered linear approximations [described by equation (A7), in Appendix A]

$$y_{mk} \approx f_{m1}\,\zeta_{1k} + f_{m2}\,\zeta_{2k} + \ldots + f_{mJ}\,\zeta_{Jk}, \qquad J < M, \qquad (6)$$

A Kaiser-type criterion was applied to assess J, by discarding PC's with variance $\lambda_j < 0.8$. The underlying reasoning is that the independent variables ζ_j with $j > J$ are less meaningful than any single variable y_m, thus should be interpreted as random noise. Table 2 shows factor loadings in matrix $\mathbf{F}$ corresponding to centroids in table 1. It is seen that 4 from 13 principal components could be adopted, since they cumulate 90% of the total variance $M = 13$. When considering four PC's, communality h is higher than 0.83 for all variables y except for texture in channel 3. Factor loadings contributing with more than 50% of variances are underlined in table 2, helping to put in evidence "hidden behaviours" or "factors" in satellite imagery. It is seen that the first two PC's are especially well correlated with variables stressing brightness temperatures (except *T54*) and with texture in channels 4 and 5, respectively, whereas the third PC is representative of texture in VIS channel. Fourth column suggests the existence of a factor better correlated with *T54*.

[Table 2. Factor analysis in principal components for centroids of table 1, 13 variables. *eigen*: eigenvalues; *f1, …, f4*: loading factors; *fr1, …, fr4*: varimax-rotated factors; *comm.*: fourth-order communality]

It is worthwhile to note that equation (5) describes a rigid rotation of orthogonal axis transforming $\boldsymbol{r}_k$ vectors in their images $\boldsymbol{Z}_k$. Equation (6) assumes that second order behaviours ($\lambda << 1$) are mainly due to random noise, but certainly they influence the process of assessing "principal axes" described by equation (A1) (in Appendix A). The so-called Varimax rotation improves factor analysis by executing a new rigid rotation, limited to the subspace of meaningful axes, maximizing factor loadings without changing communality h_{mJ} (Johnson and Wiechern, 1982). Rotated factor loadings *fr* in table 2 make evident a better defined set of correlations between factors and variables:

1) Factor 1 is mainly associated to temperature variables (*T2* to *T5*; *T24* and *T34*). *R1* is also included (with changed sign), due to higher reflectance of well developed (thus colder) cloud. Lower clouds are better observed in channel 1 since it is not so affected by the "blurring" noted in thermal windows (channels 4, 5), which is induced by water vapour absorption/emission. Channel 2 detects thermal emission as well as reflected solar radiation, so that difference *T24* shows also somewhat lower association with Factor 1.

2) Factor 2 associates with cloud texture in temperature channels. Channel 3 exhibits lower correlation, at least because: (a) spatial resolution is different; (b) brightness temperature and texture tend to be the same in channels 3 and 4 for higher cloud (due to negligible water vapour presence above top); (c) on the other hand, spatial variations in water vapour column induce smooth texture in channel 3 for cloudless and low-level cloudy scenes.

3) Texture in VIS channel is described by Factor 3 and (to a certain extent) by Factor 2. Indeed, VIS channel allows better visual inspection of lower clouds (especially cumulus fields). Table 3 also shows the lower but distinct correlation of Factor 3 with *R1* and *T24*.

4) Factor 4 is associated to *T54*. This noteworthy behaviour might be explained by the influence of water vapour continuum in channels 4 and 5 (for instance, *T54* is used in the well-known split-window technique for sea surface temperature assessment); also, liquid and ice phase in clouds have different behaviours of emission/scattering, suggesting *T54* as a valuable tool for analysis of high-level clouds, especially cirrus (Inoue 1987).

It is concluded that four original variables could be used for describing a pixel. It is suggested the set (*T4*, *X4*, *X1*, *T54*). Nevertheless, partial correlations of *R1* and *T24* with rotated factors 1 and 3 (allied to better detection of surface and lower finite clouds) suggest to include one of them as a fifth basic variable.

The method described in sections 2 and 3 was applied to the training set of images now considering 5-dimensional vectors $\boldsymbol{r}_k$ in coordinates *R1*, *T4*, *T54*, *X1*, *X4*. Centroids in table 1 were taken as initial seeds. A minor set of pixels converged to one anomalous centroid, so that only $K = 31$ final centroids are considered. The new set of 5-dimensional centroids is presented in table 4.

Figures 1 illustrate the effects of choosing $M = 5$ in lieu of $M = 13$. The multispectral image of September 1, 2002, 1609 UT was classified by MED related to centroids in table 1 [plate 1(*a*)] and table 4 [plate 1(*b*)]. The plates are RGB composite images based on a simple combination of *T4* and *R1*. *T4* is used to define grey levels between red and blue, with linear change from warmer (pure red) to cooler (pure blue) pixels, and *R1* defines the green level as pixel reflectance. Thus, the expected scene tonalities are: Sf – reddish; Cu – brownish; St – greenish; Ci – bluish; ML – bluish-cyan; Cb – bright cyan. A restricted number of *K* sets of (*R1*, *T4*) are actually used, according to the classification of the pixel and the (*R1, T4*) pair of the corresponding centroid. Note that variables *X1*, *X4* are made evident visually by image texture. It is seen that plates 1*a* and 1*b* are practically equivalent, although some details have been lost in the second one.

[FIGURE 1. Classified scene for September 1, 2002, 1609 UT. RGB image using *R1* in Green and *T4* in Red and Blue (component R increases with *T4*, B has opposite behaviour). Thirty-two clusters obtained for (*a*) $M = 13$ variables, (*b*) $M = 5$ variables (*R1*, *T4*, *T54*, *X1*, *X4*).]

4.2. Coherence analysis of automatic classification

The 320 labelled targets chosen in the set of validation images were labelled by visual nephanalysis in 6 types of scene (see section 3). They were also classified by MED related to 13-dimensional (table 1) and 5-dimensional (table 4) centroids. The results are presented in two contingency tables (tables 3). Each line in these these tables exhibit the frequency of correspondence of one class with each type of scene ("observed truth" found by nephanalysis). It is observed an outstanding monomodal concentration of frequencies around a unique type of scene, suggesting that each centroid defines, almost univocally, characteristics of one of the six types of scene. Conversely, a given type of scene may correspond to several well-defined centroids. These facts are observed for both 13- and 5-dimensional schemes. For instance, centroids grouped as Sf scenes cumulate 91% of 135 diagnosed surface targets when using $M = 13$ variables, and 90% when using $M = 5$. The remaining 10% cases were misclassified as Cu or Ci probably due to low cover or semitransparent cloud structure. An additional contingency table (not shown), was built with pixel-by-pixel direct comparison of 13- and 5-dimensional classification results. It was evident an almost one-to-one correspondence between initial 13-dimensional centroids (used as seeds) and the final ones, with scarcely observed migration to another class. The close resemblance between Figures 1 (*a* and *b*) clearly illustrates this fact.

Observing the number of misclassified targets for each type of scene, it is to be noted that: 1) 13-dimensional results better discriminate high level cloud scenes (Ci, ML, Cb); 2) surface and low and mid-level clouds (Cu, St) exhibit lower dispersion when using only 5 variables.

[Table 3. Frequency of targets objectively classified in several types of scene, using (*a*) $M = 13$ variables and (*b*) $M = 5$ variables. Total number: 320 targets. Modal frequencies (in bold) indicate closer association of centroid to a given type of scene.]

4.3. Search for a minimal group of scenes

The centroids $\boldsymbol{v}_k$ [$\boldsymbol{v}_m$] shown in table 4 were considered as $M = 31$ objects defined by $K = 5$ data. Standardized variables were adopted, leading to a 31 × 31-dimensional correlation matrix **Q**. Factor analysis lead to $J = 4$ non-redundant variables.

[Table 4. Set of 31 5-dimensional centroids, and their classification using factor analysis in principal components; *fr* are Varimax rotated factor loadings. Boldface means $fr^2 \geq 0.75$ [$fr \geq 0.86$]; underlined means $fr^2 \geq 0.6$ [$fr \geq 0.775$]; italic means $0.6 > fr^2 \geq 0.2$ [$0.775 > fr \geq 0.45$]. Groups labelled "G" include one factor contribution $fr^2 \geq 0.6$; labels "a", "b" indicate factor loading being positive and negative, respectively. The last column is the modal type of scene identified by nephanalysis.]

Table 4 summarizes data about centroids $\boldsymbol{v}_k$ and their clustering based on observation of Varimax-rotated factors. A simple criterion of separation in groups was chosen according to squared loads fr^2 (instead of the values *fr*), given that these ones are the actual contributions to communality of centroid components. Cases with $fr^2 > 0.75$ are bold-labelled, underlined cases correspond to $0.75 > fr^2 \geq 0.6$, and italics is used for $0.6 > fr^2 \geq 0.2$.

The simple condition of choosing combinations of factors cumulating more than 75% of total communality leads to define twenty groups. Groups labelled "G" contain one factor contribution $fr^2 > 0.6$; labels "a" and "b" correspond to positive and negative factor loadings, respectively. In order to compare, the most frequent type of scene classificated in labelled targets set [present in table 3 (b)] are included in the last column of table 4.

It is seen that groups with one predominant factor ("G-groups") show some definite characteristics:

• *Groups G1a, G1b*. Negative extremes of *fr1* correspond to elements with low-reflectance (*R1* < 16%, and a single centroid wit 22%), warm and somewhat homogeneous (*T4* > 286 K, *X1* < 0.38, *X4* < 0.2). They correspond to ocean, ground and desert areas. Positive *fr1* is associated to elements with higher reflectance (*R1* > 37%), somewhat low temperatures (*T4* < 275 K) and relatively high texture (*X1* > 0.84, *X4* > 0.35). They are developed Cu, thick Ci, and mixtures of Ci and St (multilayers).

• *Groups G2a, G2b*. They present similar brightness temperatures and VIS texture (*T4* ≈ 275 K, *X1* ≈ 0.9-1.2) but differ in reflectance and temperature texture. *G2a* identifies Ci over surface with low reflectance, high texture and temperature difference (*R1* = 18%, *X4* = 1.1, *T54* = –2.8 K), while *G2b* is thermally homogeneous (*X4* < –1.5) and corresponds to St top (class 19) or low clouds contamination over ocean (class 3).

• *Group G3a*. Fair weather Cu field (*R1* = 13-22%, *T4* = 278-285 K) with high texture (*X1* > 1). No "pure" group *G3b* is identified. The group with factors *g3b1a* and *g3b2b* (class 32 and 18, identified as Cb and thick stratus deck, respectively) shows the stronger opposite behaviour, with high reflectance (*R1* > 74%) and relatively low temperature (*T4* < 269 K).

• *Group G4b*. Thin Ci over surface, presenting low reflectance (*R1* = 29%), somewhat low brightness temperature (*T4* = 268 K) and high difference *T54* (between –12 K and –7 K). The opposite behaviour *G4a* does not appear.

Therefore, Factor 1 is associated to extremes between clear-sky and mid-high level high-textured clouds (multilayered structure; eventually, large Cu). Factor 2 describes cloudy systems with somewhat low brightness temperature, with extremes of thermal texture between low (St) and high (Ci over surface) values. Factor 3 describes cumuliform fields between the extremes of (warm) fair weather Cu and (cold) extended Cb. Factor 4 identifies thin Ci over surface.

Joining "pure" and mixed classes such that combinations of up to two factors cumulate more than 75% of total communality (except for the classes 9 and 16), the number of groups is reduced to seven, similarly to results published by Desbois *et al.* (1982), Sèze and Desbois (1987), Porcú and Levizzani (1992), Baum *et al*. (1997) and Gordon *et al*. (2005).

Mixed groups present the following patterns:

• Class 27 (g1b2a): the relatively low temperature, *T54* = –2.1 K and the presence of factor 2a, indicate some cirrus contamination over surface.

• Class 9 (g1b3a): the highest reflectance and texture in the factor 1b is associated to high altitude Andes deserts.

• Class 22 (g2b1a): the high texture corresponds to a transition from stratiform to developed and non-uniform cover. That is found in modal labelled target classification too.

• Class 12 (g3a2b): the reflectance and low thermal texture indicate a transition cumulus to stratiform, as shown by the factor 2b.

• Classes 16 and 11: the temperature difference (*T54* < –4.1 K) is associated to cirrus (Inoue 1987). Warm temperature (*T4* = 290 K) which lower reflectance and factor 1b contribution indicate cirrus over surface; high reflectance texture and factor 3a indicate cirrus over cumulus.

[Figure 2. Bidimensional histogram from training image sample in (*R1*, *T4*) plane. Centroids position (for *M* = 5 variables) is also represented by numbered black square. It can be seen the typical "ear-sized" distribution.]

The preceding discussion makes evident that well-known bidimensional histograms using reflectance and brightness temperature are not able to illustrate the complexity of multispectral scene descriptions. Figure 2 shows a bidimensional histogram in the plane (*R1*, *T4*), obtained from training image sample, including centroids components described in table 4. The diagram exhibits the typical "ear-sized" characteristics reported in the literature (Desbois *et al.* 1982; Sèze and Desbois 1987). It must be noted that some centroids seem closer but are actually separated by components *X1*, *X4* and/or *T54* in dimensions not included in the histogram.

One realizes that an "objective" small number of cloudy scenes is a somewhat naïf hypothesis, since a continuous series of transitions can be expected between the principal types of cloud fields (especially Cu, St, Ci, Cb) and surfaces (ocean, vegetated and bare soil, desert), with "contaminated" or mixed landscapes between "pure" situations. Clearly, the number of possible or effective scene elements may be defined by the user according to his need.

5. Conclusions

The Method of Dynamic Clusters (MDC) uses minimal Euclidean distance as clustering criterion. It was applied to multispectral information of a large population of GOES 8 Imager pixels, seeking for a proper set of clusters representative of main cloud and ground classes. Pixels properties were defined through $M = 13$ variables describing reflectance and brightness temperature (5 variables), temperature difference related to channel 4 (3 variables) and local homogeneity or texture (5 variables). MDC sought for thirty-two clusters. The method was applied to an extended area of South America, using a number of daytime multispectral images of September 2002 at 1609 UTC (about local noon over most part of the area). Two sets of images were considered, one for training of classification and the other for validation of results.

Each cluster of the training set was represented by its 13-dimensional centroid. Factor analysis in principal components for the set of clusters clearly showed that 4 from 13 principal components would cumulate 90% of total variance, suggesting the existence of only four non-redundant original variables. The non-redundant variable set could be (*T4*, *X4*, *X1*, *T54*). Nevertheless, *R1* appears as a partially independent variable and can be also included, for the benefit of a better definition of surface and lower finite clouds in channel 1.

The five-dimensional description of pixels lead by MDC to a set of thirty-two centroids (table 4) which were used for classifying images of the training set. Nephanalysis of classified images allowed to identification (labeling) of clusters through their correspondence with characteristic cloud and surface scenes. The hypothesis of their use as reference set is fairly supported by applying the classification process to the validation set of images, comparing results with the "true information" of more than three hundred targets (independently labeled by nephanalysis). Comparison was performed considering 13- and 5-dimensional pixels. It is noticeable the fact that despite a significant reduction of variables, the accuracy in target classification remains similar (as verified by frequency distribution in respective contingency tables). However, it is important to note that the whole inclusion of thermal and textural information improves classification of high level clouds (Ci, ML and Cb); on the other hand, using $M = 5$ variables it is verified an improvement in classification of surface and low-middle clouds (Sf, Cu and St). Therefore, reduction to only 5 variables could still be an acceptable choice.

The objective number of distinct objects to be considered is also a matter of discussion. Factor analysis in principal components with Varimax rotation applied to $M = 31$ (objects) defined by 5-dimensional centroids and a simplified method of clustering based on observing the contribution of factors to communality shows that basic scene behavious: 1) extremes between clear-sky and mid-high level high-textured clouds (multilayered structure; eventually, large Cu); 2) cloudy systems with somewhat high brightness temperature, with extremes of thermal texture between low (St) and high (Ci over surface) values; 3) cumuliform fields between the extremes of (warm) fair weather Cu and (cold) extended Cb or uniform thick cloud deck; and 4) thin Ci over surface. It is suggested that resulting labelling of scenes can be grouped itself in a basic set of six or seven distinct types. This number of scenes is somewhat low, but a continuous set of transitions can be expected with a number of "contaminated" or mixed landscapes defined between "pure" situations.

Appendix A. Principal components analysis

Consider a population of K objects defined by column vectors $\boldsymbol{r}$ in a Euclidean M-dimensional space. Let us suppose that components y_m of vectors $\boldsymbol{r}$ are standardized, that is, each one has a null mean $\langle y_m \rangle = 0$ and unit variance $\mathbf{Var}\{y_m\} = \langle y_m^2 \rangle = 1$. Basic Principal Components Analysis (Johnson and Wiechern 1982) shows that each vector $\boldsymbol{r}_k(y_{1k}, y_{2k}, ..., y_{Mk})$ has an image vector $\boldsymbol{z}_k(z_{1k}, z_{2k}, ..., z_{Mk})$ obeying a linear transformation

$$\boldsymbol{r}_k = \mathbf{A}\, \boldsymbol{z}_k \tag{A1}$$

where columns in matrix $\mathbf{A}$ are the M-dimensional orthonormal eigenvectors $\boldsymbol{a}_j\,(a_{1j}, a_{2j}, ..., a_{Mj})$ of matrix $\mathbf{R}$. This vector set uses to be ordered following decreasing eigenvalues λ_j of matrix $\mathbf{A}$. The set of K vectors $\boldsymbol{z}_k$ has components z_{jk} called principal components (PC's) of the original variables y_m. They have mean, variance and cross product such that

$$\langle z_j \rangle = 0, \qquad \mathbf{Var}\{z_j\} = \lambda_j, \quad \langle z_i\, z_j \rangle = \lambda_j\, \delta_{ij}, \quad \Sigma \mathbf{j}\ \mathbf{Var}\{z_j\} = M \tag{A2}$$

where δ_{ij} is the Kronecker symbol. Considering standardized PC's $\zeta_{jk} = z_{jk}/\lambda_j$, the original components of $\boldsymbol{r}_k$ are assessed *exactly* through M equations

$$y_{mk} = f_{m1}\, \zeta_{1k} + f_{m2}\, \zeta_{2k} + ... + f_{mM}\, \zeta_{Mk}, \qquad m = 1,2, \ldots, M, \tag{A3}$$

being

$$f_{mj} = a_{mj}\, \lambda_j^{1/2}, \quad \mathbf{Var}\{y_m\} = \langle y_m^2 \rangle = \Sigma_j\, f_{mj}^2 = 1. \tag{A4}$$

The standardized PC's ζ_j are orthonormal and the coefficients f_{mj} (called loading factors) are correlation coefficients between the m-th y_m and the j-th ζ_j variables. Equation (A1) can be alternatively written as

$$\boldsymbol{r}_k = \mathbf{F}\, \mathbf{Z}_k\,; \tag{A5}$$

$\mathbf{Z}_k$ are vectors composed by standardized PC's ζ_{jk} and matrix $\mathbf{F}$ is composed by factor loadings f_{mj}. Ordered eigenvalues usually show a steep decrease, attaining values $\lambda_j << 1$ for $j = J << M$. Equation (A4) make evident that factor loadings become $f_{mj} << 1$ for $j > J$, so that equation (A3) includes not more than J meaningful PC's:

$$f_{m1}^2 + f_{m2}^2 + ... + f_{mJ}^2 = h_{mJ} \approx 1. \tag{A6}$$

and variables y can be approximated by

$$y_{mk} \approx f_{m1}\, \zeta_{1k} + f_{m2}\, \zeta_{2k} + ... + f_{mJ}\, \zeta_{Jk}, \qquad J < M \tag{A7}$$

The squared factor loading f_{mj} in equation (A6) assesses the fraction of the variance of the m-th component y_m contributed by the j-th factor or principal component, and h_{mJ} ("J-th order communality") accounts for the contribution accumulated up to the J-th order approximation. The quality of an approximation [equation (A7)] (assessed by h_{mJ}) is not the same for all y_m variables but increases with higher J, being perfect if $J = M$. Since $M > J$, equation (A7) constitute a linear system with redundancy of as many as $(M - J)$ variables y_m.

References

Barton, I. J., 1995, Satellite-derived sea-surface temperatures – Current status. *Journal of Geophysical Research -Oceans*, **100**, pp. 8777-8790.

Baum, B. A., Tovinkere, V., Titlow, J., and Welch, R. M., 1997, Automated cloud classification of global AVHRR data using a fuzzy logic approach. *Journal of Applied Meteorology.,* **36,** pp. 1519-1540.

Bottino, M. J., Nobre, P., and Carneiro, G. M., 2003, Detecção de sistemas convectivos nos Trópicos utilizando imagens multiespectrais do satélite geoestacionário GOES 8. [Detection of convective systems in the Tropics using multispectral imagery of GOES 8 satellite]. *Revista Climanálise*, Ano 2, pp. 39-52.

Budyko, M. I., 1969, The effect of solar radiation variations on the climate of Earth. *Tellus*, **21**, pp. 611-619.

Ceballos, J. C., and Bottino, M. J., 1997, The discrimination of scenes by principal components analysis in multi-spectral imagery. *International Journal of Remote Sensing*, **18**, pp. 2437-2449.

Ceballos, J. C., Bottino, M. J., Souza, J. M., 2004, A simplified physical model for assessing solar radiation over Brazil using GOES-E imagery. *Journal of Geophysical Research,* **109**, doi:10.1029/2003JD003531.

Chen, D. W., Sengupta, S. K., and Welch, R. M., 1989, Cloud field classification based upon high spatial resolution textural features: Simplified vector approaches. *Journal of Geophysical Research*, **94**, pp. 14749-14765.

Chou, M. D., 1991, The derivation of cloud parameters from satellite-measured radiances for use in surface radiation calculation. *Journal Atmospheric Science*, **48**, pp. 1549-1559.

Coakley, J. A., and Baldwin, D. G., 1984, Towards the objective analysis of clouds from satellite imagery data. *Journal of Climate and Applied Meteorology*, **23**, pp. 1065-1099.

Defries, R. S., and Townshend, J. R. G., 1994, NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing*, **15**, pp. 3567-3586.

Desbois, M., Sèze, G., and Szejwach, G., 1982, Automatic classification of clouds on Meteosat imagery: Application to high-level clouds. *Journal of Applied Meteorology*, **21**, pp. 401-412.

Diday, E., and Simon, J. C. (Ed.) 1980, Clustering Analysis, Digital Pattern Recognition, pp. 47-92 (New York: Springer-Verlag).

Ellrod, G. P., 1995, Advances in the detection and analysis of fog at night using GOES multispectral infrared imagery. *Weather and Forecasting*. **10**, pp. 606-619.

França, J. R. A., Brustet, J. M., and Fontan, J., 1995, Multispectral remote sensing of biomass burning in West Africa. *Journal of Atmospheric Chemistry*, **22**, pp. 81-110.

Germán, D., Redaño, Á., Lorente, J., Magaldi, A., Bottino, M. J., Machado, L. A. T., 2006, Rainfall estimation over South America using the Precipitation RADAR product of the TRMM satellite, IR-VIS cloud classification and properties of convective systems. Fourth European Conference on Radar in Meteorology and Hidrology, September 2006, Barcelona, Spain.

Giraud, V., Buriez, J. C., Fouquart, Y., Parol, F., and Sèze, G., 1997, Large-Scale Analysis of Cirrus Clouds from AVHRR Data: Assessment of Both a Microphysical Index and the Cloud-Top Temperature. *Journal of Applied Meteorology*, **36**, pp. 664-675.

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

Henderson-Sellers, A., McGuffie, K., 1990, Are cloud amounts estimated from satellite sensor and conventional surface-based observations related? *International Journal of Remote Sensing*, **11**, pp. 543-550, doi: 10.1080/01431169008955038.

Inoue, T., 1987, A cloud type classification with NOAA 7 split-window measurements. *Journal of Geophysical Research*, **92**, pp. 3991-4000.

Johnson, R. A., and Wiechern, D. N. (Ed.), 1982, Applied Multivariate Statistical Analysis, pp. 578 (Prentice Hall).

Lau, N.-C., Crane, M. W., 1995, A satellite view of the synoptic-scale organization of cloud properties in midlatitude and tropical circulation sistems. *Monthly Weather Review*, **123**, pp. 1984-2006.

Lubin, D., and Weber, P., 1995, The use of cloud reflectance functions with satellite data for surface radiation budget estimation. *Journal of Applied Meteorology*, **34**, pp. 1333-1347.

Minnis, P., and Harrison, E. F., 1984, Diurnal variability of regional cloud and clear-sky radiative parameters derived from GOES data. Part I: Analysis method. *Journal of Climate and Applied Meteorology*, **23**, pp. 993-1011.

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

Philander, S. G. H., Gu, D., Halpern, D., Lambert, G., Lau, G., Li, T., and Pacanowski, R. C., 1996, Why the ITCZ is mostly north of the equator. *Journal of Climate,* **9,** pp. 2970–2985.

Porcú, F., and Levizzani, V., 1992, Cloud classification using METEOSAT VIS-IR imagery. *International Journal of Remote Sensing*, **13**, pp. 893-909.

Rossow, W. B., and Garder, L. C., 1993, Cloud detection using satellite measurement of infrared and visible radiances for ISCCP. *Journal of Climate*, **6**, pp. 2341-2369.

Setzer, A. W., and Verstraete, M. M., 1994, Fire and glint in AVHRR's channel 3: a possible reason for the non-saturation mystery. *International Journal of Remote Sensing*, **15**, pp. 711-718.

Seze, G., Drake, F., Desbois, M., Henderson-Sellers, A., 1986, Total and low cloud amounts over France and southern Britain in the summer of 1983: Comparison of surface-observed and satellite-retrieved values. *International Journal of Remote Sensing*, **7**, pp. 1031-1050, doi:10.1080/01431168608948907.

Sèze, G., and Desbois, M., 1987, Cloud cover analysis from satellite imagery using spatial and temporal characteristics of the data. *Journal of Climate and Applied Meteorology,* **26**, pp. 287-303.

Sèze, G., and Rossow, W. B., 1991, Time-cumulated visible and infrared radiance histograms used as descriptors of surface and cloud variations. *International Journal of Remote Sensing*, **12**, pp. 877-920.

Stuhlmann, R., Minnis, P., and Smith, L., 1985, Cloud bidirectional reflectance function: A comparison of experimental and theoretical results. *Applied Optics*, **24**, pp. 396-401.

Sun, D., and Pinker, R. T., 2003, Estimation of land surface temperature from a Geostationary Operational Environmental Satellite (GOES-8). *Journal of Geophysical Research*, **108**(D11), 4326, doi:10.1029/2002JD002422.

Tag, P. M., Bankert, R. L., and Brody, L. R., 2000, An AVHRR Multiple Cloud-Type Classification Package. *Journal of Applied Meteorology*, **39**, pp. 125-134.

Tokuno, M., and Tsuchiya, K., 1993, Classification of cloud types based on data of multiple satellite sensors. *Advance Space Research*, **14**, pp. 199-206.

Uddstrom, M. J., and Gray, W. R., 1996, Satellite cloud classification and rain-rate estimation using multispectral radiances and measures of spatial texture. *Journal of Applied Meteorology*, **35**, pp. 839-858.

Welch, R. M., Sengupta, S. K., and Kuo, K. S., 1988, Marine stratocumulus cloud fields off the coast of southern California observed using LANDSAT imagery. Part II: Textural analysis. *Journal of Applied Meteorology*, **27**, pp. 363-378.

Table 1. Centroids resulting from multispectral clustering of the training set images, September 2002.

Class	R1	T2	T3	T4	T5	T24	T34	T54	X1	X2	X3	X4	X5
1	6.7	287.1	244.2	286.3	285.1	0.8	–42.1	–1.2	–0.416	–1.182	–1.793	–1.581	–1.504
2	9.5	288.6	240.7	285.5	283.9	3.1	–44.8	–1.6	0.045	–0.468	–1.671	–0.922	–0.839
3	16.8	291.5	245.3	285.7	284.7	5.9	–40.3	–1.0	–0.066	–1.106	–2.718	–6.000	–1.495
4	20.9	318.9	251.2	310.9	310.7	8.0	–59.7	–0.3	0.398	–0.122	–1.439	–0.099	–0.144
5	12.3	302.0	245.7	297.1	295.2	4.9	–51.5	–1.9	–0.014	–0.388	–6.000	–0.756	–0.736
6	14.7	309.4	246.1	303.1	300.8	6.3	–57.0	–2.3	–0.163	–0.622	–1.758	–0.958	–0.973
7	15.5	311.2	243.7	303.6	301.0	7.6	–59.9	–2.6	0.138	0.156	–1.637	–0.001	–0.078
8	13.6	290.5	240.6	282.8	281.2	7.7	–42.2	–1.6	1.298	0.351	–1.665	–0.098	–0.173
9	22.1	311.1	251.1	301.1	300.8	10.0	–50.0	–0.2	0.880	0.711	–1.076	0.762	0.714
10	24.7	302.6	237.5	280.0	273.4	22.5	–42.6	–6.7	1.196	0.373	–0.918	1.229	1.212
11	22.9	304.8	245.4	291.3	288.1	13.6	–45.9	–3.1	1.650	–0.335	–1.693	0.527	0.424
12	25.7	299.5	251.4	283.3	282.5	16.2	–31.8	–0.8	1.535	0.097	–1.678	–0.489	–0.583
13	23.2	292.1	241.5	278.3	276.7	13.8	–36.8	–1.6	2.070	0.597	–1.495	0.881	0.768
14	13.5	301.9	248.5	296.2	294.6	5.7	–47.6	–1.6	0.069	0.010	–1.638	–0.355	–0.410
15	59.9	292.6	242.9	271.2	270.4	21.4	–28.4	–0.9	2.084	0.644	–1.319	0.798	0.695
16	17.1	305.5	239.4	295.0	291.7	10.6	–55.6	–3.3	0.424	0.656	–0.890	1.024	1.027
17	26.3	295.2	242.6	281.3	279.5	14.0	–38.6	–1.8	1.682	0.359	–6.000	0.504	0.385
18	67.3	295.8	244.3	272.9	272.5	22.9	–28.6	–0.3	1.369	–0.037	–1.623	–0.837	–0.790
19	48.8	305.0	250.7	279.9	279.7	25.0	–29.2	–0.2	1.252	–1.057	–1.814	–1.361	–1.421
20	18.5	290.8	235.6	277.1	274.3	13.7	–41.5	–2.8	1.084	1.123	–0.574	1.438	1.408
21	73.4	281.7	235.1	254.1	252.8	27.6	–19.0	–1.3	1.544	1.258	–0.732	0.829	0.772
22	50.4	300.0	247.7	276.6	276.2	23.4	–28.9	–0.4	1.404	–0.367	–6.000	–0.948	–0.966
23	66.6	290.6	234.8	252.7	251.5	37.9	–17.9	–1.2	1.119	0.404	–1.509	–0.127	–0.185
24	32.1	293.6	229.6	260.1	251.7	33.5	–30.5	–8.4	0.879	–0.058	–0.812	0.619	0.530
25	42.3	293.7	239.4	272.8	271.8	20.9	–33.3	–1.0	1.750	0.005	–1.579	–0.014	–0.137
26	65.6	255.9	227.4	235.8	234.9	20.1	–8.5	–1.0	0.747	0.348	–0.916	0.137	0.099
27	12.8	288.7	236.5	280.7	278.2	8.0	–44.2	–2.5	0.210	–0.084	–1.122	0.368	0.392
28	30.2	279.5	231.6	258.1	255.3	21.5	–26.5	–2.7	0.977	0.522	–0.866	0.718	0.658
29	33.1	280.7	232.9	259.1	256.7	21.6	–26.3	–2.4	1.641	1.420	–0.389	1.607	1.528
30	54.3	263.9	228.4	240.5	239.2	23.5	–12.1	–1.3	1.296	1.159	–0.344	1.118	1.021
31	49.6	271.3	222.6	235.4	231.6	35.9	–12.8	–3.8	0.836	0.415	–0.543	0.607	0.487
32	75.5	252.9	218.7	221.3	220.8	31.6	–2.6	–0.6	0.352	0.244	–1.002	–0.301	–0.251

Table 2. Factor analysis in principal components for centroids of table 1, 13 variables. *eigen*: eigenvalues; *f1, …, f4*: loading factors; *fr1, …, fr4*: varimax-rotated factors; *comm.*: fourth-order communality.

Eigen	var	f1	f2	f3	f4	fr1	fr2	fr3	fr4	comm
6.612	R1	0.763	–0.357	0.394	–0.041	0.759	0.040	0.484	0.234	0.867
2.643	T2	–0.807	0.350	0.267	–0.262	–0.913	–0.149	0.210	–0.117	0.914
1.518	T3	–0.861	0.028	0.451	0.002	–0.842	–0.334	0.194	0.296	0.946
0.887	T4	–0.948	0.294	0.097	0.004	–0.977	–0.166	–0.101	0.030	0.995
0.654	T5	–0.953	0.252	0.140	0.055	–0.970	–0.183	–0.094	0.107	0.994
0.319	T24	0.828	–0.106	0.197	–0.406	0.731	0.134	0.537	–0.243	0.900
0.197	T34	0.892	–0.412	0.112	–0.005	0.946	0.053	0.255	0.122	0.978
0.073	T54	–0.162	–0.495	0.556	0.631	–0.005	–0.226	0.071	0.961	0.979
0.057	X1	0.462	0.303	0.721	–0.171	0.136	0.417	0.806	0.110	0.854
0.039	X2	0.635	0.617	0.193	0.287	0.239	0.896	0.194	0.073	0.904
0.000	X3	0.442	0.324	–0.356	0.364	0.303	0.577	–0.364	–0.044	0.560
0.000	X4	0.481	0.766	0.122	0.036	0.047	0.857	0.241	–0.197	0.834
0.000	X5	0.531	0.809	–0.004	0.073	0.093	0.923	0.131	–0.252	0.942

Table 3. Frequency of targets objectively classified in several types of scene, using (*a*) $M = 13$ variables and (*b*) $M = 5$ variables. Total number: 320 targets. Modal frequencies (in bold) indicate closer association of centroid to a given type of scene.

(a)

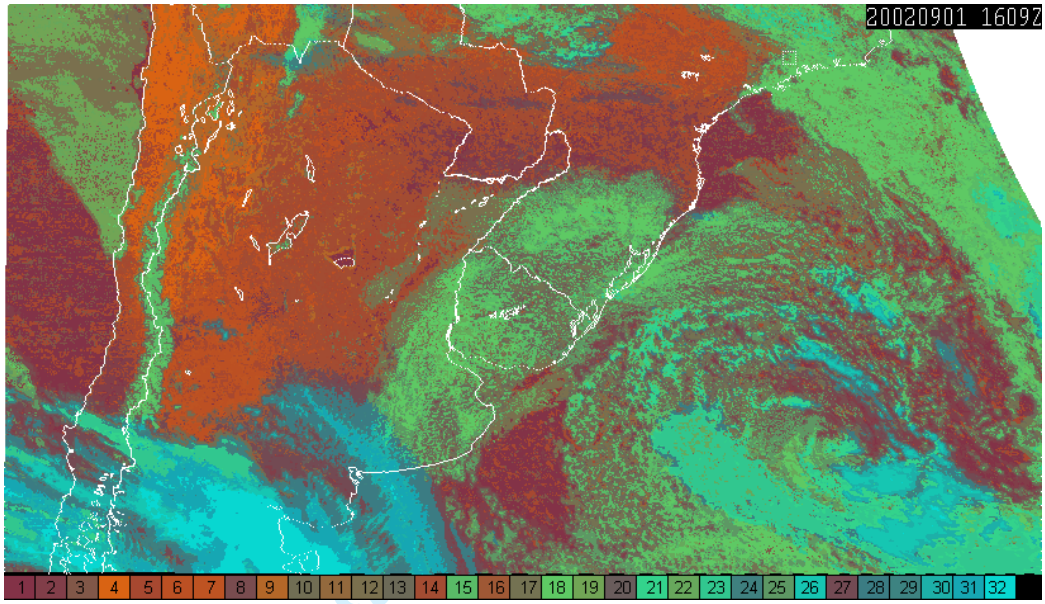
	Type of scene					
Centroid	Sf	Cu	St	Ci	ML	Cb
1	**22**	0	0	0	0	0
2	**16**	2	0	1	0	0
3	**2**	1	0	0	0	0
4	**12**	0	0	0	0	0
5	**15**	0	0	0	0	0
6	**24**	0	0	0	0	0
7	**14**	0	0	1	0	0
9	**3**	1	0	0	0	0
14	**6**	1	0	0	0	0
16	**9**	2	0	2	0	0
8	7	**14**	0	1	0	0
11	1	**9**	0	0	0	0
12	0	**5**	1	0	0	0
13	0	**16**	0	1	2	0
15	0	**4**	3	0	1	0
17	0	**10**	0	1	0	0
25	0	**6**	3	0	1	0
18	0	0	**10**	0	0	0
19	0	0	**5**	0	0	0
22	0	2	**3**	0	0	0
10	1	0	0	**6**	0	0
20	0	1	0	**9**	0	0
27	3	0	0	**4**	0	0
21	0	0	1	0	**3**	0
23	0	0	2	0	**7**	0
24	0	0	0	0	**1**	0
26	0	0	0	0	**10**	1
28	0	0	0	1	**2**	0
30	0	0	0	0	**5**	0
31	0	0	0	0	**5**	0
32	0	0	0	0	4	**8**
29	0	0	0	3	3	0
Sum	135	74	28	30	44	9

(b)

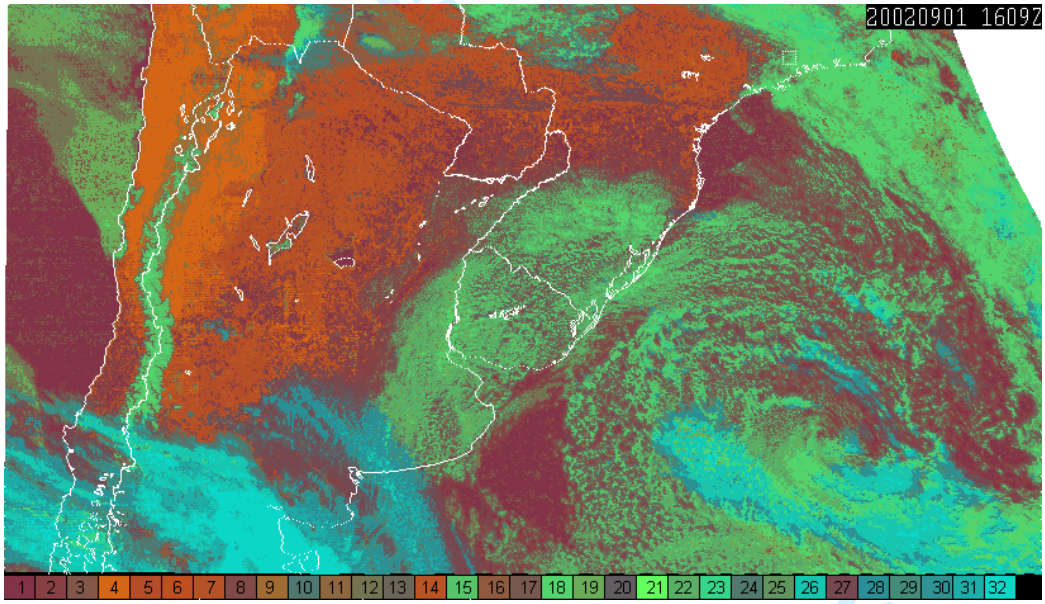
	Type of scene					
Centroid	Sf	Cu	St	Ci	ML	Cb
1	**30**	0	0	0	0	0
2	**14**	1	0	1	0	0
3	**2**	1	0	0	0	0
4	**15**	0	0	0	0	0
5	**24**	0	0	0	0	0
6	**10**	0	0	0	0	0
7	**8**	0	0	0	0	0
14	**8**	0	0	1	0	0
27	**10**	1	0	3	0	0
8	7	**11**	0	0	0	0
11	0	**2**	0	0	0	0
12	0	**5**	1	1	0	0
13	0	**7**	0	1	1	0
15	0	**3**	2	0	1	0
17	0	**24**	0	1	0	0
25	0	**14**	1	0	0	0
18	0	0	**6**	0	1	0
19	0	0	**14**	0	0	0
10	1	0	0	**3**	1	0
16	2	1	0	**5**	0	0
20	2	2	0	**7**	1	0
24	0	0	0	**1**	0	0
23	0	0	3	0	**5**	0
26	0	0	0	0	**13**	0
28	0	0	0	1	**5**	0
29	0	1	0	2	**3**	0
30	0	0	0	1	**3**	0
31	0	0	0	0	**3**	0
32	0	0	0	0	7	**9**
9	2	0	0	2	0	0
22	0	1	1	0	0	0
21	0	0	0	0	0	0
Sum	135	74	28	30	44	9

Table 4. Set of 31 5-dimensional centroids, and their classification using factor analysis in principal components; *fr* are Varimax rotated factor loadings. Boldface means $fr^2 \geq 0.75$ [$fr \geq 0.86$]; underlined means $fr^2 \geq 0.6$ [$fr \geq 0.775$]; italic means $0.6 > fr^2 \geq 0.2$ [$0.775 > fr \geq 0.45$]. Groups labelled "G" include one factor contribution $fr^2 \geq 0.6$; labels "a", "b" indicate factor loading being positive and negative, respectively. The last column is the modal type of scene identified by nephanalysis.

Class	R1	T4	T54	X1	X4	fr1	fr2	fr3	fr4	Group	type of scene
30	40.0	251	–0.4	1.61	1.39	**0.88**	0.36	0.04	0.31	G1a	ML
28	38.8	249	–2.0	0.95	0.64	**0.86**	0.34	–0.31	0.21	G1a	ML
15	64.6	270	–0.5	2.32	0.84	**0.94**	–0.33	0.00	–0.08	G1a	Cu
25	40.4	275	–1.1	1.99	0.51	0.80	–0.20	*0.57*	–0.04	G1a3a	Cu
26	65.9	237	–0.8	0.97	0.35	0.78	–0.06	*–0.56*	0.26	G1a3b	ML
23	75.1	255	–1.6	1.48	0.74	0.81	–0.17	*–0.56*	–0.07	G1a3b	ML
31	53.1	241	–4.1	0.84	0.60	0.80	0.16	*–0.54*	–0.21	G1a3b	ML
29	36.6	263	–3.6	1.73	1.28	0.78	0.37	0.21	*–0.46*	G1a4b	ML
4	21.9	309	0.2	0.38	–0.01	**–0.94**	0.04	0.08	0.34	G1b	Sf (Andes)
5	13.9	301	–1.7	–0.22	–0.73	**–0.97**	0.04	–0.05	0.23	G1b	Sf
14	15.8	303	–1.5	0.28	0.11	**–0.96**	0.22	0.12	0.14	G1b	Sf
6	14.1	304	–3.5	–0.23	–0.79	**–0.99**	0.03	–0.05	–0.09	G1b	Sf
7	15.4	302	–3.3	0.35	0.19	**–0.92**	0.27	0.15	–0.24	G1b	Sf
1	6.9	286	–1.2	–0.43	–1.54	–0.85	–0.06	–0.04	0.52	G1b4a	Sf (Ocean)
2	10.4	287	–1.4	0.24	–0.91	**–0.86**	–0.02	0.22	0.45	G1b4a	Sf
27	12.3	282	–2.1	0.04	0.40	*–0.73*	*0.62*	–0.04	0.28	g1b2a	Sf
9	24.1	296	0.2	1.21	0.94	*–0.60*	0.39	*0.55*	0.43	g1b3a	(Sf/Ci) (Andes)
20	18.3	275	–2.8	0.89	1.14	–0.04	**0.93**	0.32	–0.15	G2a	Ci
19	52.2	278	–0.1	1.20	–1.49	0.10	**–0.90**	–0.19	0.39	G2b	St
3	16.9	286	–1.0	–0.06	–6.00	–0.44	–0.81	0.05	0.38	G2b1b	Sf
22	53.5	272	–0.4	1.72	–0.34	*0.68*	*–0.69*	–0.04	0.24	g2b1a	(St/Cu)
17	17.4	282	–1.8	1.74	0.41	0.08	0.16	**0.98**	–0.07	G3a	Cu
13	21.7	278	–1.4	2.30	1.05	0.43	0.20	**0.88**	–0.09	G3a	Cu
8	13.0	285	–1.7	1.06	–0.20	*–0.53*	0.10	0.81	0.22	G3a1b	Cu
12	27.2	281	–0.6	1.53	–0.71	0.02	*–0.56*	*0.72*	0.41	g3a2b	Cu
32	74.7	226	–1.1	0.28	–0.29	*0.62*	–0.12	*–0.72*	0.27	g3b1a	Cb
18	73.8	269	–0.4	1.13	–0.55	0.42	*–0.61*	*–0.65*	0.17	g3b2b	St
10	28.5	267	–7.0	1.18	1.14	0.31	0.42	0.07	–0.85	G4b	Ci
24	29.1	268	–12.4	0.75	0.79	0.11	0.24	–0.07	**–0.96**	G4b	Ci
16	17.9	290	–5.4	0.39	0.75	*–0.57*	*0.52*	0.03	*–0.63*	g4b1b2a	Ci
11	22.3	290	–4.1	1.73	0.88	–0.04	0.25	*0.67*	*–0.69*	g4b3a	Cu



(a)



(b)

FIGURE 1. Classified scene for September 1, 2002, 1609 UT. RGB image using *R1* in Green and *T4* in Red and Blue (component R increases with *T4*, B has opposite behaviour). Thirty-two clusters obtained for (*a*) $M = 13$ variables, (*b*) $M = 5$ variables (*R1*, *T4*, *T54*, *X1*, *X4*).

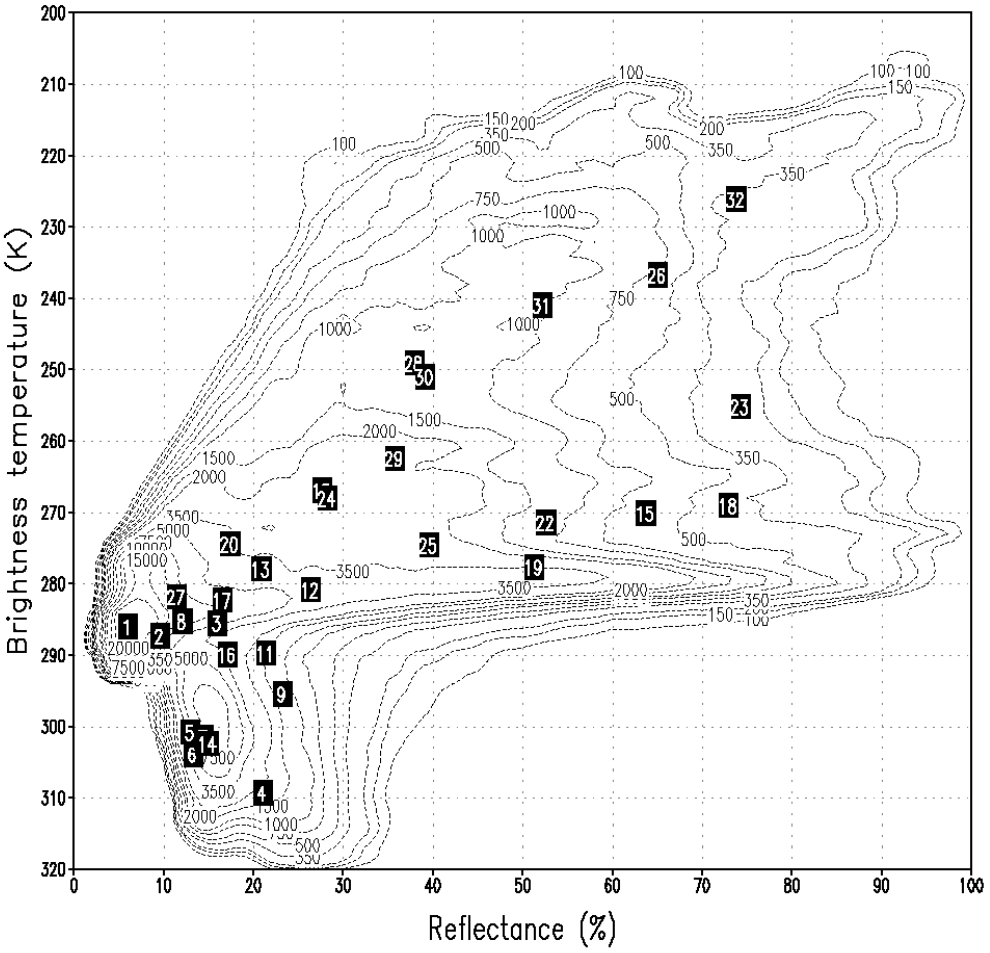


Figure 2. Bidimensional histogram from training image sample in (*R1*, *T4*) plane. Centroids position (for $M = 5$ variables) is also represented by numbered black square. It can be seen the typical "ear-sized" distribution.