Anais XI SBSR, Belo Horizonte, Brasil, 05-10 abril 2003, INPE, p. 1129-1135.

CLASSIFICATION OF SCENES IN MULTISPECTRAL GOES-8 IMAGERY

MARCUS JORGE BOTTINO
JUAN CARLOS CEBALLOS

CPTEC – Instituto Nacional de Pesquisas Espaciais
12630 – Cachoeira Paulista – SP, Brazil
{bottino, ceballos}@cptec.inpe.br

Abstract. Preliminary results are presented, concerning a method for identification of different types of clouds and clear-sky pixels in GOES-8 imagery. Based on full resolution images in five channels, thirteen different variables were designed for pixel identification. A clustering procedure ("dynamic cloud") was used for determining 32 different classes of pixel environment, defined by respective 13-dimensional centroids. Results are coherent with usual two-dimensional analysis based on reflectance (channel 1) and brightness temperature (channel 4), while eleven remaining components could offer additional information. Hierarchical clustering of centroids suggests the existence of five main groups of scenes. The question about the minimal number of non redundant variables is analyzed by factor analysis of centroids. It is found that four variables would be enough for proper description of cloud and clear-sky fields, be: brightness temperature and texture in channel 4, texture in VIS channel, and difference between channels 5 and 4.

Keywords: remote sensing, image processing, GOES-8, cloud, meteorology

1. Introduction

The amount and type of cloudiness is an important information for meteorological analysis and modeling. The Center for Time Forecast and Climate Studies (CPTEC/INPE) currently records GOES-8 imager files in five channels, which could be used for automatic cloud cover classification over South America. Usual analysis is based on visual inspection of brightness in VIS channel and corresponding brightness temperature in a thermal infrared window (channel 4). Subjective analysis includes perception of texture (spatial variability) of scenes, which characterizes different types of cloud fields. An objective analysis should include these variables numerically, as well as others emerging from combination of multispectral data (for instance, it has been found that difference between brightness temperatures in channels 2 or 3 related to channel 4 allow to better identify foggy and convective cloud scenes, respectively. A number of objective methods have been designed in the former 20 years for GOES or Meteosat imagery, but most of them are based on VIS and one thermal channel only (Rossow *et al*., 1987; Sèze and Desbois, 1987; Farki *et al*., 1993) extending from simple methods to complex neural network or fuzzy logic procedures (Baum *et al*., 1997). This paper presents preliminary results of objective analysis procedures which use information of all 5 channels, seeking for two basic questions: 1) how many basic groups should be considered for GOES scenes? 2) how many and which variables would provide enough information?

2. Clustering method

Channel 2, 4 or 5 pixels correspond to about 4×4 km at satellite nadir (75°W). Pixel size is 4×8 km for channel 3, and 1×1 km for channel 1. Images have full resolution in channels 2 to 5 but are sampled in channel 1 (one pixel for each channel 4 pixel).

Information provided by channel 1 is reflectance factor *F* (also called "bidirectional reflectance": CITA) while channels 2 to 5 provide brightness temperatures (*T2*, ..., *T5*). Two types of additional variables were defined: 1) differences between two 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 clouds; the second one quantifies local homogeneity of cloud field, helping to separate stratiform from broken cloud. Thus, one pixel may be considered as a vector in a 13-dimensional space. Reflectance R instead of reflectance factor was chosen as a variable, since previous works showed that it remains nearly constant for clear-sky conditions and not slant solar beam, *i.e.* extreme situations of sunrise and sunset (Bottino and Ceballos, 2000). Relationship between F and R is

$$F = \pi L_l / S_l, \quad R = f F / \cos Z \tag{1}$$

where L_λ is spectral radiance detected in channel 1, and S_λ the spectral specific flux (of solar origin) incident at the top of atmosphere. Reflectance R refers to the relative irradiance reflected by earth-atmosphere system (thus cos$Z$, Z= zenithal solar angle must be introduced). Factor f accounts for a correction due to anisotropy of reflecting surface; usually it does not exceed 1.2 (Lubin and Weber, 1995) and will be ignored hereafter.

Objective classification of images was performed following the "dynamic cloud" procedure described by Sèze and Desbois (1987). Basically, pixels were thought of as a 13-dimensional vector in an Euclidean space, be $\mathbf{p_n}(x_{1n}, x_{2n}, ..., x_{13,n})$, n=1,2, ..., N. Its distance to reference vectors $\mathbf{r_k}(r_{1k}, r_{2k}, ..., r_{13,k})$, k=1, 2, ..., K, or "seeds" is given by

$$d_{nk} = [\sum_i (x_{in} - r_{ik})^2]^{1/2} \tag{2}$$

Each pixel is labeled by its association to the nearest reference vector, defining K clusters with sizes N_1, N_2, ..., N_K and centroids given by the average vector $\mathbf{r}^{(1)}$ defined by

$$r_{ik}^{(1)} = \frac{1}{N_k} \sum_{n=1}^{N_k} x_{ik,n} \tag{3}$$

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

$$y_{in} = (x_{in} - \langle x_i \rangle) / s_i \tag{4}$$

where $\langle x_i \rangle$ and s_i are the average and standard deviation of the i-th variable over the N-sized sample. Hereafter, vectors $\mathbf{p}$ are considered defined by a set of standardized components.

3. Imagery data and processing

Ten days where chosen within period September 2002 and a South American sector between 42°S and 20°S, 78°W and 40°W, extended from Pacific to Atlantic oceans. All images considered were for 1609 UTC (about 1200 local time for central longitude). One pixel was sampled from each two raws and two columns. It can be expected that all characteristic types of clouds were present for the total sample (amounting more than 1.8. million pixels).

Thirty-five seeds were used, observed in a previous work corresponding to a similar region but for January 1997 (Bottino and Ceballos, 2000). A stopping value ε= 0.04 was chosen for iterations.

Considering that some variables could present redundancy (thus consuming too much computer time), factor analysis in principal components was applied to centroid vectors using SYSTAT software. After elimination of redundant variables, the classification scheme was applied again, in order to define a final set of characteristic centroids. Fortran programming was used, with double precision for floating point variables.

4. Results

Convergence of clustering procedure produced 32 clusters with centroids described by **Table 1**. Two seeds had not close enough vectors. A third centroid was affected by absence of actual information in channel 2, providing not acceptable values. Relative frequency of classes was between 1.3% and 6%, that is, they were somewhat equally distributed in the total sample.

class	R%	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 1. Centroids resulting from multispectral clustering.

Figures 1 illustrate a case of image classification using all 32 centroids with 13 variables each. The first image presents an RGB composition of temperature *T4* (using a sequence from pure Red to pure Blue, from warmer to colder pixels) and reflectance *R* (using a Green sequence from darker to brighter pixels). The second presents the same image, with classified pixels. The same RGB scale is used but discretized accordingly to *R* and *T4* components of respective centroids. It can be seen that (at least for components *R*, *T4*) the clustering procedure is coherent with the first image. It is also seen that some regions exhibit differences, probably due to the influence of the other 11 components of centroids.

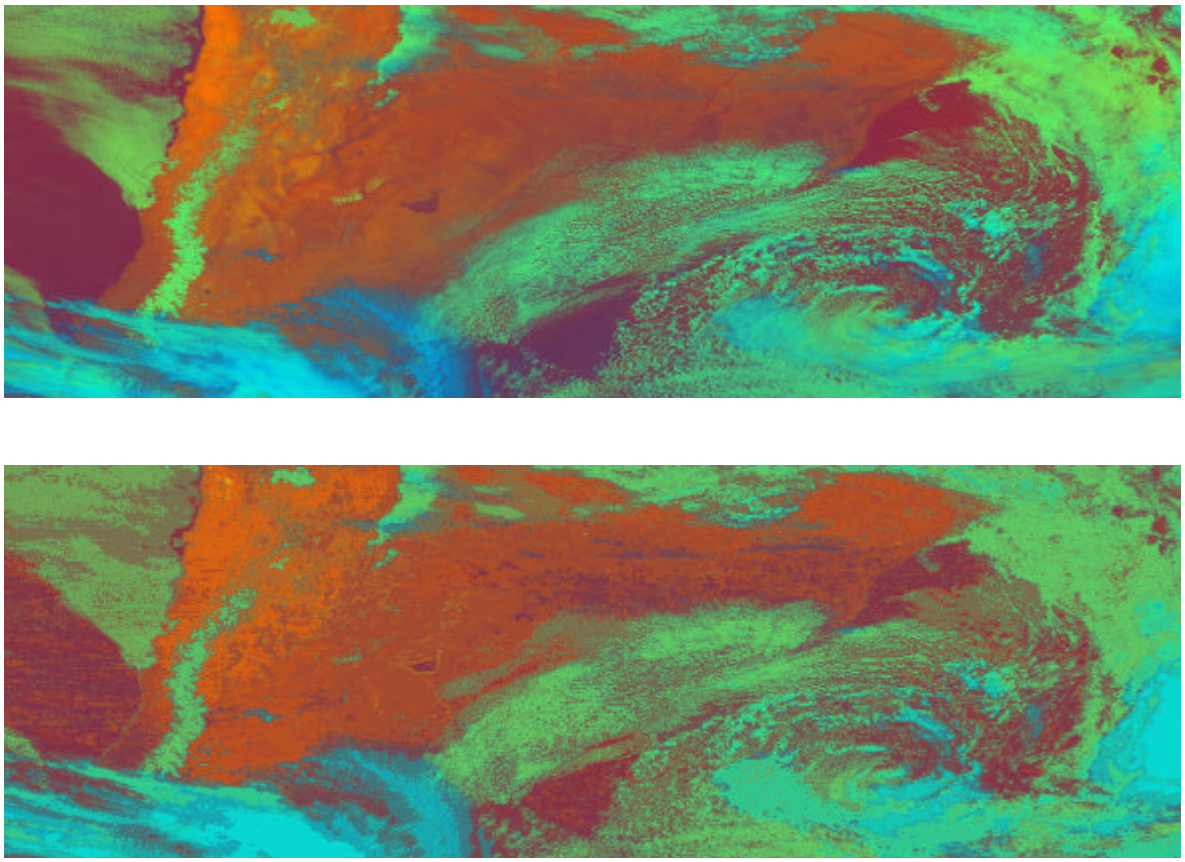

Figure 1. Results of classification for September 1st, 2002. *Top*: RGB composed image using channels 1 and 4 (see text). *Bottom*: classified image , with centroids following the same rule of RGB color code. Note that now only 32 colors are present.

From a meteorological point of view, the main cloud configurations might be classified in low, mid- and high level clouds (which should influence on observed brightness temperature), being shallow (weakly bright) or deep (very bright). In addition, cumuliform structures should present higher texture values while stratiform must present lower values. Also, some clusters should be characteristic of clear-sky conditions (identifying water or soil surfaces). In order to identify a minimal number of classes, the Ward's clustering method was applied to the set of 32 centroids in **Table 1**. Five groups were found, truncating the Ward's dendrogram just before high increases in internal "inertial moment" is verified. The centroid groups were the following:

Group 1: (26 – 30 – 31 – 32)
Group 2: a (21-23), b (24 – 28 – 29)

Group 3: (15 – 18 – 19 – 22 – 25)
Group 4: a (10 – 12 – 13 – 17 – 20), b (1 – 2 – 3 – 8 – 27)
Group 5: a (4 – 6 – 7 – 9), b (5 – 11 – 14 – 16)

Subgroups (a, b) correspond to internal clustering inside a group. **Figure 2** illustrates *R*, *T4* dispersion of centroids classified in groups 1 to 5. Labels in the figure indicate possible physical correspondence from a meteorological point of view. They roughly agree with previous works based on *R*, *T4* information only.

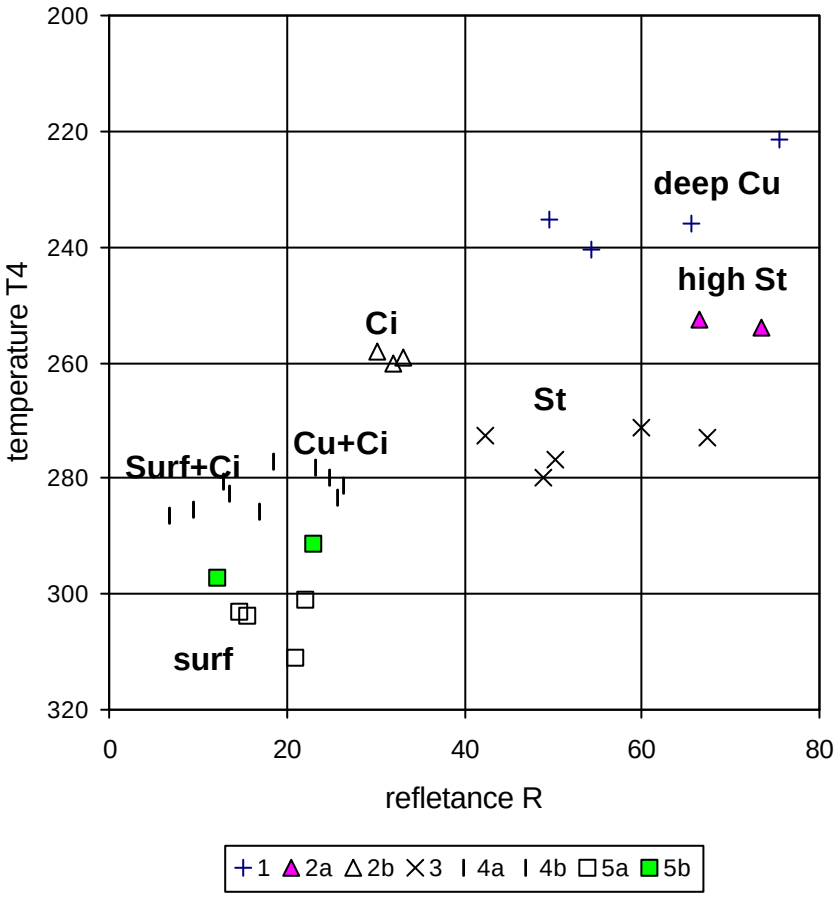

Figure 2. Results of clustering centroids (13-dimensional) using Ward's method. Only two dimensions are represented (reflectance in channel 1 and brightness temperature in channel 4). Labels (Ci= cirriform, St= stratiform, Cu= cumuliform, surf= surface) indicate possible correspondence with meteorological classification.

No further analysis was performed about additional information provided by the remaining 11 variables. Given their high number and the possible complexity involved, it seemed more important to define which of them could be non-redundant. This task was done using factor analysis in principal components. The data matrix was defined as having 13 variables with 32 samples (one for each centroid), generating a 13×13 correlation matrix. The number of eigenvalues higher than 0.8 (a Kaiser-type criterion) was four, suggesting equal number of actually independent variables.

var	fr1	fr2	fr3	fr4	comm
T4	**-0.977**	-0.166	-0.101	0.028	0.9931
T5	**-0.970**	-0.183	-0.095	0.104	0.9942
T34	**-0.946**	-0.054	-0.255	-0.124	0.9782
T2	**-0.912**	-0.149	0.210	-0.119	0.9122
T3	**-0.842**	-0.335	0.193	0.292	0.9437
R	*0.758*	0.040	*0.484*	0.236	0.8661
T24	*-0.731*	-0.134	*-0.538*	0.241	0.8998
X5	0.094	**0.923**	0.131	-0.251	0.9409
X2	0.239	**0.896**	0.195	0.075	0.9036
X4	0.047	**0.857**	0.241	-0.196	0.8332
X3	0.303	*0.577*	*-0.363*	-0.044	0.5584
X1	0.135	0.416	**0.807**	0.109	0.8544
T54	0.001	0.224	-0.071	**-0.961**	0.9787

Table 2. Factor analysis in principal components for thirteen variables. Rotated factor loadings *fr* obtained by Varimax rotation. Communality is shown in last column.

Table 2 presents factor loadings found by Varimax rotation over four principal components. Gruping variables are labeled in bold. It is seen that one fator is represented by brightness temperatures; only one (for instance *T4*) could be adopted representing all of them (even the difference *T34*). In the same manner, texture in channel 4 could be assumed as statistically equivalent to *X2* and *X5*. Texture in channel 1 (associated to cumuliform fields) and difference *T54* (probably representing the effect of water vapor and cirrus clouds over any scene) should be adopted as independent variables. Intermediate variables (labeled in italic) are *X3*, *R* and *T24*; they include contribution of two different factors.

5. Conclusions.

Clustering by "dynamic cloud" procedure is consistent with usual analysis using only reflectance in channel 1 and brightness temperature in channel 4. Although a high number of scenes could be defined a priori, five main groups (or maybe eight) seem to be representative for meteorological purposes. It is suggested that only four variables (namely, brightness temperature and texture in channel 4, texture in VIS channel and difference between channels 5 and 4) could be enough for proper description of GOES-8 scenes using multispectral imagery. With this basic information useful for saving computer time, more detailed objective analysis is being performed at CPTEC meteorological satellites group.

References

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

Bottino M.J.; Ceballos, J.C. Classificação de cenas em imagens GOES multiespectrais mediante um método de "agrupamento dinâmico". XI Congresso Brasileiro de Meteorologia, *Anais*, 3915-3923, 2000 (publicado em CD-ROM).

Farki B., D.; Dagorne, B.; Guillot, P.; Le Borgne; Marsouin, A. Classification of clouds over Africa with Meteosat 4. *Veille Climatique Satellitaire*, Centre de Meteorologie Spatiale, Meteofrance, 1993.

Lubin D.; Weber, P. The use of cloud reflectance Functions with satélite data for surface radiation budget estimation. *Journal of Appl. Meteorology*, **34**, 1333-1347, 1995.

Rossow, W.B.; Mosher, F.; Kinsella, E.; Arking, A.; Desbois, M.; Harrison, E.; Minnis, P.; Ruprecht, E.; Sèze, G.; Simmer, C.; Smith, E. ISCCP cloud algorithm intercomparison. *J. of Clim. and Appl. Meteor.* **24**, 877-903, 1987.

Sèze, G.; Desbois, M. Cloud cover analysis from satellite imagery using spatial and temporal characteristics of the data. *J. Climate and Appl. Meteor*. **26**: 287-303, 1987.