A Simple Method for the Assessment of the Cloud Cover State in High-Latitude Regions by a Ground-Based Digital Camera

M. P. SOUZA-ECHER AND E. B. PEREIRA

Center for Weather Forecasts and Climate Studies (CPTEC), INPE, São José dos Campos, São Paulo, Brazil

L. S. BINS

Earth Observation (OBT), INPE, São José dos Campos, São Paulo, Brazil

M. A. R. ANDRADE

Center for Weather Forecasts and Climate Studies (CPTEC), INPE, São José dos Campos, São Paulo, Brazil

ABSTRACT

This work describes the development of a simple method of field estimating the sky cloud coverage percentage for several applications at the Brazilian Antarctic Station, Ferraz (62°05′S, 58°23.5′W). The database of this method was acquired by a digital color camera in the visible range of the spectrum. A new algorithm was developed to classify each pixel according to a criteria decision process. The information on the pixel contamination by clouds was obtained from the saturation component of the intensity, hue, and saturation space (IHS). For simplicity, the images were acquired with a limited field of view of 36° pointing to the camera's zenith to prevent direct sunlight from reaching the internal charge-coupled device (CCD) on the camera. For a priori–classified clear-sky images, the accuracy of the method was superior to 94%. For overcast-sky conditions, the corresponding accuracy was larger than 99%.

A comparison test was performed with two human observers and our method. The results for the 29 images collected for several time of days during 50 days in 1999 summer were compared to visual observations of these same digital images by two trained field meteorologists. Correlation coefficients between human observers and the automatic method ranged from 0.84 for clear-sky conditions, and the lowest was 0.09 for undefined-sky conditions.

1. Introduction

Clouds are responsible for about 23% of the planetary albedo, which allows them to play an important role in the radiative balance in the atmosphere (Mc-Cartney 1975; Iribarne and Cho 1980; Arking 1990; Hobbs 1993). Climatic models usually require reliable information on cloud cover as input in order to produce reliable results (Seluchi and Chou 2001). Long et al. (1999) use fractional sky cover measurements obtained by a sky imager instrument to establish a relationship between the sky cover amount and diffuse cloud

effect, which is of importance in radiative balance studies of the atmosphere. Furthermore, because clouds are the most important factor modulating the solar radiation incident on the earth's surface, information on cloud cover state is also required for radiative transfer models for several applications, namely, for the assessment and usage of renewable energy (Pereira and Colle 1997). Planetary albedo has been estimated by using images from geostationary satellites [Geostationary Operational Environmental Satellite (GOES) and Meteorological Satellite (Meteosat)], and from oblique and polar orbit satellites [National Oceanic and Atmospheric Administration (NOAA), Nimbus, Satelite de Coleta de Dados (SCD)-2)] (Veissid and Pereira 2000). However, because of the large spatial and temporal variability, the routine assessment of cloud cover state is usually very imprecise. The assessment of the cloud

Corresponding author address: M. P. Souza-Echer, INPE/CPTEC, P.O. Box 515, 12201-970 São José dos Campos, São Paulo, Brazil.
E-mail: mariza@dge.inpe.br

cover state is most often accomplished by visual observation of the sky by trained meteorological technicians. Additional cloud cover information has been obtained by satellite-based image processing.

Cloud cover determination through the use of satellites has major inadequacies resulting from the large time and space variability of clouds with scales smaller than those allowed by most of today's meteorological satellites. A comparative analysis of six satellite-based techniques for the determination of cloud cover state was performed by Wielicki and Parker (1988). They have shown that the results depend on the spatial resolution and coverage. The techniques that were evaluated were the reflectance threshold, temperature threshold, bispectral technique method (BTM), hybrid bispectral threshold method (HBTM), spatial coherence, and functional box counting. Satellite time resolution also imposes difficulty in the assessment of cloud cover states. The satellites usually employed for this purpose, GOES and Meteosat, yield data with time resolutions between 15 and 30 min, and in several regions, such as over South America, it is not more than 1 h at most.

The techniques that employed the reflectance threshold and temperature threshold are based on monospectral bands, which test each pixel for reflectance (visible) or temperature (infrared) reaching a threshold that defines cloudy or clear-sky condition. The next two methods, BTM and HBTM, are bispectral; for example, they use images in both visible and infrared bands in order to determine if a pixel is clear or cloudy. The spatial coherence and functional box counting methods are based on the spatial variability inside a cloud field to determine the cloud cover state. The functional box counting technique uses the spatial variability revealed by the reflectance threshold in order to determine the cloud cover state in a lower resolution than that observed by the satellite. All these techniques require validation by ground truth data, that is, cloud cover observed at the surface (U.S. Department of Energy 1996).

Determination of the sky cloud cover fraction is usually done through visual inspection by trained human observers in meteorological stations (U.S. Department of Energy 1996). When it is necessary to determine cloud cover in a qualitative manner, photography of the sky can be used (Santos 1957). After 1970, digital photo techniques received a great advance resulting from the development of the optical astronomy and remote sensing research (Slater 1980; Balsamo et al. 1997). More recently, digital photography is becoming popular, and prices for digital cameras are falling rapidly as quality increases. Today there are several ground-based digital devices and cloud algorithms, some of which are commercially available. Davis et al. (1992) describe a region-based method for cloud imagery based on the digital interpretation of scanned slide pictures taken by a conventional camera. Their cloud identification method is based on the segmentation of the images in areas with similar color and texture. Feister et al. (2000) presented a new version of a technique earlier described by Shields et al. (1998) that applies filters in the blue, red, and near-infrared (NIR) range of the spectra. They employ ratios between the images in the visible and in the red or NIR for the cloud imagery. The whole-sky imager also provides sky radiance. Their results are reported for both daytime and nighttime conditions. Another commercially available instrument is the total-sky imager (TSI) by Yankee Environment System (information available online at http://www.yesinc.com/products/data/tsi880/). This equipment provides full-color sky images and real-time processing as well as cloud cover or sky conditions. Crowford et al. (2003) applied TSI to study cloud impacts on UV spectral irradiance.

Sabburg and Wong (1999) describe a ground-based sky camera system for studying the effect of clouds on the UV flux level. Their technique also uses a digital camera for data acquisition followed by the application of a thresholding algorithm. In this experiment, the camera is adapted to track the sun's disk and uses the area around the sun as region of interest.

The objective of the present work is to describe a simple operational methodology to determine the sky cloud cover to be continuously used at the Brazilian Antarctic Station Ferraz (62°05′S, 58°23.5′W) and other remote sites in Antarctica. This work was carried out in response to the need to better access information on clouds for several climatic and environmental studies in that area. Cloud cover status is key information for developing climatic models in Antarctica and has important direct implications in regards to the climate in Latin America. This region is important for climatic studies involving the radiative balance of the atmosphere (Harrison et al. 1993) where a correct evaluation of the cloud cover state is necessary. Usually the sky state at Ferraz is estimated by visual inspection by the station's meteorologist with all of the subjectiveness of this procedure. Sky images are collected at ground level by a digital camera and are analyzed by a computing code in order to classify the images pixel-by-pixel according to their cloud pixels and thus derive one quantitative index of cloud cover state. Our goal is to provide a simple, low-cost system to operate in the existing automatic weather stations in remote areas where the costs to maintain routine man-made observations is prohibitive, such in the Antarctic continent, although

several methods for automatic cloud cover determination do exist. The main difference of our method is the use of the information contained in the image saturation instead of colors or ratios between colors, like most other reported methods do.

2. Methodology

The images were collected at the Brazilian Antarctic Station Ferraz, in the Antarctic Peninsula. The experimental work was performed during the second part of the summer, during 50 days, from January to March 1999 (Souza 1999).

We used a Pixera professional digital camera, model PCS20232. The optical package used in this experiment was a C-mount lens (F2.1, focal length = 5 mm). The camera was mounted on the roof of one of the station's laboratories on a metal support, and it was electrically/electronically connected to a PC microcomputer inside for data acquisition and storage, with the field of half view (36°) pointing to the camera's zenith. We did not use a wide-angle lens because of the image pileup effect at the border, which may cause an overestimation of the cloud coverage. Our proposal is to compensate for the smaller area in the sky covered by each instantaneous image by increasing the number of images (trading area for short acquisition time). The images were returned to Brazil for analysis at the end of the mission. In the future, this procedure will be completely automated, and the algorithm run in the field will provide real-time information on the cloud state. All images were collected in the default mode of the equipment on a 24-bit red–green–blue (RGB) JPEG format at a resolution of 640×340 pixels. Data were collected starting at 0800 and ending at 2000 LT. We did not have to be concerned with the shadowing of the camera lens because the solar elevation never reached an angle high enough to let the direct solar beam reach the camera lens.

The principle behind this methodology is based on the optical characteristics of clouds in the visible range of the solar spectrum. Clouds are dynamic systems characterized by a high reflectance, multiple scattering, and predominantly white color, with color shades ranging from blue to red. On the other hand, the clear sky in the background is characterized by blue tones, with shades ranging from green to red, and can be considered as a quasi-static system for short time intervals (Schaefer and Day 1981). The solar radiation intensity at the top of the atmosphere is a function of solar declination and zenith angle for a given geographical region (Iqbal 1983).

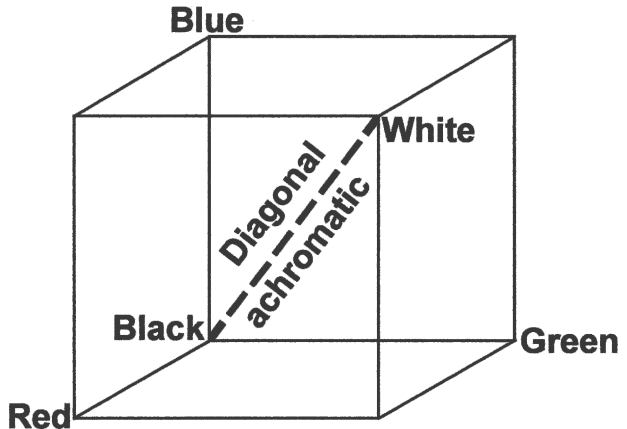We developed a method that takes into account the different wavelengths captured by the camera. The

Fig. 1. RGB color cube with primary color lines that represents red, green, and blue colors, and the achromatic diagonal with the origin in the intersect point between red, green, and blue.

method encompasses the transformation of images collected by the digital camera from the usual RGB color space to the intensity, hue, and saturation (IHS) space. The IHS color attribute space, as the name suggests, does not work directly with colors, as RGB does, but with the color's attributes, which ultimately is how the human eye perceives the sources.

The RGB model is commonly used by television manufacturers, based on the chromaticity diagram, which is, in effect, a graphic representation of color and saturation for a constant brightness (total intensity of radiation).

Figure 1 shows the RGB color cube with primary color lines that represent red, green, and blue colors. The figure shows the achromatic diagonal with the origin at the intersect point between red, green, and blue. This intersecting point represents the black color; the extreme of this diagonal represents the white color. Figure 2 shows two representations of the IHS space. Figure 2a shows a color triangle, with the vertex representing the red (R), green (G), and blue (B) colors. In the center of this triangle there is a point that represents an achromatic point linking each vertex. The color saturation increases from the center to the edge of the triangle, and the angle between hue (H) and point P shows the selected color. Figure 2b shows the combination between hue, saturation, and intensity in the three dimensions of the color space.

In the IHS system, the intensity (I) is the measurement of the total energy in all wavelengths that reaches the eye. Intensity is responsible for the brightness sensation and is obtained from Eq. (1):
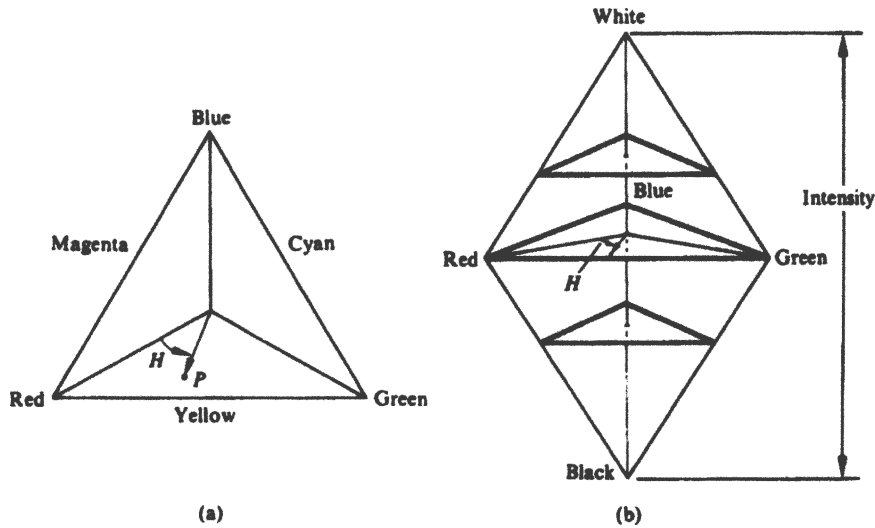
$$I = \frac{1}{3}(R + G + B). \qquad (1)$$

FIG. 2. The IHS space: (a) color triangle, with each vertex representing the red, green, and blue color. In the center of this triangle there is a point that represents the achromatic point linking each vertex. (b) The combination between hue, saturation, and intensity in the three dimensions of the color space. (From Gonzalez and Woods 1992.)

The color hue (H) is described as the measurement of the reflected or emitted spectral light, thus defining the object color and is obtained from Eq. (2):

$$H = \cos^{-1}\left\{\frac{1/2[(R-G)+(R-B)]}{[(R-G)^2+(R-G)(G-B)]^{1/2}}\right\}.$$

(2)

The saturation (S) is described by Eq. (3) and expresses color "purity";

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)].$$

(3)

Colors with high saturation are said to be pure, such as clear sky. Colors with low saturation (much gray) are pastel colors, such as clouds. More details regarding the mathematics of this transformation can be found in Gonzalez and Woods (1992), and Harley and Arthur (1993).

Through visual inspection, it is possible to see that clouds have high reflectance, a predominantly white color (caused by multiple scattering), and show color shades varying from blue to red diluted in a white matrix, which characterizes a mix of several wavelengths. On the other hand, clear skies during the daytime shows predominantly blue tones, with color shades ranging from green to red (Schaefer and Day 1981), which typically results in high saturation values. Following this scheme, clouds show lower saturation values, and clear skies shows high saturation values. Thus, saturation was selected as the component from the IHS

space to be used in the development of the classification algorithm presented in this paper.

Figure 3a shows a test image collected at Instituto Nacional de Pesquisas Espaciais's (INPE's)[1] headquarters in São Jose dos Campos, São Paolo, Brazil, selected to illustrate the method. The scheme for characterizing cloudy and clear skies can be seen in Fig. 3b and Fig. 3c. Figure 3b shows saturation for line 152 of the test image (shown in Fig. 3a). The saturation is the spectral variable that can be explored as a gauge for pixel contamination by clouds. Figure 3c is the saturation map for the entire image associated with the picture shown in Fig. 3a. The example in Fig. 3b shows for instance that pixels around 64 and 192 are characterized by saturation levels of around 100 and correspond to a clear sky, while pixels 128 and 640 have lower saturation levels and correspond to regions where clouds are undoubtedly present.

a. Pixel classification

Three classes were defined to classify image pixels:

- clear sky, identified by label 1;
- cloud sky, identified by label 2;
- undefined, identified by label 3.

Figure 4 shows the three predefined classes of sky state in an image. An undefined pixel represents the transition of image saturation values resulting from the trans-

---

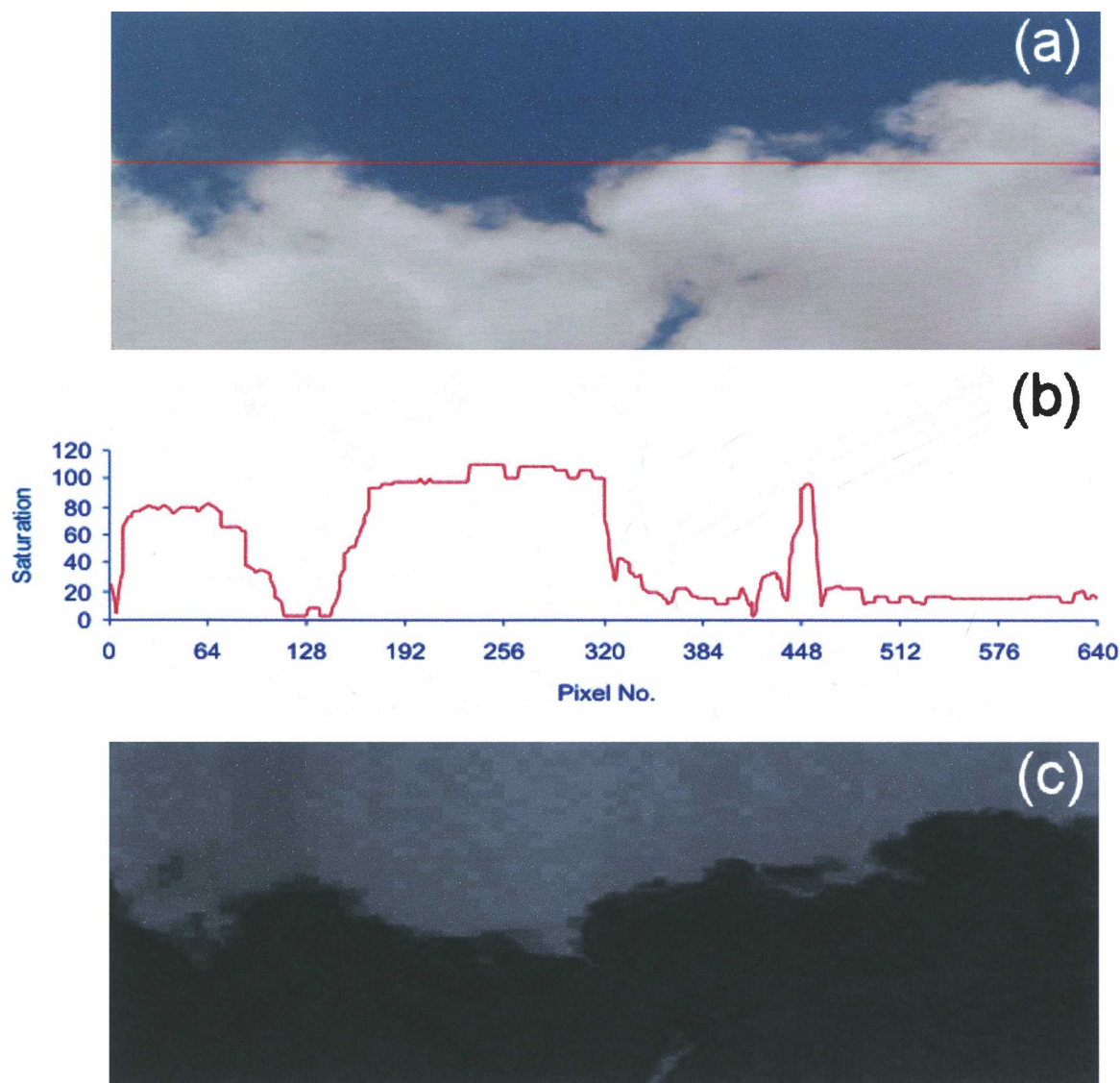[1] The Brazilian National Institute for Space Research.

FIG. 3. (a) Test image collected at INPE's headquarters in São José dos Campos, Brazil. [The red line indicates the saturation curve shown in (b).] (b) Change in saturation for line 152 of the test image, and (c) saturation map for the test.

lucent cloud zone. Other optical effects such as scattering by aerosols may also assign a pixel as undefined. A label was associated with each of the predefined cloud states according to each class. Figure 4 shows training areas chosen as classes used to teach the classifying algorithm.

b. Classification and thresholding

The classifying algorithm implemented in this work is the simplified parallelogram, where each class is represented by an upper and a lower threshold for each attribute used in the classification process. Thus, in the case of a one-dimensional space of attributes, the par-
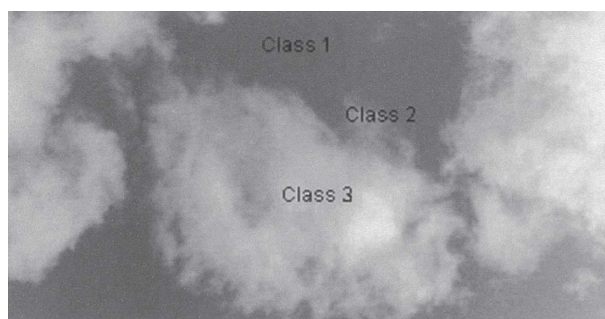

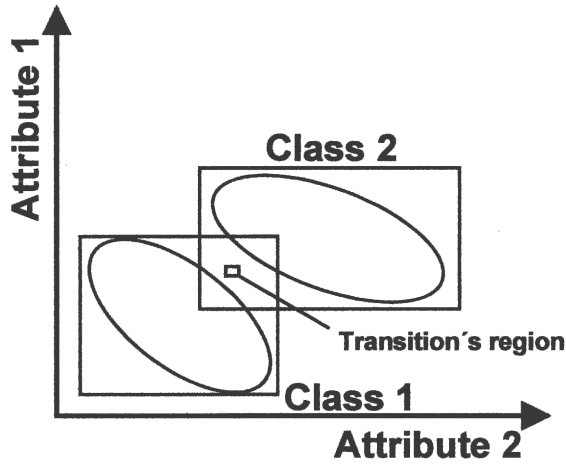FIG. 4. The three predefined classes of sky state shown in an image.

FIG. 5. Example of classification using the supervised parallelepiped method.

tition associated with a certain class will be a line segment. In the two-dimensional space, the partition will be a rectangle, and in the three-dimensional space the partition will be a parallelepiped, and so on. A pixel is then classified as belonging to a class in a region whose attributes are within predefined ranges between the upper and lower thresholds of each attribute.

Figure 5 shows an example of classification using the supervised parallelepiped method. It is possible to see the separation of the two classes through the ellipses, representing the type-1 and -2 classes in the space of attributes. It is also possible to observe a rectangle defined by solid lines. It represents both the pixels inside the ellipses and their neighboring pixels. This area inside the space of attributes is designated as being the transition region.

To determine the upper and lower thresholds for each class, the average and standard deviations of the values of the saturation attribute were used in the training area. Thus,

$$\text{lower threshold} = \mu_i - n\sigma_i,$$

$$\text{upper threshold} = \mu_i + n\sigma_i. \quad (4)$$

In these equations, μ_i is the group average, σ_i is the class standard deviation, and i equaling 1, 2, and 3 represents clear, cloud, and undefined classes, respectively. Assuming that attributes follow a normal distribution, then for $n = 3$, the area of the normal distribution in the range of $\mu - 3\sigma$ and $\mu + 3\sigma$ will be 0.9973. Thus, $n = 3$ defines thresholds that comprise 99.7% of all pixels falling within this specific class. Figure 6 shows a plot with a standard deviation and averages for the clear-sky and the cloudy-sky classes. Of course, this type of train-

ing will work only if the averages of each class are well separated; for example, a class overlapping in the chosen class interval must not occur. If this condition is met, one has

a pixel belongs to the class *cloud* if $x < \mu_n + 3\sigma_n$,

a pixel belongs to the class *clear sky* if $x > \mu_c - 3\sigma_c$,

a pixel belongs to the class *undefined* if $\mu_n + 3\sigma_n \leq x$

$$\leq \mu_c - 3\sigma_c.$$

A set of test images was acquired several times during the day. These images were processed according to the three-group classification scheme. These classes were used for the standardization procedure in the determination of the image thresholds for clear-, cloudy-, or undefined-sky-state pixels.

To construct the threshold intervals for each class, we have selected by visual inspection 20 clear-sky images and 20 overcast-sky images. From these reference datasets, the thresholds (average, standard deviation) for each class were defined, and Table 1 shows the results for μ_n, σ_n, μ_c, and σ_c. For algorithm testing, 29 images were selected in several local times and different days during the 50 days of study in the summer of 1999. These digital images were analyzed by human observers who did not know the algorithm results, and the estimates are compared.

3. Results and discussion

The validity of the method described in this work is limited to high latitudes, where the solar elevation is small all year round and the sun never enters the field of view of the camera lens setup (36°). At lower latitudes, it is necessary to add a shading device to the optical system to avoid the direct incidence of the sun over the charge-coupled device (CCD) system. In addition, there are problems arising from the scattering of solar radiation by the sky and clouds close to the apparent sun position that would need to be handled by additional image treatment techniques.

Our simple method was tested by using a package of 29 images representing several situations of cloud cover in different days and local times that we judge to be representative of the study area in the Antarctic Peninsula. The results show that the method was capable of defining the cloud cover status with an accuracy of better than 94% for clear skies and 99% for cloudy skies. These accuracies were determined from the percentage of pixels that the algorithm correctly identified according to human observers. Thus, in a set of images that were previously known by visual inspection as being
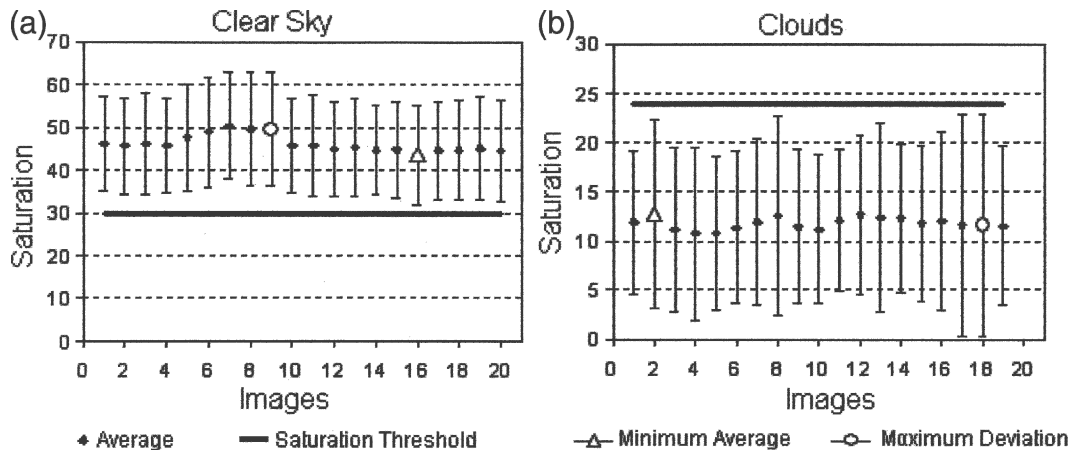
FIG. 6. (a) Clear-sky and (b) cloudy-sky cutoff threshold.

clear-sky images, the method was capable of classifying pixel by pixel as such in more than 94% of the pixels in each image. The accuracy for totally overcast skies was analogous.

Table 2 shows the result of the classification process performed over the package of 29 images. The classifications are given in percentage of pixels in relation to the total number of pixels in the full image. The automated classification results are shown in the first three columns as clear-, undefined-, and cloudy-sky percentages. The next columns correspond to the estimates made by observers for the same images. Estimates labeled on "observer 1" were made by a trained INPE field observer. The column, labeled "observer 2" corresponds to a combination of visual inspection of the same images performed by two Instituto Nacional de Meteorologia (INMET)[2] field observers. Therefore, the two reference datasets are representative of visual observation and bear their characteristic subjective nature. Because of logistic limitation and elevated cost it was not possible to carry out this part of the work in real time, with observers in the field during the experiment. Thus, a package of printable copies of the same images was sent to each observer for classification in their home institutions. In all cases, the decision on cloudiness was made by using procedures of visual inspection of images and selecting regions of interest, similar to what would have been made in the field. The idea here is not to contrast conventional with automatic methods for cloud cover estimation, but rather to understand some basic differences between the two, mainly with respect to the effects of the subjectiveness of the results.

By comparing the human classifications with the results from the automatic classifying algorithm, we can easily detect several discrepancies between their resulting values. Furthermore, classifications made by each individual observer over the same set of images often do not agree with each other as well. For example, on images 12 and 25 the visual classification indicated a totally overcast sky, whereas the automated method yielded variable values of cloud amount. Another example occurs on image 24, where the sky was classified as having variable amounts of clouds by observer 1, and was 100% undetermined as noted by observer 2. We believe the discrepancies between observers occur because of the subjective nature of the visual observations. Discrepancies between the observer and the method are mostly linked to a weakness of the method, possibly resulting from the training procedure illustrated in Fig. 4.

Figure 7 shows plots with the correlation results for the 29 images listed in Table 2. The first three plots show the clear-sky results. It is observed that data from the two human observers shown in the first plot yielded a correlation factor of 0.81 between each other, with only a small bias and one probable outlier for observer 1 at 10%. The correlation for our method versus that of observer 1 yielded a comparable correlation coefficient, but with a much larger bias for low clear-sky percentages. Results for observer 2 degraded a great deal, both

---

[2] The Brazilian National Institute of Meteorology.

TABLE 1. Saturation cutoff thresholds.

Variable	Parameter	Value
μ_n	Clouds average	12.7
σ_n	Clouds standard deviation	3.7
μ_c	Clear-sky average	45.3
σ_c	Clear-sky standard deviation	4.4

TABLE 2. Total results of the classification for cloudiness images.

| Image | Sky | Classified | | Sky | Observer 1 | | Sky | Observer 2 | |
		Undefined	Clouds		Undefined	Clouds		Undefined	Clouds
1	85.3	3.7	11.0	55.0	10.0	35.0	62.5	0.0	37.5
2	56.7	6.3	37.0	60.0	5.0	35.0	50.0	12.5	37.5
3	65.1	3.8	31.1	55.0	15.0	30.0	37.5	25.0	37.5
4	42.3	7.8	50.0	10.0	70.0	20.0	0.0	75.0	25.0
5	46.8	20.4	32.8	45.0	15.0	40.0	50.0	12.5	37.5
6	36.2	19.1	44.4	40.0	20.0	40.0	25.0	25.0	50.0
7	45.9	16.6	37.5	25.0	30.0	45.0	37.5	12.5	50.0
8	72.6	3.8	23.6	55.0	10.0	35.0	62.5	0.0	37.5
9	77.6	3.3	19.1	55.0	5.0	40.0	50.0	0.0	50.0
10	52.8	4.5	42.8	30.0	5.0	65.0	25.0	0.0	75.0
11	46.3	5.7	48.1	30.0	5.0	65.0	25.0	0.0	75.0
12	15.7	18.1	66.2	0.0	5.0	95.0	0.0	12.5	87.5
13	19.4	16.6	64.0	0.0	5.0	95.0	0.0	12.5	87.5
14	32.9	3.6	63.4	0.0	50.0	50.0	0.0	50.0	50.0
15	32.5	16.4	51.1	0.0	30.0	70.0	0.0	25.0	75.0
16	39.4	7.4	53.2	5.0	30.0	65.0	0.0	37.5	62.5
17	22.2	14.4	56.4	5.0	20.0	75.0	12.5	12.5	75.0
18	38.1	18.1	43.8	10.0	10.0	80.0	25.0	0.0	75.0
19	31.5	16.1	51.6	10.0	10.0	80.0	62.5	0.0	37.5
20	13.8	3.2	82.9	0.0	10.0	90.0	0.0	25.0	75.0
21	18.9	23.4	57.7	0.0	20.0	80.0	0.0	25.0	75.0
22	18.6	13.6	67.8	0.0	35.0	65.0	0.0	37.5	62.5
23	19.2	20.0	60.8	0.0	40.0	60.0	0.0	25.0	75.0
24	46.0	11.6	42.4	0.0	65.0	35.0	0.0	100.0	0.0
25	0.1	0.2	99.7	0.0	5.0	95.0	0.0	12.5	87.5
26	18.1	0.8	81.1	10.0	5.0	85.0	25.0	0.0	75.0
27	20.4	4.1	75.5	15.0	20.0	65.0	37.5	0.0	62.5
28	23.2	2.8	74.1	15.0	25.0	60.0	37.5	0.0	62.5
29	24.9	4.1	71.0	10.0	10.0	80.0	12.5	37.5	50.0

in terms of the correlation coefficient and in terms of bias. This indicates that both observers are producing consistent information on cloud cover independent of its amount, while the classifier is apparently producing an overestimation for low values of cloud cover in comparison with that of the observer's. Nevertheless, one cannot conclude at this stage that our method is producing results that are less accurate than those produced by the human observers, but rather that the observers are producing mutually consistent results. Classifications for cloudy pixels have shown the same 0.81 correlation coefficient between observers, but have much poorer results when compared to those of the human observers. The worst cases were observed for pixels classified as undefined. In such cases, the classifier has shown almost no correlation with the results from human observations. We believe this apparently poor result reflects the subjective character of the human observations when dealing with uncertainties.

Table 3 shows the relative values for normalized root-mean-square error (NRMSE)[3] and normalized mean bias error (NMBE)[4] for the 29 images listed in Table 2. The NRMSE shows that the variation is higher for both observers and our method to undefined conditions. Next came the clear-sky conditions with 0.6; similar results were found for observer 1. But, for cloud conditions the method of observer 1 versus our method showed the lowest value for NRMSE, 0.3. These results show that the large errors are observed for undefined-sky conditions, which was observed in the linear correlation coefficients.

For clear-sky conditions, observer 1 presented a larger bias (-0.5) when compared with our method. For undefined conditions observer 1 and observer 2 shows the similar results: overestimation with NMBE equals 1. For cloud conditions, both observers have pre-

[3] $NRMSE = [(\Sigma_{i=1}^N (y_i - x_i)^2/N)]^{(1/2)} \times (\Sigma x_i/N)^{-1}$.
[4] $NMBE = [\Sigma_{i=1}^N (y_i - x_i)/\Sigma x_i]$, where X_i is sky conditions from our method and Y_i is sky conditions from observers.
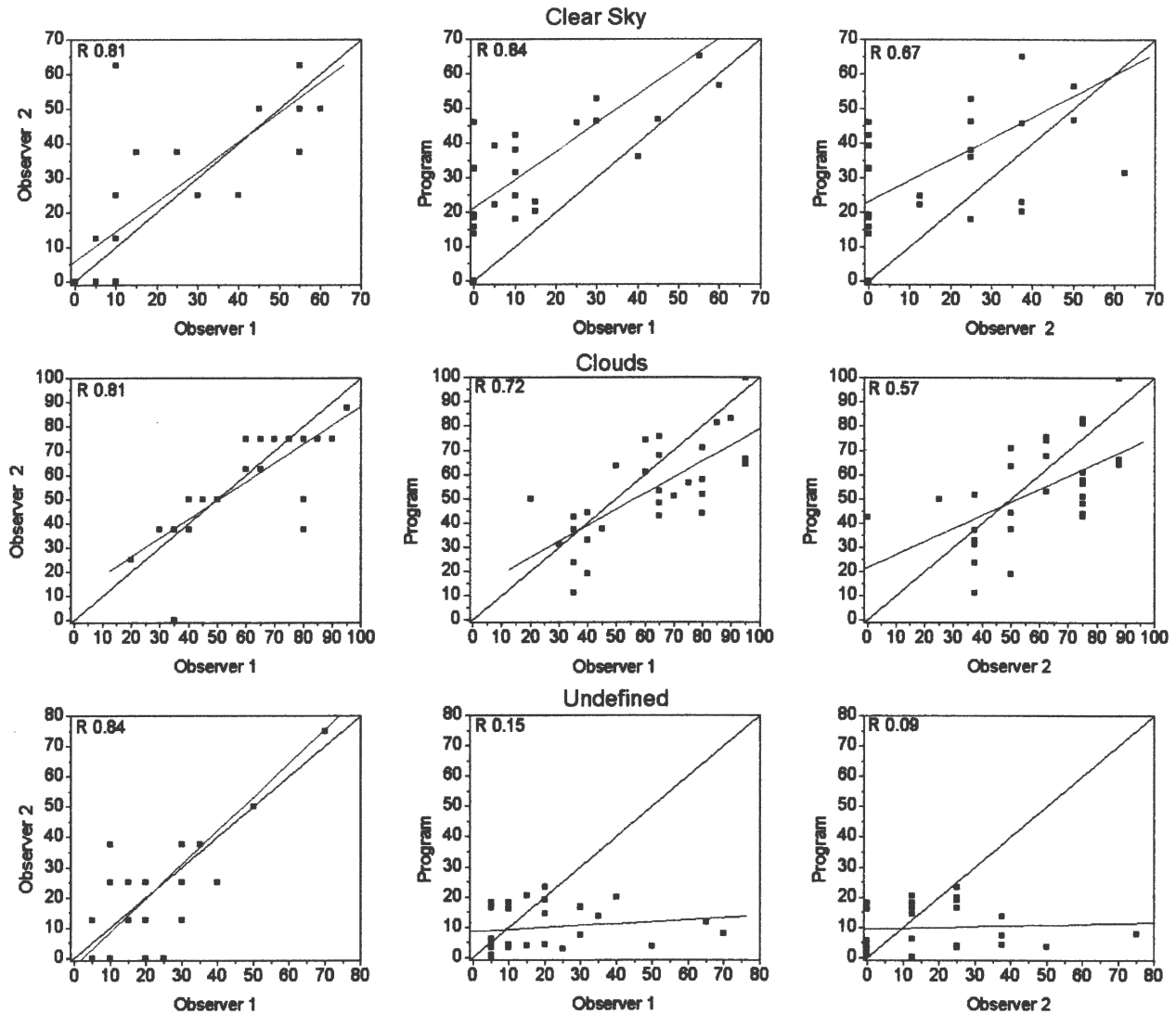
Fig. 7. Correlation among human observers and the automatic method: (top) clear-sky conditions, (middle) cloudy-sky conditions, and (bottom) undefined conditions; (left) observer 1 × observer 2, (middle) observer 1 × program, and (right) observer 2 × program.

sented good agreement with our method results. These results imply that our method overestimates clear-sky fractions given by observers 1 and 2.

To further try to understand the results from our method, we applied the technique of cluster analyses by

TABLE 3. Normalized values for rmse and MBE.

	NRMSE	NMBE
Observe1_Sky vs Method _Sky	0.6	−0.5
Observe2_Sky vs Method _Sky	0.6	−0.4
Observe1_Undefined vs Method_Undefined	2.1	1
Observe2_Undefined vs Method_Undefined	2.6	1
Observe1_Cloud vs Method_Cloud	0.3	0.2
Observe2_Cloud vs Method_Cloud	0.4	0.1

employing a tree-clustering algorithm available in the "Statistica" software package (StatSoft 1999). The purpose of this algorithm is to join together objects into successively larger clusters, using some measure of similarity or linkage distance between results (Davis 1973). By doing so, the relation between the groups of variables is revealed. Results of this analysis are usually displayed as dendrograms or tree diagrams in which the vertical axis represents the linkage distance and the horizontal axis shows the variables, wherein the larger the linkage distance, the lower the similarity between variables. Figure 8 shows a tree diagram based on the Euclidian linkage distance between the results determined from observers 1 and 2 and the classifier program. Herein O2_N, O1_N, and P_N represent cloudy
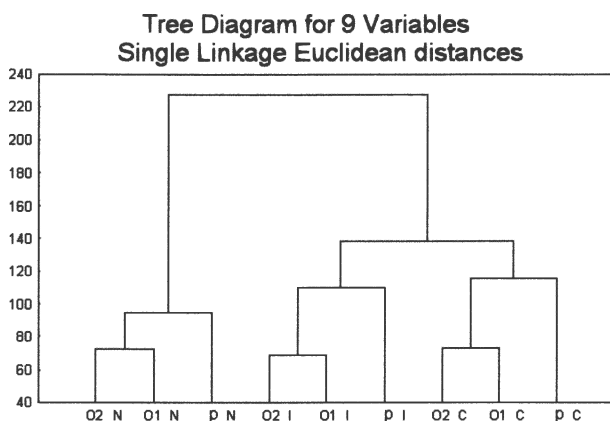
FIG. 8. Tree diagram for nine variables—single linkage Euclidian distances.

pixels variables from observers 1, 2, and the classifier program; O2_I, O1_I, and P_I represent the variables for undefined pixels, and the other variables represent clear-sky pixels. It shows that results for clear and cloudy skies belong to distinct clusters, both for our method and for the observers. This is an expected result because both sky states are mutually exclusive and constitute by far the most common observations.

The undefined pixels, however, constitute a separate cluster closer (almost the same Euclidian linkage distance) to the clear sky in all the observer's datasets. This suggests that all classification methods (human or computer) classified an undefined pixel as a special case of clear sky, or a contamination of clear-sky pixels by clouds. Going up to higher-level branches of the clustering method, we can see that the cluster analyses discriminate between observers and our method in all cases (cloudy, undefined, and clear sky). The information provided by this cluster analysis gave us a hint on the consistency of the automatic process and on the adequacy of the training process. For example, it was interesting to note that the cluster analysis "sees" the output from observers and from the method in the same low-level cluster that includes clear and undefined skies together. This is an indication that the correct training has being transferred to our method. Furthermore, it is clear in the clustering process that the observers and the method always fall into distinct higher-level clusters. In our interpretation, this is an indication of the subjective nature of visual observations. Perhaps if the automatic method was trained indefinitely it would ultimately reproduce exactly the same results as those given by human observer, but we doubt that the subjectiveness would ever be incorporated into the automatic procedure. Of course, the limited number of images precludes us from coming to a conclusion on

that. The cluster an analysis was carried out to validate our methodology and, to a certain extend it was successful.

These results support the findings of Fig. 7 by showing that the estimates obtained by human observers are consistent with each other. This consistency is probably linked to the human quality of the selection procedure with a great deal of subjectivity but is based on the same process of judgment. On the other hand, our method yielded results that are inconsistent with the observer's, predominantly in the results of undefined-sky states. Nevertheless, this does not necessarily mean that the sky-state estimates from the computer algorithm are less accurate than those from human observers. We can conclude that our method is producing results that reflect the machine nature of the selection process.

REFERENCES

Arking, A., 1990: The radiative effects of clouds and their impact on climate. *World Climate Program. Res.,* **399,** 1–39.

Balsamo, E. P., M. Candidi, G. Consolini, and L. Morici, 1997: The Italian all-sky camera for auroral observations at Terra Nova Bay (Antarctica). *Proc. Eighth GIFCO Conf.: Cosmic Physics in the Year 2000*, Vol. 58, Bologna, Italy, SIF, 101–104.

Crowford, J., R. E. Shetter, B. Lefer, B. Cantrell, W. Junkermann, S. Madronich, and J. Calvert, 2003: Cloud impacts on UV spectral actinic flux observed during the International Photolysis Frequency Measurement and Model Intercomparison (IPMMI). *J. Geophys. Res.,* **108,** 8545, doi:10.1029/2002JD002731.

Davis, C. D., 1973: *Statistical and Data Analysis in Geology.* John Wiley and Sons, 550 pp.

Davis, G. B., D. J. Griggs, and G. D. Sullivan, 1992: Automatic estimation of cloud amount using computer vision. *J. Atmos. Oceanic Technol.,* **9,** 81–85.

Feister, U., J. Shields, M. Karr, R. Johnson, K. Dehne, and M. Woldt, 2000: Ground-based cloud images and sky radiances in the visible and near infrared region from whole sky imager measurements. *Proc. Climate Monitoring–Satellite Application Facility Training Workshop*, DWD, EUMETSAT, and WMO, 10 pp.

Gonzalez, R. C., and R. E. Woods, 1992: *Digital Image Processing.* Addisson-Wesley, 509 pp.

Harley, R. M., and R. W. Arthur, 1993: *The Pocket Handbook of Image Processing Algorithms.* C. Prentice-Hall, 303 pp.

Harrison, E. F., P. Minnis, B. R. Barkstrom, and G. G. Gobson, 1993: Radiation budget at the top of the atmosphere. *Atlas of Satellite Observations Related to Global Change*, R. J. Gur-

ney, J. L. Foster, and C. Parkinson, Eds., Cambridge University Press, 19–38.

Hobbs, P. V., 1993: *Aerosol-Cloud-Climate Interactions*. Academic Press, 235 pp.

Iqbal, M., 1983: *Introduction to Solar Radiation*. Academic Press, 390 pp.

Iribarne, J. V., and H. R. Cho, 1980: *Atmospheric Physics*. Reidel Publishing, 212 pp.

Long, C. N., T. P. Ackerman, J. J. DeLuisi, and J. Augustine, 1999: Estimation of fractional sky cover from broadband SW radiometer measurements. Preprints, *10th Conf. on Atmosphere and Radiation*, Madison, WI, Amer. Meteor. Soc., 383–386.

McCartney, J. H., 1975: *Optics of the Atmosphere—Scattering by Molecules and Particles*. John Wiley and Sons, 408 pp.

Pereira, E. B., and S. Colle, 1997: A energia que vem do sol. *Ciência Hoje,* **22,** 13 024–13 035.

Sabburg, J., and J. Wong, 1999: Evaluation of a ground-based sky camera system for use in surface irradiance measurement. *J. Atmos. Oceanic Technol.,* **16,** 752–759.

Santos, J. M., 1957: Contribuição aos estudos de albedo. Ph.D thesis, Escola Superior de Agricultura Luiz de Queiroz, 47 pp.

Schaefer, V. J., and A. J. Day, 1981: *A Field Guide to the Atmosphere*. Houghton Mifflin Company, 359 pp.

Seluchi, M. E., and S. C. Chou, 2001: Evaluation of two Eta Model versions for weather forecast over South America. *Geofis. Int.,* **40,** 219–237.

Shields, J. E., R. W. Johnson, M. E. Karr, and J. L. Wertz, 1998: Automated day/night whole sky imagers for field assessment of cloud cover distribution and radiances distributions. *Proc. 10th Symp. on Meteorological Observation and Instrumentation*, Phoenix, AZ, Amer. Meteor. Soc., 165–170.

Slater, P. N., 1980: *Remote Sensing—Optics and Optical Systems*. Addison-Wesley, 575 pp.

Souza, M. P., 1999: Desenvolvimento de uma nova metodologia para determinação de cobertura efetiva de nuvens. M.S. thesis, Geophysics Space–INPE, 107 pp.

StatSoft, 1999: STATISTICA for Windows III. 2d ed.

U.S. Department of Energy, 1996: Science plan for the Atmospheric Radiation Measurement program (ARM). DOE/ER-0670T.

Veissid, N., and E. B. Pereira, 2000: Estimativa do Albedo Planetário Empregando Dados do Experimento Célula Solar do Satélite Brasileiro SCD2. *Braz. J. Geophys.,* **18,** 25–28.

Wielicki, B. A., and L. Parker, 1988: Cloud properties observed using Landsat satellite data. *Current Problems in Atmospheric Radiation*, J. Lenoble and J. Geleyn, Eds., A. Deepak, 161–165.